

Terminologien in der Medizin im Spannungsfeld von Logik und Ontologie

Ingenerf J

Institut für Medizinische Informatik, Universität zu Lübeck, Deutschland
 ingenerf@imi.uni-luebeck.de

Einleitung und Fragestellung

Die Notwendigkeit des Einsatzes standardisierter Terminologien in der medizinischen Informatik gilt als unbestritten. In diesem Beitrag werden verschiedene Einsatzbereiche der Daten- und Wissensverarbeitung vorgestellt. Aufgrund verschiedener Aufgabenstellungen unterscheiden sich die Anforderungen an eine terminologische Standardisierung. Besonders herausfordernd sind jene externen Terminologien, die einen interessierenden Wertebereich mit seinen Begriffen hierarchisch organisieren. Die enorme Produktivität medizinischer Begriffsbildung erfordert wiederum die Berücksichtigung von Relationen zwischen Begriffen. Für genau diese Aufgabenstellung haben sich Beschreibungslogiken etabliert, wie sie etwa in SNOMED CT verwendet werden. Diese begriffslogischen Ansätze wurden bereits mehrfach diskutiert und von statistischen Klassifikationen wie ICD oder Thesauren wie MeSH abgegrenzt [6]. Der vorliegende Beitrag setzt diese Betrachtung in gewisser Weise fort. Die Korrektheit logischer Deduktionen von Begriffswissen ist essentiell angewiesen auf das Beachten ontologischer Prinzipien. Den Potentialen reichhaltiger und korrekter Schlussfolgerungen steht eine Reihe von Konsequenzen gegenüber, die in diesem Beitrag diskutiert werden [7].

Material und Methoden

Im Vordergrund stehen in diesem Beitrag formale Terminologien, die möglichst ausdrucksstark und durch den Computer verarbeitbar sind. Zur formalen Definition von Begriffen haben sich mittlerweile Beschreibungslogiken durchgesetzt (z.B. GALEN, SNOMED CT). Die logische Sprache bietet gerade solche Ausdrucksmittel zur Komposition von Begriffen (genauer: von Begriffsintensionen), die entscheidbar und effizient analysierbar sind. Gleichzeitig kann im Rahmen der deduktiven Logik Vollständigkeit und Korrektheit garantiert werden, d.h. syntaktisch abgeleitete Sachverhalte entsprechen semantischen Folgerungen bei gegebenen Axiomen.

Nun haben die Namen von Begriffen und Relationen über die innerlogischen Bezüge hinaus auch fachsprachliche Bedeutungen für Menschen, die formale Terminologien erstellen und verwenden. Bei dieser Konfrontation mit der Wirklichkeit ergeben sich zwei aus anderen Gebieten bekannte Fehler: falsch-positive Ableitungen, d.h. logisch ableitbare Fakten sind falsch (z.B. multiple Magengeschwüre *isa* Magengeschwür) und falsch-negative Ableitungen, d.h. gültige Beziehungen sind logisch nicht herleitbar (z.B. Magenentzündung *isa* Magenschleimhautentzündung). Um diese Fehler zu reduzieren, hat sich seit einiger Zeit das Gebiet der Ontologie etabliert. Mit einem System von ontologischen Prinzipien lassen sich Subsumptionsbeziehungen existierender Begriffssysteme überprüfen [4]. Beachtet man etwa verschiedene Identitätskriterien für die Begriffe „*Helicobacter*“ und „*Infektionserreger*“, so kann man den so genannten ISA-Overload vermeiden. Die oben genannten formalen Terminologien (auch: „*conceptualist ontologies*“) werden qualitativ verbessert. Eine radikale Alternative ist der seit einiger Zeit propagierte Ansatz einer formalen Ontologie, die über eine extensionale Logik beschrieben wird. In bewusster Abkehr von Begriffsintensionen werden Beziehungen zwischen Instanzen in Raum und Zeit in Prädikatenlogik oder diversen Erweiterungen axiomatisiert (auch: „*Real Ontologies*“). Neben einem schlüssigen Kategoriensystem (Upper Model) werden insbesondere algebraische Eigenschaften von non-ISA-Relationen definiert, z.B. Transitivität oder Symmetrie [9].

Es sei festgehalten, dass die aktuell sehr virulente Diskussion von logischen und ontologischen Prinzipien für das effiziente und korrekte Deduzieren von Begriffswissen von fundamentaler Bedeutung ist. Allerdings werden dieselben Erkenntnisse auf verwandte Bereiche übertragen, die in Tabelle 1 aufgelistet sind. Das gilt für ein „*konzeptuelles*“ Datenbank- bzw. XML-Schema, für ein Data Dictionary im Rahmen kontrollierter Studien sowie für alternative externe Terminologien wie ICD oder MeSH.

