

Ludger Jansen^{*,†}

Chinesische Tiere und biomedizinische Datenbanken. Logische und technische Bedingungen wissenschaftlicher Klassifikationen

In unserer Zeit wird soviel neues Wissen produziert wie nie zuvor: Jährlich erscheinen etwa 10 Millionen neue wissenschaftlich-technische Publikationen. Konnte ein Arzt vor hundert Jahren noch durch eigene Lektüre mit dem wissenschaftlichen Fortschritt der Medizin Schritt halten, würde er heute mehrere Wochen benötigen, um eine Tagesproduktion an wissenschaftlichen Publikationen seiner Profession zu lesen (Gaus 2003, 11-12). Doch wie findet ein Arzt in diesem Berg von Veröffentlichungen die für ihn relevanten Informationen? Er wird versuchen, sie über computergestützte online verfügbare Literaturdatenbanken zu suchen. Ganze Wissenschaftszweige wie die Genetik sind ohne Computerunterstützung nicht mehr denkbar, so groß ist die von ihnen angesammelte Datenmenge. In unserer Zeit ist damit neben die Entdeckung neuen Wissens eine andere Herausforderung getreten: Wie können wir unser wissenschaftlich-technisches Wissen so organisieren und darstellen, dass wir es auch wiederfinden, wenn wir es benötigen? Medizininformatiker begegnen dieser Herausforderung, indem sie versuchen, das Wissen über biomedizinische Entitäten in sogenannten Ontologien systematisch darzustellen. Elementar dafür ist das klassische Instrument der Klassifikation oder Taxonomie. Dinge zu klassifizieren ist seit alters her eine Standardtätigkeit für Wissenschaftler, und so verwundert es nicht, dass sich Philosophen seit zweieinhalb Jahrtausenden Gedanken über das Klassifizieren gemacht haben, angefangen bei Platon und Aristoteles bis hin zur modernen Wissenschaftstheorie.

Im folgenden werde ich zunächst die Parodie einer Klassifikation diskutieren: die „chinesische“ Klassifikation der Tiere von Jorge Luis Borges.

* Institut für Philosophie, Universität Rostock.

† Dieser Beitrag wurde geschrieben mit Hilfe der Unterstützung des Wolfgang-Paul-Programms der Alexander-von-Humboldt-Stiftung, des European Union Network of Excellence on Medical Informatics and Semantic Data Mining und der Volkswagen Stiftung im Rahmen des Projektes „Formen des Lebens“. Dank an Barry Smith für hilfreiche Kommentare zu früheren Fassungen und an mein Publikum in Saarbrücken und Berlin für die Gelegenheit, anlässlich zahlreicher Fragen und Diskussionbeiträge vieles noch deutlicher herauszuarbeiten.

Das, was uns an dieser Klassifikation komisch erscheint, zählt natürlich zu dem, was in einer idealen Klassifikation vermieden werden sollte (Abschnitt 1). Ich werde dann aber zeigen, dass vieles von dem, was an dieser Parodie komisch erscheint, auch in dezidiert wissenschaftlichen Klassifikationen vorkommt. Dazu werde ich Beispiele aus der Terminologie-Datenbank des amerikanischen *National Cancer Institute* (NCI) heranziehen, dem *NCI Thesaurus* (Abschnitt 2). Zum Teil sind diese Abweichungen vom klassifikatorischen Ideal abhängig von den technischen Randbedingungen einer solchen Datenbank (Abschnitt 3). Eine Arbeitsteilung zwischen mehreren Informationssystemen – zwischen einer auf Vollständigkeit angelegten Referenzklassifikation oder Referenzontologie auf der einen und auf bestimmte Anwendungen zugeschnittenen Anwendungsklassifikationen oder Anwendungsentologien auf der anderen Seite – könnte einige der Probleme lösen (Abschnitt 4).

1 Chinesische Tiere: Wie man eine gute Klassifikation erstellt

In einer „gewissen chinesischen Enzyklopädie“, so erzählt Jorge Luis Borges, seien die Tiere in die folgenden Gruppen eingeteilt:

1. Tiere, die dem Kaiser gehören,
2. einbalsamierte Tiere,
3. gezähmte Tiere,
4. Milchschweine,
5. Sirenen,
6. Fabeltiere,
7. herrenlose Hunde,
8. in diese Gruppierung gehörige,
9. die sich wie Tolle gebärden,
10. unzählbare,
11. die mit einem ganz feinen Pinsel aus Kamelhaar gezeichnet sind,
12. und so weiter,
13. die den Wasserkrug zerbrochen haben,
14. die von weitem wie Fliegen aussehen. (Borges 1966, 212)

Wir haben es hier mit einer ausgezeichneten Parodie von Klassifikationen zu tun.¹ Um mich leichter auf sie beziehen zu können, will ich diese erfundene Klassifikation im folgenden als „CTT“, als „Chinesische Tier-Taxonomie“ bezeichnen. CTT wirkt dadurch komisch, dass sie einige Sachen grundlegend anders macht, als man es in einer wissenschaftlichen Klassifikation machen sollte. Gute Klassifikationen beruhen also im Gegenschluß darauf, genau diese Fehler zu vermeiden. Was ist es also, das uns in CTT lustig vorkommt? Was können wir von CTT lernen? Hier sind einige der Regeln für das Anfertigen von guten und nützlichen Klassifikationen, gegen die CTT verstößt:

