

# Kategorientheoretische Ontologieentwicklung und Wissensmodellierung für die Digital Humanities

**Gerstorfer, Dominik**

gerstorfer@linglit.tu-darmstadt.de  
TU Darmstadt, Germany

## Einleitung

Formale Modellierung von Wissen spielt in den Digital Humanities eine nicht zu unterschätzende Rolle. Daten müssen organisiert, kategorisiert, gespeichert, abgerufen und verarbeitet werden. Hierzu werden häufig Datenbank- und Ontologiesprachen wie SQL und RDF/OWL oder auch semantische Netzwerke eingesetzt, die als *informatische* Systeme robust und etabliert sind. Ihre Verwendung setzt zumeist nach der eigentlichen geisteswissenschaftlichen Arbeit an, d.h. erst, wenn die konzeptuelle Arbeit abgeschlossen ist, werden die so gewonnenen Ergebnisse formalisiert. Dieses Vorgehen birgt das Risiko, dass sich die, im ersten Arbeitsschritt gewonnenen Ergebnisse nicht ohne Weiteres formalisieren lassen. Die Gründe hierfür sind vielfältig und reichen von unzureichender Rigidität der informellen Analyse bis zur mangelnden Expressivität des verwendeten Formalismus. Hinzu kommt, dass bestehende Formalismen die kreative Beschäftigung mit den Inhalten erschweren und die Beherrschung der technischen Systeme erfordert häufig eine steile Lernkurve.<sup>1</sup>

In diesem Beitrag werden *ontology logs* (*ologs*) als Framework für konzeptuelle Arbeit und Wissensmodellierung in den Digital Humanities vorgestellt. *Ologs* sind im Bereich der angewandten mathematischen Kategorientheorie als Framework entwickelt worden, mit dem Ziel ein gleichermaßen mächtiges wie benutzerfreundliches Modell zur Wissensrepräsentation bereitzustellen, das in der Lage ist, die geisteswissenschaftliche Forschung zu unterstützen und leichter an mathematisch-informatische Theorien und Techniken anschlussfähig zu machen.

## Kategorientheorie

Die mathematische Kategorientheorie wurde in den 1940er-Jahren von Saunders Mac Lane und Samuel Eilenberg entwickelt, um verschiedene mathematische Felder und Theorien zu vergleichen. Die grundlegende Einsicht hinter der Kategorientheorie formuliert Mac Lane (1998: 1) wie folgt: “Category theory starts with the observation that many properties of mathematical systems can be unified and simplified by a presentation with diagrams of arrows.”

Eine Kategorie  $C$  besteht aus einer Sammlung von Objekten  $Ob(C)$  und Pfeilen ( $\rightarrow$ ) zwischen Objektpaaren  $A, B \in Ob(C)$ , so dass  $f: A \rightarrow B$ . Des Weiteren existieren Pfeile zwischen zwei Kategorien  $C$  und  $D$ , sodass  $C \rightarrow D$  eine strukturerhaltende Abbildung ist, die Funktor genannt wird. Diese einfache Systematik und die Möglichkeit, Pfeile zu kombinieren eröffnen weitgehende Abstraktionsmöglichkeiten, die so weit gehen, große Teile der Mathe-

matik (Algebra, Topologie, Mengentheorie, Graphentheorie, Gruppentheorie, etc.) zu fundieren.

In der Regel wird die Kategorientheorie in der Mathematik verwendet, um Theorien zu analysieren und weiter zu verallgemeinern. Man könnte auch davon sprechen, dass es sich um Meta-mathematik handelt, die Mathematik als Anwendungsgegenstand hat.

