

Medizinische Ontologien: das Ende des MeSH

Medical ontologies: the end of MeSH

Abstract

Since the beginning of information technology the complexity of medical questions and medical information management is an important topic which challenges computer scientists.

In the eighties of last century artificial intelligence went awry. Though some core ideas of AI have brought up fruitful results. After all congruent development in a number of different scientific disciplines and the exponential development in computer hardware could meet the high requirements in medical information search. In 2000 Tim Berners-Lee's programmatic request for a Semantic Web gained the ontology topic broader attention.

Already 20 years ago NLM started to develop the Unified Medical Language System (UMLS). So in medicine (PubMed) ontology integrated into a semantic net is in operation. Hence it is high time for medical librarians and documentalists to get into this topic although it is covered by a smoke screen of terminology from IT. Ontologies can be understood as tools for classification. So essential contributions from library and documentation science could be expected.

This paper should open an entrance to the topic. It will explain fundamental elements of UMLS and includes an annotated list of literature for further studies.

Keywords: classification, ontology, Semantic Web, UMLS, PubMed, information retrieval

Zusammenfassung

Die Komplexität medizinischer Fragestellungen und des medizinischen Informationsmanagements war seit den Anfängen der Informatik immer ein besonders wichtiges Thema. Trotz des Scheiterns der Künstlichen Intelligenz in den 80er Jahren des vorigen Jahrhunderts haben deren Kernideen Früchte getragen. Durch kongruente Entwicklung einer Reihe anderer Wissenschaftsdisziplinen und der exponentiellen Entwicklung im Bereich Computerhardware konnten die gestellten, hohen Anforderungen bei der medizinischen Informationssuche doch noch erfüllt werden. Die programmatische Forderung von Tim Berners-Lee betreffend "Semantic Web" im Jahr 2000 hat dem Thema Ontologien für maschinenlesbare Repositorien in Allgemein- und Fachsprache breitere Aufmerksamkeit gewonnen. Da in der Medizin (PubMed) mit dem von NLM schon vor 20 Jahren entwickelten Unified Medical Language System (UMLS) eine funktionierende Ontologie in Form eines semantischen Netzes in Betrieb ist, ist es auch für Medizinbibliothekare und Medizindokumentare hoch an der Zeit, sich damit zu beschäftigen. Ontologien können im Wesen, trotz der informatisch vernebelnden Terminologie, als Werkzeuge der Klassifikation verstanden werden. Hier sind von seiten der Bibliotheks- und Dokumentationswissenschaft wesentliche Beiträge möglich. Der vorliegende Bericht bietet einen Einstieg in das Thema, erklärt wesentliche Elemente des UMLS und schließt mit einer kommentierten Anmerkungs- und Literaturliste für die weitere Beschäftigung mit Ontologien.

Constantin Cazan¹

1 Wien, Österreich

Schlüsselwörter: Klassifikation, Ontologie, Semantisches Netz, medizinische Literatursuche

Medizinische Ontologien: das Ende des MeSH

Vorbemerkung

Versierte Nutzerinnen (Bis zur vollständigen Gleichberechtigung der Frau auf der Welt habe ich beschlossen die weibliche Endung in meinen Texten zu wählen.) der Datenbank Medline haben vielleicht bemerkt, dass schon seit einiger Zeit das Hauptordnungselement MeSH (Medical Subject Heading) nicht mehr als Keyword bezeichnet wird, sondern als Concept. Für die Masse der Benutzerinnen mag das nicht weiter beachtenswert sein, aber die aufmerksame und professionelle Nutzerin wird das sicher stutzig machen.

So ist es auch dem Autor dieser Zeilen gegangen. Diese eigenartige Veränderung spiegelt eine Entwicklung, die bis ins Jahr 1986 zurückreicht, ja eigentlich noch weiter. Man muss leider wieder einmal ganz emotionslos feststellen, dass die traditionelle Dokumentationswelt und schon gar nicht die Bibliothekswelt in deutschen Landen davon keine merkbare Notiz genommen hat.

Das ist per se keine Überraschung. Aber in diesem speziellen Fall könnte das Übersehen und das nicht sorgfältige Mitlernen des Themas Ontologie, sowie weiterer, unter dem Begriff „Semantische Technologie“ subsummierter Entwicklungen, erstere Auswirkungen auf die heutige (wissenschaftliche) Bibliothekswelt haben.

Google und andere Mitbewerber haben des Thema Suchen und Finden kämpferisch und technologisch innovativ bereits global besetzt. Dadurch könnte das Thema „Semantische Technologie“ zu einer technisch dominierten Disziplin werden. Dies führt unter Umständen dazu, dass eine Kernkompetenz bibliothekarischer und dokumentarischer Tätigkeit, nämlich jene der sachlichen Erschließung, die Berufsgruppe oder die Disziplin wechselt. Das ist sicher ein Worst-Case-Szenario und meistens kommt es anders als man denkt. Aber vorstellbar ist es allemal. Die Behebung der semantischen Defizite bei der Suche im Internet und in anderen Informationssammlungen sehe ich eng verknüpft mit der Zukunft bibliothekarisch-dokumentarischer Dienstleistungen. Je bessere und je ausgereifere semantische Werkzeuge zu günstigen Preisen auch in Spezialdomänen angeboten werden, desto enger wird es für bibliothekarisch-dokumentarische Dienste werden.

