

Eine Ontologie der Wissenschaftsdisziplinen. Entwicklung eines Instrumentariums für die Wissenskommunikation

Elena Semenova, Martin Stricker (Hermann von Helmholtz-Zentrum für Kulturtechnik, Humboldt-Universität zu Berlin)

Die drastische Veränderung der Wissenschaftslandschaft ist zu einem markanten Zeichen unserer Zeit geworden. Die Wissenschaftsentwicklung der letzten Jahrzehnte zeichnet sich neben der Etablierung neuer wissenschaftlicher Disziplinen sowie ständiger Erneuerung und Erweiterung der Inhalte von bereits bestehenden Einzelwissenschaften, insbesondere durch eine ausdrückliche Tendenz zur Interdisziplinarität, aus. Die Logik dieser Entwicklung macht Wissenschaftskommunikation zum unentbehrlichen Teil des wissenschaftlichen Lebens. Dementsprechend wächst die Rolle der Kommunikationsinstrumente. Eine gelungene Kommunikation kann nur dann erfolgen, wenn alle Kommunikationspartner eine gemeinsame Sprache beherrschen oder zumindest mit ihren Grundstrukturen vertraut sind. Die Begriffssysteme, die auch als Dokumentationssprachen bezeichnet werden, wie Klassifikationen, Thesauri und zunehmend Ontologien, erhalten dabei neue Funktionen und gewinnen dadurch an Bedeutung. Längst werden diese Begriffssysteme nicht nur für die Erschließung und Recherche der Wissensbestände verwendet. Im Bereich des Datenaustausches, einem wichtigen Teil der Wissenschaftskommunikation, ist die Anwendung systematisch erfasster Begriffssysteme heute unabdingbar geworden. Mit zunehmender Rolle, die Begriffssysteme für den Wissensaustausch spielen, vermehren sich auch die Anforderungen an sie. Die modernen Instrumentarien sollten möglichst adäquat die gewählte Domäne in ihrer Komplexität abbilden, zeitnah, flexibel und gleichzeitig eindeutig, konsistent, skalierbar und effizient in der Benutzung sein.

Die wachsende Zahl der interdisziplinären Vorhaben verlangt besonders nach einer solchen Struktur. Bei der Zusammenarbeit von Wissenschaftlern aus unterschiedlichen Bereichen erschwert die semantische Pluralität die Arbeit enorm. Die interdisziplinäre Kommunikation stößt auf den verschiedensten Ebenen auf zahlreiche Hindernisse, die größtenteils mit Hilfe eines normierten Begriffssystems überwunden werden können. Die Notwendigkeit eines für alle Beteiligten verständlichen und klar strukturierten Vokabulars, welches die Fachsprache

kanalisiert, das Wissensaustausch – vor allem im Rahmen technischer Informationssysteme - ermöglicht, die Kommunikation auf allen Ebenen erleichtert und in einer transparenten und nachvollziehbaren Struktur repräsentiert wird, ist kaum zu überschätzen. Die heute vorhandenen Strukturen entsprechen jedoch diesen Anforderungen nicht: Keines der verfügbaren Vokabularien kann die zunehmende Komplexität der wissenschaftlichen Erkenntnisse und Strukturen zufriedenstellend repräsentieren.

Traditionell wird das System der Wissenschaftsdisziplinen in Form von Klassifikationen dargestellt, d.h. dass als Ordnungssystem eine Baumstruktur zugrunde gelegt wird. Diese Jahrtausende alte Tradition ist jedoch seit geraumer Zeit an ihre Grenzen gestoßen. In der Wissenschaftswelt, wo fast jede Disziplin mit einer anderen verknüpft ist, lassen sich Wissenschaften nur bedingt zu großen Gruppen zusammenfassen und die Hierarchien auf den zweiten, dritten und weiteren Ebenen können oft nur unter zahlreichen Vorbehalten gebildet werden.

Vor einigen Jahren war eine Tendenz zu beobachten, die Nachteile der starren Baumform mit Hilfe der Bildung von polyhierarchischen Vokabularien zu überwinden.