“Eilenberg and Mac Lane introduced very abstract tools into mathematics, which seemed even too abstract. Nevertheless, they motivated their work with both technical merits, which allow for an effective study of the phenomenon of naturality, and conceptual advantages. They noted that the proposed conception is so general that it allows for the detection of the same structures in fundamentally different fields of mathematics. By finding new analogies between different fields of mathematics it suggests new results. Thanks to the fact that categorical glasses allow for the observation of the same structures in both topology and algebra, these glasses allow for a unifying view of mathematics. Already in 1945 it was clear that CT had the power to unify mathematics.” (Landry 2017: 3)

Und obwohl die Kategorientheorie als *allgemeine Theorie mathematischer Strukturen* ihren Anfang nahm, beschränken sich ihre Anwendungsmöglichkeiten nicht auf die Mathematik, sie kann auf viele weitere Gegenstandsbereiche appliziert werden, zu denen nun auch die Digital Humanities gehören.

## Angewandte Kategorientheorie

Die angewandte Kategorientheorie oder *applied category theory* (ACT) ist eine neuere Entwicklung,<sup>2</sup> mit dem Ziel kategorientheoretische Ideen aus der Mathematik in andere wissenschaftliche Disziplinen zu übertragen. So heißt es programmatisch auf der ACT-Website im Proposal zu einem Workshop mit dem klingenden Titel „Towards An Integrative Science“:

„[...] we should treat the use of categorical concepts as a natural part of transferring and integrating knowledge across disciplines. The restructuring employed in applied category theory cuts through jargon, helping to elucidate common themes across disciplines. Indeed, the drive for a common language and comparison of similar structures in algebra and topology is what led to the development category theory in the first place, and recent hints show that this approach is not only useful between mathematical disciplines, but between scientific ones as well.“ (ACT2018)

Dies ist unter anderem in der Physik (Abramsky / Coecke 2007; Baez / Stay 2010), Linguistik (Coecke et al. 2010), den Neurowissenschaften (Brown / Porter 2008), Informatik (Ehrig et al. 2001) und der Philosophie (Landry 2017) geschehen.

Gerade in der Informatik spielt die Kategorientheorie eine ausgezeichnete Rolle, da formale Logik und Mengentheorie in der theoretischen Informatik und funktionale Programmiersprachen wie Haskell oder Module wie Catlab.jl für Julia in der angewandten Informatik durch sie verbunden sind.

In den oben genannten Fällen hat sich gezeigt, dass die angewandte Kategorientheorie einige für die Digital Humanities attraktive Eigenschaften aufweist: Der hohe Abstraktionsgrad und der Umstand, dass nur Objekte und Pfeile in einer Kategorie vorkommen, erlauben einen leichten Einstieg in kategorientheoretische Modellierung und der Einsatz von kommutativen Diagrammen zur Analyse macht die Verwendung benutzerfreundlich. Gleichzeitig können Modellierungen aufgrund der Modularität

und Kompositionalität der Theorie auch sehr komplexe Sachverhalte repräsentieren, ohne selbst unüberschaubar zu werden. Darüber hinaus bietet die Kategorientheorie vielfältige Anschlussmöglichkeiten an andere mathematische Bereiche. Sollte sich herausstellen, dass formale Logik oder Graphentheorie benötigt wird, ist es ein Leichtes die nötigen Übergänge herzustellen.

In diesem Sinne kann die Kategorientheorie in den Digital Humanities als leichtgewichtiges Modellierungstool eingesetzt werden, das nach Bedarf erweitert und skaliert werden kann. Dies soll nun anhand der von Spivak und Kent (2012) entwickelten *Ontology Logs* illustriert werden.

## Was sind und was können *ologs*?

*Ologs* sind eine konkrete Anwendung der Kategorientheorie zur Wissensmodellierung. Der Ausgangspunkt ist natürliche Sprache, darum ist dieses Modell besonders für die textbasierten Digital Humanities geeignet. Ein *olog* besteht zunächst nur aus einer Sammlung gelabelter Objekte und Pfeile, die funktionale Relationen zwischen den Objekten ausdrücken. Diese einfachen *ologs* können bei Bedarf um einfache und komplexe *Typen* erweitert werden, d.h. um die Charakterisierung einer Menge, in der das Objekt Element ist. Ebenso können die Pfeile weiter spezifiziert werden als *Aspekte* oder *Attribute*.