Einsatzbereich	Beobachtung	Eingesetzte Methoden	Externe Terminologien	Wesentliche Ziele
1) Datenverarbeitung a) Modellierung eines Gegenstandsgebietes, u.a. klassische Datenmodelle, semistrukt. Dokumente. b) Standardisierte Dateneingabe, unter a) bereits genannte Bezüge zu externen Terminologien.	Festlegung von Objekttypen, Attributen und Wertebereichen. Vollständigkeit wird durch eingehende Person bestimmt; Vergleichbarkeit durch „ <i>Logik</i> “.	Konzeptuelle Modellierung im Datenbankbereich, u.a. ER- oder UML-Standard, u.a. auch HL7 V3. Beschreibungslogiken; gerne auch mächtigere Logiken, u.a. für ontologisch Inferenzen (Raum, Zeit).	Offen für alle Standards; u.a. wenn fokussiert auf einzelne Wertebereiche. Ausdrucksstarke Terminologien wie SNOMED CT.	Interoperabilität verteilter Anwendungen; z.B. Zusammenführung verteilter Daten. Deduzierbarkeit von Begriffsbeziehungen für definierte und postkoordinierte Begriffe.
b*) Standardisierte Dateneingabe, erweiterte Festlegung ausgewählter Anteile, u.a. klinische Studien.	Vollständigkeit und Vergleichbarkeit werden bestimmt durch vorab def. Auswertungsfragen.	Uneinheitliche Methoden, u.a. bekannt als Data Dictionary, Merkmalskatalog, usw.	Prädefinierte Wertebereiche, u.a. Klassifikationen wie ICD und OPS.	Stat. Auswertbarkeit verteilter Datenbestände; vor allem für vorab definierte Fragen.
2) Wissensverarbeitung a) Rechner-interpretierbare Wissensrepräsentation, inkl. Interpretation von zugrunde liegenden Daten. b) Mensch-interpretierbare Wissens-Texte in Dokumenten, evtl. ergänzt um Metadaten.	Möglichst zuverlässiger Zugriff auf Attribut-Wert-Kombination relevanter Objekte. Ergänzung von Attribut-Wert-Kombinationen auf der Meta-Ebene.	Logikformalismen, u.a. ARDEN-Syntax oder GLIF. Metabeschreibungsstandards, z.B. Dublin-Core, RDF und OWL (<i>Semantic Web</i>).	Ausdrucksstarke Terminologien wie SNOMED CT. Thesauren wie MeSH als Schlagworthierarchie mit Gebrauchbedingungen.	Verknüpfbarkeit von Wissen und Daten; z.B. Welche linke Regelseite wird instanziiert? Information Retrieval von Dokumenten. Mit RDF usw. existieren weitere Ziele.

Tab.1 Terminologische Standardisierung in verschiedenen Bereichen der Daten- und Wissensverarbeitung in der Medizin

Ergebnisse und Diskussion

Die Konfrontation zwischen den genannten Alternativen „*conceptualist versus real Ontologies*“ verdeutlicht bereits vorab ein Ergebnis: die bekannte Problematik, zwischen zwei Übeln wählen zu müssen, wird neu entfacht [3]. Entweder man schränkt mit den Beschreibungslogiken die Ausdrucksmächtigkeit zugunsten der Berechenbarkeit ein und akzeptiert ein gewisses Maß an falsch positiven bzw. falsch negativen Deduktionen. Oder man öffnet die „*Dose der Pandora*“ mit allen Konsequenzen, die aus den 80er und 90er Jahren aus den theoretischen und vor allem praktischen Arbeiten zur (symbolischen) Künstlichen Intelligenz resultierten.

Mit Blick auf Tabelle 1 ergeben sich weitere Erkenntnisse, wenn man formale Terminologien (oder Ontologien) zur standardisierten Repräsentation von medizinischen Daten oder Wissen verwendet. Im Folgenden werden für die Bereiche nur stichwortartig einige Erkenntnisse genannt. Bei der Diskussion vor allem von Datenbankschemata und Merkmalsverzeichnissen (Data Dictionaries) geht es im Wesentlichen um die „*epistemologische*“ Frage: „*Was können wir wissen?*“ (oder: Welche Daten sind verfügbar?) im Unterschied zur ontologischen Frage „*Was ist?*“. Es stellt sich heraus, dass die Abgrenzung für das Funktionieren von logischer Deduktion sehr wichtig ist. In formalen Terminologien wie SNOMED CT ist auf epistemologisch angereicherte Begriffe wie „*Tuberkulose, histologisch nicht gesichert*“ zu verzichten [1]!