Die Entwicklung von Ontologien als Begriffssysteme war die logische Folge dieser Tendenz, ging aber darüber hinaus und führte zu einem gewissen Paradigmenwechsel in diesem Bereich: Seit der Entwicklung von Ontologien verdrängt die Vorstellung einer Netzstruktur die alten Muster. Die Ontologien entwickeln sich rasch. In der Wirtschaft längst Norm geworden, gewinnen sie seit einiger Zeit sichere Positionen im Wissenschaftsalltag. In vielen Teilbereichen der Wissenschaft werden fachspezifische Ontologien entwickelt. Wenn jedoch die Struktur der Wissenschaftslandschaft selbst repräsentiert werden soll, wird nach wie vor zumeist auf die konventionelle Systematik in Form der Klassifikation zurückgegriffen. Die Gründe dafür sind ohne Zweifel in der Geschichte zu suchen. Durch die historische Entwicklung hat es sich so ergeben, dass die Bibliotheken im Bereich der informationswissenschaftlichen Technologien eine Vorreiterrolle übernommen haben. Dementsprechend wurden alle bekannte Systematiken in erster Linie für die Erschließung und Recherche der Bibliotheksbestände erstellt. Als Melvil Dewey in den 70er Jahren des 19. Jahrhunderts seine Dezimalklassifikation entwickelte, entsprach diese einerseits den Anforderungen, eine ortsgebundene,

klare und einheitliche Ordnung in die Bibliothek zu bringen, andererseits ist in ihr die zeittypische Ordnung der Wissenschaftsdisziplinen ablesbar.

Durch die Wandlung der Inhalte einzelner Wissenschaftsdisziplinen sowie durch die drastische Veränderung der gesamten Wissenschaftslandschaft erscheinen heute - mehr als hundert Jahre nach Deweys Entwicklung – trotz beständiger Überarbeitung einige Positionen der DDC hinsichtlich der Wissenschaft nicht zeitgemäß. Außerdem ist das Ziel der DDC nicht, explizit wissenschaftliche Disziplinen zu erfassen, sondern eher Wissens- und Wirkungsbereiche. Schon in der ersten Übersichtsebene der letzten Fassung der DDC¹ wird dieses deutlich:

000 Informatik, Informationswissenschaft, allgemeine Werke
100 Philosophie und Psychologie
200 Religion
300 Sozialwissenschaften
400 Sprache
500 Naturwissenschaften und Mathematik
600 Technik, Medizin, angewandte Wissenschaften
700 Künste und Unterhaltung
800 Literatur

Wissenschaftliche Disziplinen (*Informationswissenschaft, Philosophie und Psychologie, Religion* u.a.) befinden sich auf einer Ebene mit Phänomenen (*Sprache*) und Erzeugnissen der menschlichen Kreativität (*Künste und Unterhaltung, Literatur*). Von der zweiten Ebene an wird diese Art der uneinheitlichen Klassenbildung in Hinsicht der Wissenschaftsdisziplinen noch deutlicher: In der Klasse *Sprache* steht *Linguistik* (eine Disziplin) neben der Auflistung verschiedener Sprachen (einem Phänomen) wie z.B. *Englisch, Altenglisch* oder *Deutsch, germanische Sprachen allgemein* etc. Die hierarchische Gliederung in der DDC wird manchmal sogar regelrecht verunstaltet: in der Klasse *Naturwissenschaften* stehen *Biowissenschaften/Biologie* auf der gleichen Ebene wie *Pflanzen (Botanik)* und *Tiere (Zoologie)*. Ein solches Ordnungssystem orientiert sich an der Besonderheit des Bibliothekssystems und ist den Erfordernissen der Erschließung von Bibliotheksbeständen angepasst. Dieses ist jedoch nur bedingt auf andere Bereiche übertragbar. Nach dem gleichen Prinzip wurden die Internationale Universale Dezimalklassifikation (UDK) und die Regensburger Verbundklassifikation (RVK)² aufgebaut. Obwohl beide Klassifikationen in der Auswahl der Bezeichnungen viel konsequenter vorgehen als die DDC, weisen aus Sicht der wissenschaftlichen