- *Strukturiertheit.* Gute Taxonomien sind strukturiert. Sie berücksichtigen, dass Typen sich in Untertypen aufteilen. In CTT sind alle Gruppen gleichberechtigt. Nun sind Sirenen aber doch wohl Fabeltiere, (5) mithin eine Unterart von (6). Gefragt ist also eine Hierarchie von Gattungen und Arten.
- *Disjunkтивität.* Wenn wir eine solche Hierarchie von Typen und Subtypen haben, dann gilt: Alles, was unter einen Subtyp fällt, fällt auch unter den dazugehörenden Typ. In einer zoologischen Klassifikationen (die CTT ja wäre) gilt beispielsweise, daß alle Tiere, die Säugetiere sind, auch Wirbeltiere sind. Auf derselben Ebene der Klassifikation sollten die verschiedenen Typen jedoch disjunkt sein: Kein Ding sollte zu zwei Typen derselben Ebene gehören, wie ja auch kein Tier sowohl ein Säugetier und ein Reptil ist oder sowohl ein Wirbeltier und ein Wirbelloses. Wenn Typen derselben Ebene diese Forderung erfüllen, sind sie disjunkt. Die Typen in CTT erfüllen diese Forderung nicht (zumindest nicht notwendigerweise): Typ (1) der Tiere, die dem Kaiser gehören, kann ja auch gezähmte Tiere umfassen, also Tiere, die zu Typ (3) gehören.
- *Exhaustivität.* Taxonomien enthalten idealerweise alle Entitäten desjenigen Bereiches, den sie zu klassifizieren vorgeben. CTT ist jedoch weit entfernt davon, alle Tiere zu umfassen, wenn wir davon absehen, dass wir beliebige Tiere unter Typ (8) oder auch unter den Typ (13) „und so weiter“ subsumieren können. Exhaustivität und Disjunkтивität werden oft mit der Formulierung „paarweise disjunkt und gemeinsam exhaustiv“ in einem Atemzug genannt; im Englischen wird

¹ Als Quelle für die Taxonomie gibt Borges „Dr. Franz Kuhn“ an. Es gibt zwar einen deutschen Sinologen und Übersetzer zahlreicher chinesischer Texte dieses Namens, aber unter seinen Veröffentlichungen ist keine Quelle für CTT zu finden. Vermutlich hat Borges hier in parodistischer Absicht eine falsche Fährte gelegt. Vgl. <http://www.linguistlist.org/issues/7/7-1446.html>. Meine Zweifel an der Seriosität von Borges' Verweis auf Kuhn und den Hinweis auf diese Seite verdanke ich Jan Westerhoff.

diese Forderung gerne mit „JEPD“ abgekürzt (für „jointly exhaustive, pairwise disjunct“).

- *Verzicht auf Ambiguität.* Gute Taxonomien spielen nicht mit Mehrdeutigkeiten ihres Oberbegriffs. Sind Fabeltiere, gemalte Tiere und tote Tiere im gleichen Sinn Tiere wie herrenlose Hunde? Wohl nicht. Die Kategorien (2), (5), (6) und (11) passen aus diesem Grund nicht in das Ordnungsschema hinein. Gemalte Tiere sind eben keine Tiere, sondern Bilder, die Tiere darstellen.
- *Einheitlichkeit.* Gute Taxonomien haben einen wohldefinierten Bereich, aus dem sie ihre Unterscheidungsmerkmale wählen. CTT hingegen verwendet auf höchst kreative Weise Ordnungsmerkmale aus möglichst verschiedenen Bereichen: (1) sortiert nach dem Eigentümer, (4) nach der Artzugehörigkeit, (7) nach der Abwesenheit eines Eigentümers plus Artzugehörigkeit, (9) nach dem Verhalten, (13) nach den Auswirkungen des Verhaltens und (14) nach der Wirkung auf einen entfernten Beobachter.
- *Keine Meta-Kategorien.* Gute Taxonomien vermeiden „Meta-Kategorien“, die nur durch die Kategorisierung selbst zustande kommen. Denn solche Meta-Kategorien können zu paradoxen Effekten führen. In CTT ist (8) eine solche Meta-Kategorie: „Tiere, die zu CTT gehören“. Gehört ein Tier zu CTT? Dann gehört es zu (8). Gehören alle Tiere zu CTT? Dann gehören alle Tiere zu (8). Alle Tiere, die zu (1) bis (7) oder (9) bis (14) gehören, gehören also zu (8). Gehört ein Tier zu CTT, aber nicht zu diesen Klassen, ist dies auch kein Problem: Es kann ja auch zu CTT gehören, wenn es nur zu (8) gehört. (8) ist also eine sehr merkwürdige Typ Tier. Klassifikationen, die (8) enthalten, führen zu Problemen, die strukturell dem semantischen Paradox des sogenannten Wahr-Sagers „Dieser Satz ist wahr“ entsprechen: Dieser Satz ist hinsichtlich seines Wahrheitswertes unbestimmt, weil er beliebig bewertet werden kann. Wenn wir annehmen, er sei wahr, dann ist der Fall, was er behauptet, nämlich dass er wahr ist – und das ist, was für einen wahren Satz verlangt ist. Nehmen wir hingegen an, er sei falsch, dann ist nicht der Fall, was er behauptet, wie dies für einen falschen Satz auch angemessen ist. Jeder der beiden Wahrheitswerte „wahr“ und „falsch“ kann dem Wahr-Sager also konsistent zugeordnet werden. Genauso konsistent können Tiere, die nicht zu den übrigen Typen von CTT gehören, (8) zugeordnet werden oder nicht. Eine solche Möglichkeit zur Willkür ist freilich sehr merkwürdig.