Zu beachten ist, dass *ologs* immer perspektivisch an einen bestimmten Standpunkt gebunden sind, d.h. ausgehend von einer individuellen Lesart wird eine Konzeption entwickelt und Schritt für Schritt mit mehr Informationen angereichert, bzw. in Zusammenarbeit mit anderen Lesarten abgeglichen. Die Konstruktion von *ologs* stellt sicher, dass das Ergebnis strukturell valide ist und nicht faktisch korrekt. Dies ermöglicht es, im Modell mit Diskrepanzen umzugehen und Gemeinsamkeiten und Differenzen zwischen Modellierungen mit den, durch die Kategorientheorie bereitgestellten mathematischen Werkzeugen zu bearbeiten. Die Darstellung als kommutative Diagramme erleichtern die Kommunikation über unterschiedliche Interpretationen des Untersuchungsgegenstandes und unterstützen den geisteswissenschaftlichen Reflexionsprozess.

Da es jederzeit möglich ist ein bestehendes *olog* um weitere Pfeile und Objekte zu erweitern, können neue Informationen – wie zusätzliche Aspekte und Attribute – hinzugefügt werden, ohne das bereits entwickelte Schema von neuem aufbauen zu müssen.

Da *ologs* von Grund auf modular sind, können genaue Schnittstellen durch Funktoren spezifiziert werden, die kollaboratives Arbeiten erleichtern. *Ologs* können klar getrennte Abstraktionsebenen enthalten, die tief miteinander verlinkt sind, ebenso können mehrere *ologs* miteinander verbunden werden, um unterschiedliche Facetten des Untersuchungsgegenstands oder Sichtweisen auf ihn zu integrieren.

## Abgrenzung der *ologs* zu anderen Wissensrepräsentationen

*Ologs* sind eng verwandt mit anderen Wissensrepräsentationen wie RDF/OWL, Datenbanken und semantischen Netzen, unterscheiden sich von ihnen jedoch in einigen wichtigen Punkten:

Im Vergleich zu RDF/OWL zeichnen sich *ologs* durch eine höhere Expressivität aus, da in *ologs* Kommutativität ausgedrückt werden kann. Aussagen der Art “Die Schwester meines Vaters ist meine Tante” können also ausgedrückt werden, da die Äquivalenz

der beiden Pfade  $\vdash \text{ich} \rightarrow \vdash \text{Tante}$  und  $\vdash \text{ich} \rightarrow \vdash \text{Vater} \rightarrow \vdash \text{Schwester} \rightarrow$  spezifiziert werden können, sodass  $\vdash \text{Tante}$  und  $\vdash \text{Schwester}$  Label des gleichen Objekts (Entität) sind.

Im Vergleich mit Datenbanken zeichnen sich *ologs* durch höhere Flexibilität aus, sie sind einfacher zu lesen, weniger präskriptiv und lassen sich leichter um neue Informationen erweitern. Dennoch lassen sich *ologs* direkt als Datenbanken implementieren, indem Objekte und Pfeile den Reihen und Spalten von Tabellen zugewiesen werden.

Semantische Netze und *ologs* sind sich sehr ähnlich, unterscheiden sich jedoch hinsichtlich der Robustheit: In semantischen Netzen müssen bei Veränderungen oft viele Links synchronisiert werden, was zu Fehlern führen kann, während bei *ologs* neue Pfade ohne Veränderung des bestehenden kreiert werden können.

## Vortrag

Der Vortrag ist zweigeteilt: Im ersten Teil werden *ologs* mit den zugehörigen Formalismen und Diagrammen eingeführt und das Potential der kategorientheoretischen Ontologieentwicklung und Wissensmodellierung für die Digital Humanities herausgearbeitet. Besonderes Gewicht wird hierbei auf der präzisen Formulierung perspektivisch gebundener Sichtweisen auf Untersuchungsgegenstände, flexible Erweiter- und Kombinierbarkeit von *ologs* und den kollaborativen Möglichkeiten der Modellierung liegen. Darüber hinaus wird die Unterstützung des geisteswissenschaftlichen Denkens durch *ologs* und damit verbundene *rules of good practice* untersucht werden.