¹ <http://www.ddc-deutsch.de/produkte/uebersichten/summaries1.htm> [02.08.2007]

² http://www.bibliothek.uni-regensburg.de/rvko_neu/ [02.08.2007]

Fachgebiete diese Klassifikationssysteme einen Mangel an Homogenität der Beschreibung, an ausgewogenen Strukturen und an präziser Abbildung der Relationen zwischen den Disziplinen auf.³ Mit anderen Worten: Diese Klassifikationssysteme sind streng zweckbezogen, vertreten die Sichtweise einer bestimmten Zielgruppe und erfüllen auf keine Weise den Anspruch auf Universalität.

Die primäre Position, die Bibliotheken im Informationsbereich innehatten, wird zunehmend durch die Verbreitung des Internet in Frage gestellt. Die Verbreitung der neuen Technologien und das ständig anwachsende Wissensgut führt zu einem neuen Selbstverständnis der Gesellschaft, die heute als Wissensgesellschaft bezeichnet wird. Wissensportale, die der Bündelung der Informations- und Wissensströme dienen, besetzen heute in der Wissenschaftskommunikation eine wichtige Stellung. Virtuelle Bibliotheken, Subject Trees bzw. Gateways und Suchmaschinen für Wissenschaftsserver bieten neben den Suchmöglichkeiten auch eine Browsing-Funktion, die hauptsächlich auf einer Wissenschaftssystematik basiert. Die überwiegende Zahl dieser Portale sind Fachportale, die jeweils nur ein Fachgebiet präsentieren und dementsprechend eine fachspezifische Systematik verwenden.⁴ Viele dieser Portale entwickeln eine eigene Systematik. Bei aller Differenz haben sie ein wesentliches Merkmal gemeinsam: Die Bezeichnungen der Wissenschaftsdisziplinen werden hier in erster Linie für den thematischen Einstieg in die Recherche genutzt. Die Systematiken gehen also nicht in die Tiefe, sondern sind relativ grob gegliedert und bilden die Fachgebiete nicht in vollem Maße ab. Die Fächerauswahl von ‚vascoda‘ z. B. bietet nur 30 Fächer an, die in 4 großen Klassen zusammengefasst sind. Obwohl die Suchfunktion von ‚vascoda‘ Informationen für eine erheblich größere Anzahl der Fächer recherchieren lässt, mangelt es an Navigationsmöglichkeiten. Aus den ausgegebenen Treffern kann man nur bedingt die nötigen Zusammenhänge herauslesen. Als Folge dessen kann diese Systematik nur beschränkt für eine erfolgreiche wissenschaftliche Kommunikation verwendet werden. Offensichtlich wurde bei der Erstellung der Systematik ein Bottom-Up

³ Vgl. Konrad Umlauf, Einführung in die bibliothekarische Klassifikationstheorie und –praxis: <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h67/> (02.08.07)

⁴ Vgl. Clio-online, Ein Fachportal für die deutschsprachigen historischen Wissenschaften; EVIFA, die Virtuelle Fachbibliothek Ethnologie; Anglistik Guide: Virtual Library of Anglo-American Culture - Literature at SUB Göttingen (GoeAng); DAINet - German Agricultural information Network, Germany (DAINet); Geo-Guide: Virtual Library of Earth Sciences, Geography, Thematic Maps, and Mining at SUB Göttingen (GoeGeo); History Guide: Virtual Library of Anglo-American Culture - History at SUB Göttingen (GoeHist); MathGuide: Part of the Virtual Library of Mathematics at SUB Göttingen (GoeMath) u.a.

Prinzip gewählt. Die Vorteile dieses Prinzips bestehen darin, dass das Erschließungs- und Rechercheinstrument in Umfang und fachspezifischem Standpunkt an die zu erschließende Datensammlung angepasst wird. Über die Datensammlung hinaus verliert ein solches Instrument seine Relevanz.