Ad fontes oder über die Wurzeln semantischer Technologien

Versierte Kolleginnen wissen aus vielerlei Erfahrungen, dass in der Medizin samt ihren Teil- und Nachbardisziplinen tagtäglich, im Spannungsfeld von Notfällen mit begrenzten Ressourcen und großem Informationsanfall,

viele schwierig zu handhabende Situationen entstehen. Daher wurden schon sehr früh aus der und rund um die Medizin Überlegungen und Projekte entwickelt, möglichst viel automations- oder computerunterstützt durchzuführen.

Eines der Gebiete, auf dem seit vielen Jahrzehnten gearbeitet wird, verfolgt das Ziel, Ärztinnen bei der Diagnosestellung möglichst umfassend zu unterstützen. Bei der täglichen Flut an neu bekannt werdenden Ergebnissen und der Tatsache, dass jede Patientin letztendlich ein Spezialfall ist, wurde schon früh versucht, mit den Methoden der sog. Künstlichen Intelligenz (KI) technische Hilfssysteme zu entwickeln. Obwohl hier viele Hoffnungen, Geld und Expertinnen-Jahre hineingeflossen sind, haben diese Ansätze keine durchbrechenden Erfolge gebracht. Zu komplex - muss man einfach feststellen - waren die zu überwindenden Hürden und Hindernisse. Diese lagen in der damals z.T. noch mageren oder schwachen Leistung der eingesetzten Hardware, die höchst komplexe und zeitintensive Berechnungen großer Text- und Daten corpora nicht entsprechend durchführen konnte. Netzwerke, verteilte Datenbanken und elektronische Diagnose-Systeme standen nicht zur Verfügung. Last but not least schaffte die menschliche Sprache in ihrer Vieldeutigkeit bedeutende, um nicht zu sagen kaum bewältigbare Hindernisse.

Doch mit den Jahren und vielen z.T. parallel laufenden Entwicklungen entstand und entsteht doch ein mächtiges Netzwerk aus Know How und technischer Infrastruktur, das viele der Hürden der 80er Jahre überwindet oder zumindest deren Überwindung möglich erscheinen lässt. Neben höchst leistungsfähiger Hardware, weltweiten Hochgeschwindigkeitsnetzen und einer weitgehend elektronifizierten Medizin in den großen Industriestaaten wurden auch im Bereich der Sprachwissenschaft, der Computerlinguistik, der Textanalyse, der Bilderkennung, der Robotik und anderer Disziplinen an der Schnittstelle von Technik und Humanwissenschaften bedeutende Fortschritte erzielt.

Es soll natürlich nicht verschwiegen werden, dass es viele schwierige und schlecht oder gar unbearbeitete Themen in diesem Kontext gibt und viele Irr- oder Umwege gegangen wurden.

So etwa ist das ganze Theoriegebäude rund um das Thema Ontologien unter kompletter Ignorierung der Ergebnisse der Dokumentations- und Thesauruswissenschaften aufgebaut worden, mit allen negativen Folgen, die hier nur denkbar sind. Es muss also festgestellt werden, dass zwar viele verschiedene Disziplinen, wie oben angeführt, eine gewisse Mächtigkeit in ihrem Methodeninventar erreicht haben um die Probleme z. B. rund um natürlichsprachige Computersysteme bewältigbar erscheinen zu lassen. Es gibt aber zuwenig Kommunikation und zuwenig Schnittstellen zwischen all diesen Disziplinen, um die Ergebnisse wechselseitig nützen zu können.

Was ist eine Ontologie im informatischen Sinne ?

Nicht wenige Kolleginnen in unserer Profession haben geisteswissenschaftliche oder andere humanistische Bildungswege beschritten, bevor sie im Informationsgeschäft gelandet sind. Ihnen ist der Begriff Ontologie besser oder schlechter, klarer oder diffuser als philosophische Disziplin geläufig. Den Import des Begriffs in die moderne Web-Informatik kann man im besten Falle noch als Glück im Unglück bezeichnen. Glück insofern, als man im antiken und mittelalterlichen Verständnis (Aristoteles, Thomas von Aquin) in der Ontologie eine Denkschule der Klassifizierung der Welt durch Bezeichnungen mit Symbolen und Worten identifizieren kann, also Elemente wiederfindet, um die es der modernen Web-Informatik letztendlich auch geht.