Im zweiten Teil werden dann die im ersten Teil gewonnen Einsichten anhand von Beispielen erläutert und gezeigt, wie die Wissensmodellierung durch *ologs* in den Workflow textbasierter Digital Humanities Projekte – wie zum Beispiel CATMA<sup>3</sup> – integriert werden kann.

Das übergreifende Ziel des Vortrags ist es zu zeigen, dass eine mathematisch-theoretische Fundierung der Ontologieentwicklung und Wissensmodellierung sowohl das geisteswissenschaftliche Denken und Arbeiten unterstützen kann, als auch die technische Entwicklung und informatische Implementation treiben kann. Die Kategorientheorie hat in der Mathematik und Informatik sowie in Naturwissenschaften wie Physik, Chemie und Genetik bereits erfolgreich unter Beweis gestellt, ein geeignetes Denkwerkzeug zu sein. Der größte Nutzen in diesen Feldern ist durch Systematisierung und Vergleichbarkeit mit bzw. Anschlussfähigkeit an andere Forschungsgebiete entstanden. Diese Effekte gilt es auch für die Digital Humanities nutzbar zu machen.

## Fußnoten

1. Ein kurzer Blick in “Ontology Development 101: A Guide to Creating Your First Ontology” ([https://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](https://protege.stanford.edu/publications/ontology_development/ontology101.pdf)) vermag das zu bestätigen.
2. Wenngleich Kategorientheorie schon immer auch angewendet wurde, so hat sich in den letzten Jahren eine lebhaftere *scientific community* gebildet (vgl. <https://www.appliedcategorytheory.org> und <https://golem.ph.utexas.edu/category/>), seit 2018 gibt es auch ein OpenAccess Journal (<https://compositionality-journal.org>)
3. CATMA – Computer Assisted Text Markup and Analysis (<https://www.catma.de>)

## Bibliographie

**ACT2018** (2018): “Towards An Integrative Science”, <http://www.appliedcategorytheory.org/workshops/> [letzter Zugriff 15. Juli 2021].

**Abramsky, Samson / Coecke, Bob** (2007): “A categorical semantics of quantum protocols”, <https://arxiv.org/pdf/quant-ph/0402130.pdf> [letzter Zugriff 15. Juli 2021].

**Baez, John C. / Stay, Mike** (2010): “Physics, Topology, Logic and Computation: A Rosetta Stone”, <https://arxiv.org/pdf/0903.0340.pdf> [letzter Zugriff 15. Juli 2021].

**Brown, Ronald / Porter, Timothy** (2008): “Category Theory and Higher Dimensional Algebra: potential descriptive tools in neuroscience”, <https://arxiv.org/pdf/math/0306223.pdf> [letzter Zugriff 15. Juli 2021].

**Coecke, Bob / Sadrzadeh, Mehrnoosh / Clark, Stephen** (2010): “Mathematical Foundations for a Compositional Distributional Model of Meaning”, <https://arxiv.org/pdf/1003.4394.pdf> [letzter Zugriff 15. Juli 2021].

**Ehrig, Hartmut / Mahr, Bernd / Große-Rhode, Martin / Cornelius, Felix / Zeitz, Philip** (2001): *Mathematisch-strukturelle Grundlagen der Informatik*. Berlin: Springer.

**Landry, Elaine M.** (2017): *Categories for the working philosopher*. Oxford: Oxford University Press.

**Mac Lane, Saunders** (1998): *Categories for the working mathematician*. New York: Springer.

**Spivak, David I. / Kent, Robert E.** (2012): “Ologs: A Categorical Framework for Knowledge Representation”, in: *PLoS ONE* 7(1): e24274. <https://doi.org/10.1371/journal.pone.0024274>