Gleiches gilt für andere Portale. Die vorgenommene Systematisierung erfüllt durchaus das eigentliche Ziel des einzelnen Portals, ist dennoch für andere Datensammlungen nicht zu verwenden. Als Folge dessen stellen sie keineswegs ein komplettes Bild der Wissenschaftsdisziplinen dar und können nur begrenzt in anderen Kontexten genutzt werden.

Für die Erstellung einer neuen Struktur sprechen viele Gründe, die im folgenden zu skizzieren sind:

- Als erste sind quantitativen Gründe zu nennen: Es gibt wesentlich mehr Fachbereiche, als in den oben genannten Systematiken aufgelistet sind. Allein nach Angaben des Statistischen Bundesamtes⁵ werden an deutschen Hochschulen 619 Fächergruppen, Lehr- und Forschungsbereiche sowie Fachgebiete gezählt, wobei einige Disziplinen in dieser Aufstellung nicht ausdifferenziert werden. So wird z.B. die Sprachwissenschaft nur im Hinblick auf verschiedene Sprachen wie *Anglistik*, *Niederlandistik*, *Skandinavistik*, *Slavistik* präsentiert und nicht in Teilgebiete wie *Phonetik/Phonologie*, *Morphologie*, *Syntax/Grammatik*, *Semantik/Lexikographie* gegliedert. Unter Berücksichtigung der fehlenden Gliederung könnte die Anzahl der Fachbereiche etwa auf eintausend geschätzt werden. Um die Wissenschaftsdisziplinen adäquat darstellen zu können, ist die Entwicklung eines möglichst vollständigen Vokabulars unabdingbar.

- Die qualitativen Gründe, die für die Entwicklung einer neuen Beschreibungssystematik sprechen, lassen sich in zwei Bereiche untergliedern: Einerseits gilt es Probleme zu bearbeiten, die für jedes kontrollierte Vokabular typisch sind, andererseits gilt es Ansätze zu entwickeln, die den spezifischen Anforderungen bei der Beschreibung von Wissenschaftsdisziplinen genügen:

- Synonymie, Polysemie und Homonymie. Die für jedes kontrollierte Vokabular typischen Probleme sind auch für den Bereich der wissenschaftlichen Disziplinen relevant. Das heißt: Alle Bezeichnungen der Fachgebiete müssen berücksichtigt

⁵<http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Klassifikationen/Bildung/Kultur/StudentenPruefungsstatistik,property=file.pdf> (02.08.07)

werden. So sollte im Vokabular nicht nur der Begriff *Phylogenie*, sondern auch *Phylogenese*, neben *Stammesentwicklung* auch *Stammesgeschichte* enthalten sein und damit das Synonymproblem bewältigt werden.

– Das Polysemieproblem wird in vielen vorhandenen Systematiken mit Hilfe von polyhierarchischen Strukturen gelöst, was durchaus berechtigt wäre, wenn alle relevanten Bedeutungen vollständig abgebildet würden. Allerdings findet man in keiner der Systematiken eine komplette Darstellung beispielsweise des klassischen Polysems *Morphologie*. Die *Morphologie* ist ein Teilgebiet, das nicht nur in der *Sprachwissenschaft* und der *Biologie* eine Rolle spielt, sondern auch in der *Geomorphologie* oder in der *Stadtmorphologie*.

– Ein grundlegendes Problem der Präsentation von Wissenschaftsdisziplinen ist die Vielfältigkeit der Forschungsaspekte. Die *Anthropologie* beispielsweise unterscheidet *Philosophische Anthropologie*, *Theologische Anthropologie*, *Biologische Anthropologie*, *Forensische Anthropologie*, *Historische Anthropologie*, *Kultur- und Sozialanthropologie*, *Kybernetische Anthropologie* und *Visuelle Anthropologie*. Der Forschungsgegenstand ist jedes mal der Gleiche - der Mensch - aber je nach Forschungsbereich wird er unter verschiedenen Aspekten betrachtet. Bis jetzt gibt es keine Systematik, die diese Relationen anschaulich und umfassend abbildet.