Noch schlimmer wird es mit einer Russellschen Variante von CTT, CTT*, die statt des Typs (8) den Typ (8*) enthält:

8*. Tiere, die nicht zu CTT* gehören.

Ein solcher Typ führt zu Problemen, die der Russellschen Antinomie oder dem Lügner-Paradox strukturell entsprechen: Gehört ein Tier zu den Typen (1) bis (7) oder (9) bis (14), dann gehört es zu CTT* und damit nicht zu (8*). So weit ist die Sache klar. Gehört ein Tier aber nicht zu diesen Typen, geraten wir in eine paradoxe Situation. Denn wenn das Tier nun auch nicht zu (8*) gehören würde, würde es zu gar keinem Typ von CTT* gehören, und damit nicht zu CTT*. Tiere, die nicht zu CTT* gehören, gehören aber zu (8*). Wenn wir davon ausgehen, dass das Tier nicht zu den übrigen Typen gehört, folgt also: Wenn etwas nicht zu (8*) gehört, gehört es zu (8*). Aber was auch immer zu (8*) gehört, gehört zu CTT*. Dann jedoch gehört das Tier *nicht* zu (8*). Wieder vorausgesetzt, dass das Tier nicht zu den übrigen Typen gehört, gilt also auch: Was zu (8*) gehört, gehört nicht zu (8*). Das Auftreten solcher Situationen sollte in wissenschaftlichen Klassifikationen tunlichst vermieden werden.

- *Expliztheit und Präzision.* Gute Taxonomien sind explizit und präzise. „Und so weiter“-Kategorien wie (12) genügen diesen Ansprüchen offensichtlich nicht.
- *Ontologische Fundierung.* Gute Taxonomien klassifizieren Dinge aufgrund von diesen zukommenden Merkmalen. Das schließt Meta-Kategorien genauso aus wie Typ (12), „und so weiter“. Denn zur „und so weiter“-Gruppe gehören Dinge nicht, weil sie ein sie einigendes Merkmal besitzen. Und auch Typ (14), „Die von weitem aussehen wie Fliegen“, klassifiziert nicht direkt aufgrund von Merkmalen der Dinge selbst, sondern aufgrund ihrer Wirkung auf einen entfernten Beobachter.

2 Medizindatenbanken: Wie man eine schlechte Klassifikation macht

Bisher diente uns CTT als heuristisches Mittel, einige der systematischen Fehler zu benennen, die man beim Erstellen einer Klassifikation machen kann. Diese Fehler treten nicht nur in der parodistischen Literatur auf, sondern auch in der wirklichen Wissenschaft. Dies will ich im folgenden anhand des *National Cancer Institute Thesaurus* (NCIT) zeigen.² Dies gibt mir die Gelegenheit, ausführlicher auf die bisher nur angerissenen

² Der NCIT ist frei zugänglich unter:
<http://ncit.nih.gov/NCIBrowser/Dictionary.do>. Alle Verweise auf den NCIT in diesem Kapitel berufen sich auf die Version 05.11f vom November 2005.

Probleme des Klassifizierens einzugehen und Lösungsmöglichkeiten zu diskutieren.

Der NCIT wurde vom amerikanischen *National Cancer Institute* erstellt, um seinen Kampf gegen den Krebs dadurch zu unterstützen, dass der Thesaurus sowohl ein online zugängliches Nachschlagewerk darstellt, als auch ein kontrolliertes Vokabular zur Annotation und Indexierung der für die Erforschung und Heilung des Krebs einschlägigen Literatur bietet (Ceusters, Smith & Goldberg 2005). Er enthält dazu mehr als 110.000 Ausdrücke und 36.000 Begriffe, die für Krebs und die Krebsforschung wichtig sind. Darunter sind 10.000 Arten medizinischer Diagnosen oder Störungen, mehr als 5.000 anatomische Arten, mehr als 3.500 Chemikalien und Medikamente, etwa 2.000 Arten von Genen etc. (Fragoso et al. 2004).

2.1 Strukturiertheit: Gruppen und Tiere

Anders als CTT, besitzt der NCIT eine Hierarchie von Ober- und Untertypen. Doch ist diese Hierarchie oft nicht oder nicht richtig ausgearbeitet. Ein Beispiel für eine fehlende Hierarchisierung bietet der Eintrag „Subgroup“, der vom NCIT als „A subdivision of a larger group with members often exhibiting similar characteristics“ definiert wird. Nun sollte man annehmen, dass solche Teile von Gruppen auch Gruppen sind, und in der Tat wird dies auch von der Definition von „Group“ im NCIT impliziert: „Any number of entities (members) considered as a unit.“ Aber die Verbindung zwischen „Subgroup“ und „Group“ fehlt im NCIT. Hier fehlt also ein Teil der hierarchischen Struktur. Aber der NCIT ist in vielen Fällen auch falsch strukturiert. Als Obertyp von „Group“ wird beispielsweise „Grouping“ angegeben, was definiert wird als „A system for classifying things into groups or the activity of putting things together in groups“. Aber, wie die philosophische Tradition weiß,³ muß die Definition des Supertyps auch von den Subtypen ausgesagt werden können. Aus der Definition von „Grouping“ und der Tatsache, dass „Group“ als Untertyp von „Grouping“ angesehen wird, erhalten wir jedoch die folgende absurde Konklusion:

A subgroup is a system for classifying things into groups or the activity of putting things together in groups.

