

Ontologie-basierte Qualitätssicherung medizinischer Terminologien

Barry Smith

*Institute for Formal Ontology and Medical Information Science (IFOMIS),
Universität des Saarlandes, Saarbrücken, Deutschland
Department of Philosophy, University at Buffalo.
Email: phismith@buffalo.edu, Internet: <http://ifomis.org>*

Einleitung

Die medizinische Ontologie soll helfen, die Flut medizinischer Information soweit zu kanalisieren, dass sie ohne Vereinfachungen oder Verzerrungen computer-gestützt verarbeitet werden kann. Nach ontologischen Kriterien aufbereitete Terminologiesysteme sollen darüber hinaus die Integration von Daten aus heterogenen Quellen unterstützen. Das im Jahr 2002 an der Universität Leipzig gegründete Forschungsinstitut IFOMIS stellt sich der Aufgabe, das gängige Instrumentarium der medizinischen Informatik mit Mitteln der philosophischen Ontologie zu erweitern. Im folgenden wird skizziert, wie die Anwendung philosophischer Prinzipien die Qualität medizinischer Terminologien erhöhen kann.

Systematische Schwächen medizinischer Terminologien

Die Mehrzahl der gebräuchlichen medizinischen Terminologien krankt an ungenau formulierten Begriffen; sie sind vielfach zirkulär und inkonsistent, da die zugrunde liegenden Regeln der Klassifikation unscharf sind. Diese Mängel können durch Menschen mit entsprechendem theoretischem Wissen oder praktischer Erfahrung ausgeglichen werden; in der elektronischen Verarbeitung sorgen sie aber verstärkt für Schwierigkeiten. Dadurch geht entscheidendes Potential für die automatisierte Verknüpfung von Information verloren.

In den 90er Jahren richteten sich umfangreiche Bemühungen auf die Standardisierung der Fachtermini und Kategorien verschiedener Domänen, um die

Verfügbarkeit und Wiederverwendbarkeit elektronisch gespeicherter Information dauerhaft zu sichern. Die medizinische Informatik brachte Ansätze wie GALEN, SNOMED, das Unified Medical Language System (UMLS) sowie dessen Semantic Network (SN) hervor. Dabei wurde zunehmend auf Formalismen wie Description Logic zurückgegriffen, um Konsistenz und Stringenz der Terminologiesysteme zu sichern und automatisches Schließen zu unterstützen. [1, 2] Kontrollierte Vokabulare wie die Gene Ontology (GO) trugen dazu bei, eine neue Phase der terminologischen Präzision in der biologischen Forschung einzuläuten, die die Integration biologischer und medizinischer Information ermöglicht.

Aber die mit kontrollierten Vokabularen angestrebte bloße syntaktische Exaktheit der Termini und die sie ergänzende Strukturierung mittels logischer Formalismen können nur bestimmten Mängeln von Klassifikationen – z. B. Zirkelbezügen – vorbeugen. Um tieferliegenden ontologischen Fehlern abzuhelfen, ist eine weiter reichende Methode erforderlich. [3] Für GO und SN ist diese Methode schon ausführlich dokumentiert. [4–7] Hier sollen Ergebnisse ihrer Anwendung auf SNOMED-CT (in der Version vom Juli 2003) skizziert werden. Dazu werden zunächst philosophische Prinzipien ermittelt, deren Nichtbeachtung die Qualität von Terminologien gefährdet, und dann Terminologien vor diesem Hintergrund analysiert, um die aus der Verletzung dieser Prinzipien resultierenden Fehler aufzudecken und zu kategorisieren. (Details über diese Methode finden sich in: [8, 9].)

Ontologische Grundrelationen

Ein Grundprinzip der Klassifikation lautet: *Gleiche Termini sind genau dann zu verwenden, wenn tatsächlich von denselben Entitäten die Rede ist.* Dieses Prinzip wird selbst im Hinblick auf die fundamentalsten Relationen, die die formale Architektur heutiger Klassifikationssysteme prägen, nämlich die *is_a*-Beziehung, also die Inklusionsverhältnisse zwischen Klassen, und die *part_of*-Beziehung zwischen Teilen und Ganzen, oft nicht beachtet.

Die philosophische Begriffsanalyse zeigt, dass die Verwendung von *is_a* und *part_of* in biomedizinischen Ontologien und Terminologien vielfach an gravierenden Inkonsistenzen leidet. [6] Mitunter werden die beiden Relationen sogar gleichgesetzt. So gilt in SNOMED-CT: *beide_Hoden is_a Hoden, beide_Gebärmutter is_a Gebärmutter* und *teilweise_Entfernung_der_Blase is_a Entfernung_der_Blase*.

Ontologische Grundkategorien

Ein weiteres zentrales Klassifikationsprinzip ist: *A is_a B gilt nur, wenn A und B zu denselben ontologischen Grundkategorien gehören.* Da aber die gebräuchlichen Terminologiesysteme aufgebaut wurden, ohne die implizit verwendeten Grundkategorien systematisch zu spezifizieren, finden wir auch hier eine Reihe vermeidbarer Fehler. In SNOMED-CT existiert z. B. *diagnostische_endoskopische_Untersuchung_eines_Mediastinums_NOS is_a Mediastinoskop*. Hier wird eine Prozedur als Instrument klassifiziert; ein Fehler, der durch eine klare Abgrenzung zwischen

Prozessen und Objekten zu vermeiden wäre. Analoges gilt für *Kontrazeption is_a funktionaler_Befund*, wo ein Verfahren mit dem Wissen über dieses Verfahren verwechselt wird.