– Manche Grundwissenschaften sind Bestandteil unterschiedlicher Disziplinen. Ein Beispiel ist die *Papyrologie*, die sowohl für die *Altphilologie* als auch für die *Archäologie* relevant ist. Auch diese Relationen müssen zum Ausdruck gebracht werden.

– Viele moderne Fachgebiete haben sich aus unterschiedlichen Disziplinen heraus entwickelt. So ist etwa die *Computerlinguistik* ein Teilbereich der *Künstlichen Intelligenz* und gleichzeitig Schnittstelle zwischen *Sprachwissenschaft*, *Kognitionswissenschaft* und *Informatik*. Solche Schnittstellen müssen genau dokumentiert werden.

– Es gibt kaum Wissenschaftsdisziplinen, die keine Kooperationen mit anderen Fachgebieten eingehen. In der *Telekommunikation* beispielsweise setzen sich die Forschungsteams aus den Bereichen *Ingenieurwissenschaften*, *Nachrichten- und Elektrotechnik*, *Mathematik* und *Physik* sowie *Informatik* zusammen. Heutzutage wird zwar oft über die Vernetzung der Wissenschaft gesprochen, es fehlt jedoch

eine entsprechende Abbildung solcher Netze. Mit dem traditionellen Instrumentarium der Hierarchie- und Äquivalenzbeziehungen lässt sich dieses Problem nicht lösen.

– Auch die Beziehungen zwischen Studiengängen und Wissenschaftsdisziplinen sind zu berücksichtigen. Der Studiengang *Religionswissenschaft* z.B. beinhaltet u.a. auch das Studium der *Philosophie*, *Soziologie* und *Psychologie*. Eine Systematik der universitären Disziplinen hat auch solche Beziehungen repräsentativ darzustellen.

Diese Aufgaben können nur mit Hilfe eines Begriffssystems bewältigt werden, das die Möglichkeiten der traditionellen Erschließungs- und Rechercheinstrumente – Klassifikationen und Thesauri – erweitert und Relationen zwischen Konzepten, sowie pragmatische Aspekte wie unterschiedliche Perspektive in die Struktur mit einbezieht. Ein solches Instrumentarium sollte nicht nur möglichst unmissverständlich, sondern auch anschaulich und intuitiv nachvollziehbar sein, um größere Benutzerkreise mit der Struktur der Domäne vertraut zu machen und mögliche Schnittstellen für potenzielle Kooperationspartner zu verdeutlichen. Die aktuelle Entwicklung im Bereich Ontologieaufbau bildet das Fundament für die Erarbeitung eines solchen Instrumentariums. Die Ontologien erlauben, unterschiedlichste Formen von Beziehungen abzubilden und die Wissenschaftslandschaft, die sich selbst als eine vernetzte Welt versteht, in einer geeigneten Form zu repräsentieren.

Die Erarbeitung der Ontologie der wissenschaftlichen Disziplinen wird, wie bei solchen Projekten üblich, in zwei Phasen ablaufen: Modellerstellung mit Bildung von Klassen für Konzepte, Relationen, Namen und Rollen sowie anschließend die Erstellung der Ontologie selbst. Zugleich erfolgen parallel die Entwicklung spezifischer Software und Software-Schnittstellen für die Implementierung.

Unter einer „Ontologie“ verstehen wir für unser Projekt ein verzweigtes, verflochtenes Begriffssystem, das die Elemente eines kontrollierten Vokabulars in sich einschließt und das auf einem standardisierten Referenzmodell für einen ausgewählten Bereich⁶ (in diesem Fall für die wissenschaftlichen Disziplinen) beruht. Dabei sollen die

⁶ Vergl. z.B. Wolfgang Hesse, Barbara Krzensk: Ontologien in der Softwaretechnik <http://www.mathematik.uni-marburg.de/~krzensk/mod04/WorkshopBand.pdf> (31.07.07). Definition der Ontologie von Wolfgang Hesse in Glossar der Gesellschaft für Informatik; Diskussion: Katrin Weller : „Kooperativer Ontologieaufbau“ http://www.walt.phil-fak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/35/1166536740katrinwell.pdf (31.07.07)

Vorteile der Potentiale der maschinellen Begriffsermittlung (*Reasoning*) genutzt werden, es ist jedoch davon auszugehen, dass der Anteil intellektueller Arbeit sehr hoch sein wird.