Nahezu von ähnlicher Qualität wie in Borges' CTT ist die Klassifikation der Tiere im *NCI Thesaurus*. Der Typ *Tier* hat dort die Untertypen *Wirbelloses*, *Labortier*, *Wechselwarmes*, *Wirbeltier*. Eine höchst merkwürdige

³ Vgl. z.B. Aristoteles, Kategorien 3.

Reihung fürwahr, ist doch erstens schon das Paar *Wirbelloses* / *Wirbeltier* eine vollständige Aufteilung aller Tiere. Zweitens nimmt sich der Typ *Labortier* neben drei natürlichen Klassen höchst seltsam aus: Labortiere bilden eben keine natürliche Art; die Einteilung verwendet hier also Merkmale ganz verschiedenen Typs. Und drittens ist *Wechselwarmes* ein Untertyp von *Wirbeltier*, gehört also gar nicht auf die gleiche Ebene wie sein Obertyp.

2.2 Disjunkтивität und Exhaustivität: Patienten

Häufig gibt es im NCIT Untertypen desselben Obertyps, die nicht disjunkt sind. Ein Beispiel: Als medizinische Terminologie hat der NCIT natürlich einen Eintrag „Patient“. Dieser Eintrag hat zwei Untertypen: „Krebspatient“ und „Ambulanter Patient“ („Cancerpatient“, „Outpatient“). Diese beiden Einträge sind augenscheinlich nicht disjunkt: Viele Krebspatienten werden ambulant behandelt. Und natürlich sind diese beiden Untertypen auch keine exhaustive Klassifikation der Patienten. Denn es gibt zahlreiche Patienten, die weder Krebspatienten noch ambulante Patienten sind.

Normalerweise würde man dieses Beispiel als einen typischen Anwendungsfall einer Kreuzklassifikation ansehen. Wir haben es mit zwei voneinander unabhängigen Merkmalen zu tun, die jeweils vorliegen oder nicht vorliegen können. Miteinander kombiniert, bilden sie vier Klassen von Patienten:

Abb. 1: Vier Klassen von Patienten – eine Kreuzklassifikation

Patienten	Ambulant? Ja.	Ambulant? Nein.
Krebs? Ja.	Ambulante Krebspatienten	Stationäre Krebspatienten
Krebs? Nein.	Ambulante Nichtkrebspatienten	Stationäre Nichtkrebspatienten

In den Informationswissenschaften stellt man sich ideale Klassifikationen als Baumstrukturen vor. Das steht in der langen Tradition der *Arbor Porphyriana*, des Porphyrischen Baumes. Mit Hilfe solcher Bäume können gemäß der Formel *Definitio fit per genus proximum et differentiam specificam* leicht Definitionen nach Aristotelischem Muster erstellt werden: Eine Spezies wird definiert, indem man die ihr als nächstes übergeordnete Gattung (das *Genus proximum*) gemeinsam mit dem sie konstituierenden spezifischen Merkmal (der *Differentia specifica*) angibt.

Solche Baumstrukturen nennt man in der Informatik auch „gerichtete Graphen“. In ihnen ist eine Richtung ausgezeichnet: Bäume haben einen eindeutigen „Stamm“, das *Genus ultimum*, von dem ausgehend immer feiner werdende Äste abzweigen, die schließlich in den „Blättern“, den Spe-

zies, enden. Jedes Element in einem solchen Baum (jeder „Knoten“ des Graphen) hat einen eindeutigen Obertyp. Versucht man nun aber, eine Kreuzklassifikation in einen solchen Graphen umzuwandeln, erlebt man zwei Überraschungen. Erstens geht die Eindeutigkeit des Obertyps verloren. Ambulante Krebspatienten gehören sowohl zum Obertyp „Krebspatienten“ als auch zum Obertyp „Ambulante Patienten“. Die Verzweigungen gehen in einem solchen Diagramm nicht mehr nur in eine Richtung. Ein Element des Diagramms kann sowohl mehrere Untertypen, als auch mehrere Obertypen haben. Man sagt in diesen Fällen, dass diese Diagramme eine „multiple Vererbung“ erlauben (*multiple inheritance*) bzw. dass in ihnen „diamonds“ vorkommen können, also rautenförmige Strukturen.

Zweitens ergibt sich die Frage, welches der beiden Merkmale wir in der Hierarchie der Klassifikation zuerst abarbeiten sollen: Sollen wir die Patienten zuerst in solche mit bzw. ohne Krebs unterteilen und diese beiden Patientenklassen jeweils in solche, die ambulant und solche, die stationär behandelt werden? Oder sollen wir vielmehr die Patienten zuerst in solche, die ambulant und solche, die stationär behandelt werden, unterteilen, und dann diese beiden Klassen jeweils in Unterklassen von Patienten, die an Krebs leiden oder die nicht an Krebs leiden? Wir haben also die Wahl zwischen zwei verschiedenen Baumstrukturen.