1) Datenverarbeitung

1a) Ein Datenmodell ist nicht primär eine Abbildung der Wirklichkeit, sondern ein Modell mit relevanten Sichten auf Repräsentationen der Wirklichkeit im Sinne einer zu lösenden Aufgabe. Beim Einsatz externer formaler Terminologien ergeben sich komplexe „*Zuständigkeitsfragen*“ (z.B. Familienanamnese Diabetes in DB oder Term?).

1b) Wenn logikbasierte Systeme ihr Potential zur Postkoordination ausschöpfen, stellt sich die Frage der softwaretechnischen Integration des Beweisystems in der medizinischen IT-Infrastruktur (Terminologieserver?). Der Einsatz einer formalen Ontologie mit gänzlich neuen Herausforderungen wird hier kurz angesprochen [2].

1b) Data Dictionaries bzw. Merkmalsverzeichnisse im Zusammenhang mit geplanter Dokumentation (u.a. für Studien) teilen mit statistischen Klassifikationen wie ICD die Zielsetzung einer möglichst vergleichbaren Datenerfassung. Der Anteil epistemologisch geprägter Inhalte ist groß (z.B. fehlende Werte, negative Fakten). Dieses kollidiert mit ontologischen Prinzipien und logischer Deduzierbarkeit [8]. Umso verwunderlicher sind die Arbeiten zu formalen Ontologien im Kontext von Studien [5].

2) Wissensverarbeitung

2 a) Der Einsatz symbolischer Wissensverarbeitung ist auf standardisiert repräsentierte klinische Daten angewiesen, siehe MEDWIS-Projekt in den 90er Jahren. Mit möglichst formalen Terminologien ergeben sich verwandte Probleme zu 1a) sowie Abgrenzungsfragen von terminologischem und kontingentem (Welt-)Wissen.

2b) Für ein verbessertes Dokumenten Retrieval werden Thesauren wie MeSH verwendet. Die Anwendung ontologischer Prinzipien ist auch hier zu hinterfragen [7].

Zusammenfassend sei festgehalten, dass die Anwendung logischer und ontologischer Prinzipien für definierte Zwecke unbedingt erforderlich ist. Diese Prinzipien werden für die beim praktischen Einsatz tangierten Bereiche in möglichst ganzheitlicher Weise hinterfragt und diskutiert. Dabei werden eine Reihe von Erkenntnissen zusammengetragen, z.B. ein prinzipieller Konflikt zwischen SNOMED CT („*Epistemologie-frei*“) und ICD („*Epistemologie-behaftet*“) mit Konsequenzen fürs Mapping [7].

Literatur

[1]	Bodenreider O, Smith B, Burgun A. The Ontology-Epistemology Divide: A Case Study in Medical Terminology. In: Varzi A, Vieu, L., editor. Proc. of the Int. Conf. on Formal Ontology and Information Systems (FOIS 2004), Turin, 4-6. November 2004; Amsterdam: IOS-Press; 2004. p. 185-195.
[2]	Ceusters W, Smith B. Tracking referents in electronic health records. Stud Health Technol Inform 2005;116:71-6.
[3]	Doyle J, Patil RS. Two theses of knowledge representation: language restrictions, taxonomic classification, and the utility of representation services. Artificial Intelligence 1991;48:261-297.
[4]	Guarino N, Welty CA. Evaluating Ontological Decisions with OntoClean. Communications of the ACM 2002;45(2):61-65.
[5]	Heller B, Herre H, Lippoldt K. Domain-Specific Concepts and Ontological Reduction within a Data Dictionary Framework. In: Rahm E, editor. Data Integration in the Life Sciences (DILS 2004), März 2004 in Leipzig; Lecture Notes in Bioinformatics, Vol. 2994. Heidelberg: Springer; 2004. p. 47-62.
[6]	Ingenerf J. Standardisierte Terminologien in der Medizin zwischen Begriffssemantik und Anwendungspragmatik. In: Proc. der 50. GMDS-Jahrestagung, September 2005 in Freiburg; Meeting-Abstract, online unter: http://www.egms.de/en/meetings/gmds2005/05gmds492.shtml ; 2005.
[7]	Ingenerf J, Linder R. Ontological Principles Applied to Biomedical Vocabularies. In: Proc. of FCTC 2006 (erscheint erst 12 April 2006), siehe http://medinfo.umft.ro/stc2006/workshop1.htm
[8]	Rector AL. Faithfulness or comparability. Methods of Information in Medicine 1996;35(3):218-219.
[9]	Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. Genome Biology 2005;6(5):R46, siehe http://genomebiology.com/2005/6/5/R46 .