Zusammengesetzte Termini

Ein drittes Prinzip: *Nicht jede mögliche Kombination von Termini darf selbst als Terminus zugelassen werden.* Vor allem Termini, die sich aus der Negation anderer ergeben, z. B. solche, die sich auf das nicht-Vorhandensein (eines Pulses, einer Niere) beziehen, sind mit Sorgfalt zu behandeln. Dafür spricht, dass verschiedene der in SNOMED-CT unter Verwendung von Operatoren der Verneinung wie *without* oder *with_no* konstruierte Termini fehlerhaft sind; so *Dupuytren-Krankheit, Knoten_ohne_Kontraktur is_a Kontraktur_der_Palmarfaszie* (*Dupuytren's disease of palm, nodules with no contracture is_a contracture of palmar fascia*).

Auch SNOMED-CTs Verwendung von Operatoren wie *complete* und *partial* für die Konstruktion von Termini ist fehleranfällig; z. B. gilt *Halbextraktion_aus_Steißlage is_a Extraktion_aus_Steißlage* und gleichzeitig *Extraktion_aus_Steißlage is_a vollständige_Steißgeburt* (*partial breech extraction is_a breech extraction is_a complete breech delivery*). Derartigen Problemen könnte SNOMED durch den Bezug auf die Prinzipien einer formalen Mereologie, d. h. eine Theorie der Teil-Ganzes-Beziehungen, vorbeugen. [10]

Raum und Topologie

Ein wichtiger Bestandteil philosophisch fundierter Ontologien sind leistungsfähige formale Theorien des Raums und seiner Topologie. [11] Deren Anwendung kann wiederum Fehlern wie dem folgenden vorbeugen: SNOMED-CT zufolge steht *Struktur_des_Nervus_tibialis* in der *is_a*-Beziehung sowohl zu *Oberschenkelteil* als auch zu *Unterschenkelstruktur* (*structure of tibial nerve is_a thigh part* und *structure of tibial nerve is_a lower leg structure*).

Nachbemerkung

Die dargestellten Beispiele deuten auf Unklarheiten im Verständnis der entsprechenden Ausdrücke, die durch die Beachtung ontologischer Prinzipien weitgehend ausgeräumt werden könnten. Die fehlende einheitliche Behandlung selbst der fundamentalsten Elemente von Terminologiesystemen wie der Grundrelationen *is_a* und *part_of* erschwert nicht nur die Arbeit derjenigen, die diese Systeme pflegen oder verwenden. Auch die dringend notwendige Integration mit anderen biomedizinischen Terminologiesystemen und somit die automatische Verknüpfung der in ihnen enthaltenen Information wird dadurch behindert.

Literatur

- [1] Horrocks I, Rector A, Goble C. A Description Logic Based Schema for the Classification of Medical Data. In: Baader F, et al. (eds.). Proc 3rd Workshop (KRDB 1996), 24-28.
- [2] Wroe CJ, Stevens R, Goble CA, Ashburner M. A Methodology to Migrate the Gene Ontology to a Description Logic Environment Using DAML+OIL. In: Pac Symp Biocomput 2003: 624-35.
- [3] Ceusters W, Smith B, Flanagan J. Ontology and Medical Terminology: Why Description Logics Are Not Enough. Proc TEPR 2003 – Towards an Electronic Patient Record. San Antonio, Texas. Boston, MA: Medical Records Institute (CD-ROM).
- [4] Smith B, Kumar A. The Formal Architecture of the Gene Ontology. In: Silico Biology, 2004.
- [5] Smith Barry, Jennifer Williams, Steffen Schulze-Kremer: The Ontology of the Gene Ontology. In: Proc AMIA Symp 2003: 609-613.
- [6] Smith, Barry, Cornelius Rosse: The Role of Foundational Relations in the Alignment of Biomedical Ontologies. In: Proc MedInfo 2004.
- [7] Smith, Barry, Jakob Köhler, Anand Kumar: On the Application of Formal Principles to Life Science Data: A Case Study in the Gene Ontology. In: Data Integration in the Life Sciences (DILS 2004) (Lecture Notes in Computer Science) Springer, Berlin, 79-94.
- [8] Bodenreider O, Smith B, Kumar A, Burgun A. Investigating Subsumption in DL-Based Terminologies: A Case Study in SNOMED CT. In: Proc KR-MED 2004.
- [9] Ceusters W, Smith B, Kumar A, Dhaen C. Ontology-Based Error Detection in SNOMED-CT. In: Proc Medinfo 2004..
- [10] Bittner, T. Axioms for Parthood and Containment Relations in Bio-
] Ontologies. In: Proc KR-MED 2004.
- [11] Cohn AG, Hazarika SM. Qualitative Spatial Representation and Reasoning:
] An Overview. Fundamenta Informaticae 2001; 46: 1-29.

Danksagung: Diese Arbeit wurde durch die Alexander von Humboldt-Stiftung im Rahmen des Wolfgang Paul-Programs unterstützt.