Auf der untersten Ebene der Bäume, der Ebene der „Blätter“, ergibt sich in beiden Fällen dasselbe. Vielleicht wäre es für die medizinische Praxis völlig irrelevant, welche der beiden Möglichkeiten wir wählen würden. Aus philosophischer Sicht ist eine solche Arbitrarität, also die Möglichkeit zu einer willkürlichen Entscheidung dieser Frage, allerdings ein bemerkenswertes Phänomen. Sie kommt dadurch zustande, dass eine Kreuzklassifikation gerade darauf beruht, dass sie nach dem Vorliegen bzw. Nichtvorliegen zweier voneinander unabhängiger Merkmale fragt. Im Falle der Patienten im NCIT sind dies die Fragen:

(Q1) Aus welchem Grund wird der Patient behandelt?

(Q2) Bleibt der Patient über Nacht?

Während die erste Frage nach dem Grund für die Behandlung fragt, fragt die zweite Frage nach der Weise der Behandlung. Für den Arzt im Krankenhaus ist beides wichtig. Aber es sind Dinge aus ganz unterschiedlichen Kategorien, und diese sollte man streng auseinanderhalten.⁴

Eine Möglichkeit, solche Aspekte voneinander zu trennen, ist die „mehrachsige“ oder „mehrdimensionale“ Klassifikation. Dieses Verfahren wird z.B. von SNOMED angewandt, der *Systematized Nomenclature of*

⁴ Vgl. Auch Kumar & Smith (2005): „Where distinct specification factors are combined within a single tree errors often result.“

Human and Veterinary Medicine, die vom amerikanischen *College of American Pathologists* entwickelt worden ist.⁵ In der Version 3 unterscheidet SNOMED elf verschiedene Klassifikationsachsen, die mit 17 Qualifikatoren verknüpft werden können. Abbildung 2 listet einige dieser „semantischen Achsen“ von SNOMED auf. Nicht für jedes Krankheitsbild wird jede dieser Achsen benötigt. Aber indem mehrere dieser Achsen herangezogen werden, kann z.B. eine Virusenzephalitis bei einem Waldarbeiter kodiert werden als

TX2000 M40000 E30000 J63230,

wobei der mit „T“ beginnende Teil der Kodierung den Ort der Krankheit, der mit „M“ beginnende Teil den betroffenen Körperteil, der mit „E“ beginnende Teil die Ursache der Krankheit (den Virus) und der mit „J“ beginnende Teil den Beruf des Betroffenen angibt.

Abb. 2: Mehrdimensionale Klassifikation in SNOMED II (Dugas & Schmidt 2003, 80)

Welche morphologische Struktur?	Morphologie	M.
Findet sich wo?	Topographie	T
Was ist die Ursache?	Ätiologie (etiology)	E
Mit welchem Effekt?	Funktion	F
Bei welcher Krankheit?	Krankheit (disease)	D
Welche Prozeduren sind erfolgt?	Prozedur	P
Zusammenhang mit dem Beruf?	Beruf (job)	J

2.3 Uniformität: Labortiere

Patienten zugleich nach ihrer Krebsdiagnose und ihrem Übernachtungsstatus zu klassifizieren, führt nicht nur zu Problemen mit der Disjunktivität und Exhaustivität, sondern verstößt zugleich auch gegen das Uniformitätsgebot: Merkmale aus zwei ganz verschiedenen Bereichen werden hier zusammengebracht. Noch drastischer fällt dies bei der Klassifikation der Labortiere im NCIT auf, die, soviel steht fest, für die Krebsforschung äußerst wichtig sind. Das merkt man auch daran, dass der Obertyp „Laboratory animal“ im NCIT in zwölf vielfältige Untertypen unterteilt wird. Die Merkmale, mit Hilfe derer diese Untertypen gebildet werden, stammen allerdings aus ganz unterschiedlichen Bereichen. Der NCIT verstößt also auch hier gegen das Uniformitätsgebot.

⁵ Vgl. <http://www.snomed.org> und Winger (1984).

Einige der Untertypen von „Labortier“ basieren auf Dingen, die den fraglichen Tieren widerfahren sind. Eine „Genetically_Engineered_Mouse“ wird z.B. vom NCIT definiert als eine „Maus, die durch das Hinzufügen neuer genetischer Charakteristika genetisch verändert wurde“. In diesem Beispiel wird das Erleiden der DNA-Manipulation als Wesen der „Genetically_Engineered_Mouse“ angegeben. Andere Untertypen wie „Control_Animal“ basieren auf der jeweiligen Rolle, die das Tier innerhalb eines Versuchsaufbaus einnimmt. Denn Kontrolltiere sind, so die Definition im NCIT, „die Tiere in einer Studie, die nicht die zu testende Behandlung erfahren“.

Diese Definitionen berufen sich also auf Unterscheidungsmerkmale aus ganz unterschiedlichen Bereichen. Sie stammen aus ganz unterschiedlichen Kategorien: natürliche Arten, Leiden, funktionale Rollen. Diese kategorialen Unterscheidungen sollten in einer guten Klassifikation berücksichtigt werden.

2.4 Meta-Kategorien und „Anderes“

Auch hinsichtlich der Explizitheit und Präzision läßt der NCIT zu wünschen übrig. So, wie es in CTT einen Eintrag „und so weiter“ gibt, gibt es im NCIT einen Eintrag „Other“. Es ist ein Untertyp des Typs „General_Modifier“ (der ein Untertyp von „Qualifier“ ist, was wiederum ein Untertyp von „Properties_and_Attributes“ ist) und ist definiert als „Different than the one(s) previously specified or mentioned“. Außer dem bloßen „Andere(r/s)“ gibt es im NCIT noch ca. 80 andere Einträge der Form „Andere F“, z.B. „Carcinoma, Other, of the Mouse Pulmonary System“.⁶

Ganz wie CTT besitzt der NCIT auch Meta-Kategorien, also Kategorien, die ihrerseits von der Klassifikation abhängig sind, von der sie ein Teil sind. So enthält der NCIT z.B. die Kategorie „NCI-Thesaurus_Property“, die ein Untertyp von „Property“ ist und definiert wird als „A specific terminology property present in the NCI Thesaurus“.