Für die Bestimmung der Anwendungsdomäne wird ein Verständnis des Begriffs „Wissenschaft“ als differenzierter und institutionalisierter Teilbereich der Gesellschaft zugrunde gelegt. Zusätzlich beschränkt sich derzeit das Projekt auf Forschung und Lehre in Deutschland (beim Referenzmodell soll auf eine weitgehende internationale Anwendbarkeit geachtet werden). Für eine kohärente Definition des Gegenstandsbereiches sollen Wissenschaftsdisziplinen von Wissensfeldern, Wissensgebieten und Wissensbereichen deutlich abgegrenzt werden. Die Anerkennung eines Wissensgebietes als wissenschaftliche Disziplin ist mitunter sehr umstritten. Da unser Projekt keinen normativ-setzenden, sondern einen deskriptiven Charakter hat, ist hier eine praktische Lösung gefragt, die durch einen Rückgriff auf die Verfahren gesellschaftlicher Legitimierung (zum Beispiel Institutionalisierung an Universitäten) Konsens zu erzielen sucht.

Die Festlegung der Vorgehensweise für die Erstellung der Ontologie wurde aus diesen zwei Grundverständnissen abgeleitet. Für unser Begriffssystem bevorzugen wir als Arbeitsweise einen Top-Down Ansatz. Dieses Verfahren hat eine Reihe von spezifischen Merkmalen, die bei der Erstellung dieser Art von Ontologien von Vorteil sind: Unter anderem ermöglicht die Unabhängigkeit von einem festgelegten Datenbestand eine breite Perspektive; auf diese Weise können unterschiedliche Auffassungen nebeneinander zum Ausdruck gebracht werden.

Die projektierte Ontologie wird mit dem standardisierten Datenmodell des *Semantic Web* (OWL/RDF) kompatibel sein bzw. auf einem Subset davon beruhen. Eine Kompatibilität mit weiteren Ontologien oder entsprechenden Anwendungen ist somit gewährleistet. In der derzeitigen Anfangsphase der Modellierung, die, da von standardisierten Verfahren bei der Ontologierstellung noch nicht die Rede sein kann, sehr stark experimentell und iterativ geprägt ist, kommt der ISO-Standard *Topic Maps* zum Einsatz. Dessen Paradigma sowie entsprechende Softwareinstrumente ermöglichen aufgrund des – im Vergleich zu OWL/RDF – Reichtums an konkreten Konzepten einen flexiblen Entwicklungsprozess. Beabsichtigt ist, Referenzmodell und Ontologie in beiden Standards zu publizieren.

Zu den weiteren Projektzielen zählt die Entwicklung und Bereitstellung einer Web-basierten Anwendung, die hypertextualisiert und grafisch-visuell zur Recherche der Ontologie sowie der im Laufe des Projektes erarbeiteten damit verknüpften Informationsquellen genutzt werden kann. Ein dritter Weg, auf die Ontologie zuzugreifen oder sie in kollaborativen Verfahren zu bearbeiten oder zu ergänzen (von Interesse vor allem für Kooperationen und die Integration in verteilte Grid-Systeme), werden (voraussichtlich Web-basierte) API-Schnittstellen sein, die gegebenenfalls gemeinsam mit konkreten Kooperationspartnern entwickelt werden sollen.

Kontaktinfo

Dipl. Phil., Wiss. Dok. Elena Semenova elena.semenova@culture.hu-berlin.de

Martin Stricker martin.stricker@rz.hu-berlin.de

Humboldt-Universität zu Berlin Hermann von Helmholtz-Zentrum für Kulturtechnik
Unter den Linden 6

Tel. 030/2093-2522 oder –2563

Telefax 030/2093-2727 oder -1961

10099 Berlin

<http://www2.hu-berlin.de/hzk/>