Meta-Kategorien kommen im NCIT sogar in der obersten Ebene der Klassifikation vor. In dieser Ebene listet der NCIT nämlich den Typ „Retired_Concept“, den er definiert als „Concept has been retired, and should not be used except to deal with old data“. Hier geraten, um Freges Terminologie zu verwenden, eindeutig *Begriffseigenschaften* und *Begriffsmerkmale* durcheinander (Frege 1884, § 53; 1892). Während es ohne Zweifel

⁶ Für eine eingehende Kritik an solchen Formulierungen wie „other“ bzw. „not otherwise specified“ („NOS“) vgl. Bodenreider, Smith & Burgun (2004). In deutschsprachigen Informationssystemen wird oft der Ausdruck „o.n.A.“ („ohne nähere Angabe“) verwendet.

nützlich ist, in einer Klassifikation auch Eigenschaften der vorkommenden Ausdrücke oder Begriffe festzuhalten, so sollten sie doch nicht so behandelt werden, als ob sie Merkmale wären, die Dinge haben müssen, um die fragliche Universalie zu instantiieren.

3 Einschränkende Randbedingungen für Klassifikationen

Bei der Kritik von Borges' CTT und dem *NCI Thesaurus* habe ich mich von einem Idealbild von Klassifikationen leiten lassen. Nach diesem Idealbild ist eine Klassifikation eine Hierarchie von Klassen, die innerhalb einer Hierarchieebene paarweise disjunkt sind, gemeinsam den Gegenstandsbereich ausschöpfen und durch ontologisch fundierte Merkmale konstituiert werden. Es gibt eine Reihe von Gründen, warum real existierende Klassifikationen von diesem Idealbild abweichen.

Eine erste Gruppe von einschränkenden Randbedingungen hängt von dem Gegenstandsbereich ab, der zu klassifizieren ist. Gerade wenn es um die Klassifikation biologischer Arten geht, haben wir es mit dem Problem zu tun, dass eine große Zahl von Tier- und Pflanzenarten wissenschaftlich noch nicht beschrieben oder noch gar nicht entdeckt worden sind. Zudem werden mit neuen genetischen Methoden Artunterschiede entdeckt, die durch traditionelle phänotypische Methoden nicht unterschieden werden konnten. Schon die schiere Anzahl der Arten bietet eine Garantie dafür, dass den Biologen auf absehbare Zeit die Arbeit nicht ausgehen wird: Allein die Anzahl der Tierarten wird auf etwa 30 Millionen geschätzt. Es mag Bereiche geben, die, wie z.B. die Anatomie des Menschen, so gut wie vollständig und abgeschlossen wissenschaftlich erfasst sind. Andere Bereiche hingegen haben ein stetes Wissenswachstum zu verzeichnen, wie die Zoologie, die Botanik und ganz besonders die Genetik, die aufgrund der Datenintensität der genetischen Analysen ohne computergestützte Analysen heute kaum mehr möglich wäre. Prinzipiell muß aber immer mit der Entdeckung weiterer Arten gerechnet werden, nicht zuletzt auch deswegen, weil neue Arten entstehen können. Dies wäre eine Grenze für die Exhaustivität einer Klassifikation, die durch den zu klassifizierenden Gegenstandsbereich bedingt ist. Manche Gegenstandsbereiche geben dem Klassifikator auch grundsätzliche Probleme auf. Da die Gene von Bakterien auf vielfache Weise untereinander ausgetauscht werden können und sie sich durch die hohe Teilungsrate schnell ändern können, ist es besonders schwer, bei Bakterien stabile und trennscharfe Gattungen und Arten zu unterscheiden. Auch dies sind Grenzen der Klassifikation, die sich aufgrund des Gegenstandsbereiches ergeben.

Eine zweite Gruppe von einschränkenden Randbedingungen hängt ab von der technischen Seite der Realisierung und Anwendung der Klassifikation. Egal, ob diese in traditioneller, gedruckter Form oder in Form einer

computergestützten Datenbank vorliegt: Der zur Verfügung stehende Speicherplatz ist endlich, egal ob er in Druckseiten oder Festplattenkapazität besteht. Sollen Computerprogramme automatische Schlußfolgerungen aufgrund der eingegebenen Daten durchführen (*automated reasoning*), gibt es zusätzlich zum Speicherproblem das Problem der Berechenbarkeit. Die nötige Rechenzeit wächst an mit der Anzahl der Klassen, die zu verwalten sind, und mit der Anzahl der Relationen zwischen den Klassen, die repräsentiert werden sollen, und je nach gewählter Programmiersprache und der ihr zugrundeliegenden Beschreibungslogik und ihrer Ausdruckskraft (Nardi & Brachman 2002) bleibt stets die Gefahr, dass eine gestellte Aufgabe überhaupt nicht berechenbar ist.

Unabhängig von der Leistungsfähigkeit der verwendeten Hard- und Software ergeben sich auch durch den menschlichen Benutzer einschränkende Randbedingungen für Klassifikationen. Denn während es immer leichter und billiger wird, die Speicher der Computer zu erweitern, haben die kognitiven Fähigkeiten ihrer menschlichen Benutzer enge Grenzen. Menschlichen Dokumentatoren in Archiven und Bibliotheken wird empfohlen, nicht mehr als etwa tausend systematisch angeordnete Schlagworte zum Indexieren von Büchern oder Dokumenten zu verwenden (Gaus 2003, 93-94). Computer können natürlich deutlich mehr Termini verwalten; der *NCI Thesaurus* mit seinen 36.000 Termini ist dabei keineswegs die größte Terminologie-Datenbank. Das *Unified Medical Language System* (UMLS) umfaßte z.B. schon 2001 1,9 Millionen Ausdrücke mit über 800.000 unterschiedenen Bedeutungen (Dugas & Schmidt 2003, 81). Aber es sind menschliche Kuratoren, die diese Datenbanken erstellen, und menschlichen Benutzer, die sie später verwenden. Die Kuratoren sind Datenbank-Experten, die sich oft ausschließlich der Entwicklung dieser Art von Wissensspeichern widmen. Aber spätestens dann, wenn ein Hausarzt seine Abrechnungen aufgrund der Diagnosen-Codierung einer solchen umfangreichen Datenbank erstellen soll, stellt sich wieder die Frage, wieviele Diagnosen-Codes er in seinem Praxisalltag beherrschen kann.

Es sieht so aus, als ob das anvisierte Ideal nicht vollständig realisiert werden kann, weil unsere Wünsche sich wechselseitig beschränken: Wenn wir eine vollständige Repräsentation eines Wissenschaftsbereiches wollen, wird dies vermutlich weit von einem einfachen, verständlichen und benutzerfreundlichen System entfernt sein. Wenn Vollständigkeit bedeutet, eine große Menge an Daten anzuhäufen und viele Relationen zwischen den Einträgen zu definieren, können wir zudem Probleme mit der Berechenbarkeit bekommen. Wenn wir andererseits vereinfachende Typen wie „Andere“ oder „Nicht anderweitig bestimmt“ verwenden, bekommen wir Probleme beim Aktualisieren der Klassifikation, denn in unterschiedlichen Versionen kann „Andere“ ganz unterschiedliche Extensionen haben. Aber wenn wir davon absehen, solche „Andere“-Typen zu verwenden und kurzerhand das JEPD-Kriterium aufgeben, verlieren wir beträchtliche Möglichkeiten,

Schlüsse aus unseren Daten zu ziehen. Denn dann wissen wir nicht mehr, ob eine Entität, die zu einem Obertyp gehört, auch zu einem der entsprechenden Untertypen gehört, und so weiter.

4 Aufgabenteilung: Ein Lösungsversuch

Ein jüngerer Vorschlag zur Lösung dieses Dilemmas besteht in einer klaren Aufgabenteilung. Wir brauchen schlicht zwei Arten von Systemen: Zum einen benötigen wir Klassifikationen, die zu Referenzzwecken erstellt werden – Referenzklassifikationen oder, wie sie auch genannt werden, Referenzontologien. Zum anderen benötigen wir Klassifikationen, die für bestimmte Anwendungsgebiete maßgeschneidert sind – Anwendungsklassifikationen bzw. Anwendungsontologien. Die Referenzontologien sollen ohne Rücksicht auf Speicherprobleme und Rechenzeiten entwickelt werden. Sie sollen den Stand der jeweiligen Wissenschaftsdisziplin, aus der sie stammen, möglichst vollständig widerspiegeln:

Eine Referenzontologie ist analog zu einer wissenschaftlichen Theorie; sie hat einen einheitlichen Gegenstandsbereich, der aus Dingen besteht, die unabhängig von der Ontologie existieren, und sie versucht, die deskriptive oder repräsentationale Angemessenheit ihrem Gegenstandsbereich gegenüber zu optimieren, in dem höchsten Grad, der mit formaler Strenge und informationstechnischer Nützlichkeit kompatibel ist. Da eine Referenzontologie analog zu einer wissenschaftlichen Theorie ist, besteht sie aus Repräsentationen der biologischen Realität, die korrekt sind, wenn sie im Licht unseres gegenwärtigen Verständnisses betrachtet werden (und sie sollte daher einem regelmäßigen Update im Lichte des wissenschaftlichen Fortschritts unterworfen werden).⁷

Eine Anwendungsontologie hingegen ist analog zu einem technischen Produkt wie einem Computerprogramm. Bisher ist es üblich gewesen, für jeden Anwendungsfall von Grund auf neue Ontologien zu entwickeln. Das macht den Austausch und das Vergleichen von Daten natürlich schwierig. Anders ist es, wenn man auf eine bereits bestehende Referenzontologie zurückgreifen kann, von der ausgehend man die Anwendungsontologie erhält, indem man eine auf einen bestimmten Zweck hin zugeschnittene Auswahl oder Kombination von Typen der Referenzontologie vornimmt. Verschiedene

⁷ OBO Evaluation Criteria, Version 7 (January 14, 2006), http://sourceforge.net/mailarchive/forum.php?forum_id=12140&max_rows=25&style=nested&viewmonth=200602; meine Übersetzung. Dort auch das folgende Zitat.

solcher Anwendungsontologien sind dann durch ihre jeweilige Rückbindung an eine gemeinsame Referenzontologie leicht aufeinander abbildbar.

Während man also die Aufgabe der maximal-adäquaten Realitäts-Repräsentation der Referenzontologie zuweist, übernehmen es die Anwendungsontologien, das Speicher- und Berechenbarkeitsproblem in den Griff zu bekommen und Rücksicht auf die menschlichen Benutzer zu nehmen. Die zur *OBO-Foundry* zusammengeschlossenen Wissenschaftler⁸ sehen darin einen entscheidenden Fortschritt: „Die Methode, Anwendungsontologien ausschließlich vor dem Hintergrund von formal robusten Referenzontologien zu entwickeln und ihre Aktualisierung im Lichte der Aktualisierung der zugrundeliegenden Referenzontologie sicherzustellen, kann sowohl diesen Tendenzen der Vermehrung von Ontologien entgegenwirken, als auch die Interoperabilität der Anwendungsontologien sicherstellen, die nach ihren Vorgaben konstruiert worden sind.“

5 Exotisches Denken? Oder ein ungeeignetes Werkzeug?

Manche Philosophen haben gemeinsam mit Foucault darin übereingestimmt, dass Borges' CTT den „exotischen Zauber eines anderen Denkens“ offenbart (Foucault 1974, 17).⁹ Ich habe gezeigt, dass CTT in der Tat darin bezaubernd ist, dass es dazu dienen kann, eine große Bandbreite von Fehlern und Problemen zu illustrieren, die bei der Konstruktion von Taxonomien vorkommen können. Davon abgesehen ist CTT ein großartiges Stück Literatur, aber kein Zeugnis der Wissenschaft. Selbst wenn es ein solches wäre, wäre es kein Beweis für ein exotisches Denken, sondern eher ein Beispiel für eine sehr unpraktische Taxonomie. Es wäre jedenfalls ebensoviel oder ebensowenig ein Beispiel für exotisches Denken, wie der *NCI Thesaurus*, der seinem Anspruch nach ein wissenschaftliches Produkt ist, aber dieselben Eigentümlichkeiten aufweist, die CTT so seltsam erscheinen lassen. Die diskutierten Teile des NCIT sind aber keineswegs Beispiele exotischen Denkens, sondern tatsächliche Beispiele für höchst unpraktische Taxonomien. Zumindestens ist das *National Cancer Institute*, das den NCIT geschaffen hat und unterhält nicht mit seinem Zustand und seinem mutmaßlichen exotischen Charme zufrieden und will ihn dahingehend verbessern, solche Klassifikationsfehler zu vermeiden.¹⁰ Wie ich gezeigt habe, ist eine solche Verbesserung

⁸ Die *OBO Foundry* hat es sich zum Ziel gesetzt, durch Standardisierung die Interoperabilität der im Rahmen des OBO-Projektes („Open Biomedical Ontologies“) entwickelten Informationssysteme zu verbessern. Vgl. <http://obofoundry.org>.

⁹ Vgl. dazu auch die Beiträge in Jullien (2004).

¹⁰ Review of NCI Thesaurus and Development of Plan to Achieve OBO Compliance, <http://www.cbd-net.com/index.php/search/show/938464> (zuletzt besucht am

ein hervorragendes Beispiel dafür, dass auch technische Anwendungen auf den von der Philosophie gelegten Fundamenten aufbauen können.

LITERATUR

- Bodenreider, O., Smith, B. & Burgun, A. (2004). The ontology-epistemology divide: A case study in medical terminology, *Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004)*, hrsg. von A. Varzi & L. Vieu, 185-195. Amsterdam: IOS Press.
- Borges, J.L (1966). Die analytische Sprache John Wilkins', *Das Eine und die Vielen. Essays zur Literatur*, übers. von K.A. Horst. München: Hanser.
- Ceusters, W., Smith, B. & Goldberg, L. (2005). A terminological analysis and ontological analysis of the NCI Thesaurus. *Methods of Information in Medicine* 44, 498-507.
- Dugas, M., & Schmidt, K. (2003). *Medizinische Informatik und Bioinformatik. Ein Kompendium für Studium und Praxis*. Berlin u.a.: Springer.
- Foucault, M. (1974). *Die Ordnung der Dinge*. Frankfurt a.M.: Suhrkamp.
- Fragoso G. et al. (2004). Overview and utilization of the NCI Thesaurus. *Comperative and Functional Genomics* 5, 648-654.
- Frege, G. (1884). *Grundlagen der Arithmetik*. Breslau: Koebner.
- Frege, G. (1892). Begriff und Gegenstand. *Vierteljahrschrift für wissenschaftliche Philosophie* 16, 192-205.
- Gaus, W. (2003). *Dokumentations- und Ordnungslehre. Theorie und Praxis des Information Retrieval*, 4. verb. und erw. Aufl. Berlin, Heidelberg: Springer.
- Jullien, F. (2004). *Die Kunst, Listen zu erstellen*. Berlin: Merve Verlag.
- Kumar, A. & Smith, B. (2005). Oncology ontology in the NCI Thesaurus. *Proceedings of AIME 2005*.
- Nardi, D. & Brachman, R.J. (2002). An introduction to description logics, *The Description Logic Handbook. Theory, Application, Implementation*, hrsg. von F. Baader et al., Kap. 1. Cambridge: Cambridge University Press.
- Winger, F. (1984). *SNOMED – Systematisierte Nomenklatur der Medizin*. Berlin: Springer 1984.

24.01.2006). Für weitere Kritik am NCIT vgl. Ceusters, Smith & Goldberg (2005).