Unglücklich, weil die philosophische Ontologie mit vielen schwierig zu durchschauenden Denk- und Geistesstraditionen überladen ist. Unglücklich, weil der Begriff viele davon abhält, sich im Bewusstsein eben dieser Überladung und des Gefühls des „Nichtdurchschauenskönnens“ damit auseinander zu setzen. Unglücklich, weil er letztlich verhindert, eine wichtige Innovation für viele Menschen leicht verständlich zu machen [1].

„An ontology is a specification of a conceptualization“ [2]. Diese immer wieder in der informatischen Welt zitierte Definition einer Ontologie von Tom Gruber, einem aus der KI-Forschung kommenden Informatiker, ist keine didaktische Meisterleistung und höchstens für studierte Informatiker eine verwertbare Aussage.

Besser verständlich ist die folgende Definition von Neches et al: *“An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary“* [3].

Es handelt sich also um ein kontrolliertes Vokabular. Ein Thesaurus definiert aber nur ein fixes Set von Beziehungen zwischen Vorzugsbegriffen, die im wesentlichen dazu dienen, einen definierten Corpus von Information, also eine Dokumentensammlung, besser zu erschließen und zu durchsuchen. Im Unterschied dazu ist die Ontologie darauf angelegt, die gesamte natürlichsprachige Repräsentation mit all ihren Facetten semantischer und syntaktischer Natur stimmig abzubilden. Das Set an Beziehungen zwischen den Begriffen ist dabei ebenso frei und unbestimmt, wie die Möglichkeiten die Begriffe selbst zu kategorisieren.

Ein mir persönlich hilfreicher Vergleich ist etwa der zwischen HTML und SGML(XML). Im Falle von HTML gibt es ein fixes Set von Elementen, die vordefiniert sind. In SGML(XML) können über die sog. DTD (Document Type Definition) beliebig viele Elemente und ihre Beziehungen definiert werden.

Das heißt, eine Ontologie ist zu aller erst einmal ein Corpus von Wörtern mit allen notwendigen syntaktischen und semantischen Beziehungen, um die natürliche Sprache zu repräsentieren. Dies kann entweder ein klei-

nes oder mittleres Fachgebiet betreffen. Man spricht dann von einer Domänen-Ontologie. Oder es betrifft kein Fachgebiet im Speziellen, sondern den gesamten Basis-Wortschatz um eine allgemeine Verständigung in einer Sprache zu ermöglichen. Dann wird von „upper ontologies“ gesprochen.

Aus informatischer Sicht fehlt aber noch etwas Wesentliches. Hier wäre vielleicht anzumerken, dass dieses, aus informatischer Sicht fehlende Element, letztendlich zwei Entwicklungen oder Teildisziplinen trennt, die durchaus nebeneinander existieren:

Das fehlende Element aus informatischer Sicht ist das Faktum, dass die Ontologie, wie sie etwa von Tim Berners-Lee in seinem programmatischen Entwurf des „Semantic Web“ eingeführt wurde, ein von Agenten maschinenlesbares Wörterbuch ist, das die sogenannte Semantische Interoperabilität zwischen verschiedenen Informationssystemen ermöglicht. Das heißt, Berners-Lee u.v.a. Informatiker verstehen unter einer Ontologie vor allem einmal etwas rein Informatisches. Berners-Lee etwa verstieg sich anlässlich einer Keynote der XML-Konferenz am 6. 12. 2000 sogar zu der imho etwas überspannten Definition *„that the word semantic meant machine processable. He explicitly rules out the sense of natural language semantics“* [4]. Dies als weiterer Beleg für die wenig sensitive Wahl von Begriffen in der informatischen Welt der Ontologien.

Trotz dieses rigoros wirkenden Ansatzes zeichnen sich schon jetzt, wie oben angedeutet, zwei Disziplinen ab.

Einerseits die informatische Disziplin, die Werkzeuge zum Erstellen von Ontologien, Standards und sog. Ontologiesprachen entwickelt, also die ganze informatische Infrastruktur, die notwendig ist, um das von Berners-Lee postulierte „Semantic Web“ überhaupt maschinell zu ermöglichen.

Die andere Richtung wird beeinflusst von der philosophischen Ontologieschule eines Barry Smith und von den sprachwissenschaftlichen Disziplinen aus dem Bereich Computerlinguistik und Textanalyse sowie den informationswissenschaftlichen Disziplinen aus dem Bereich der Dokumentation, Thesaurusforschung und Katalogisierung. Selbst den Informatikern beginnt also langsam zu dämmern, dass die Sprache in ihrer natürlichen Ausprägung als Umgangssprache des Alltags, die vielen Fachsprachen (Medizin, Chemie, Technik, Handel etc.) und nicht zu vergessen auch alle Weltsprachen, Dialekte und Schriftsysteme nicht so ganz simpel „agentisierbar“ sind.

So werden zunehmend Projekte und Konzepte entwickelt, um neue Werkzeuge für sog. „Domänenexperten“ zu schaffen, Spezialisten, die keine Informatiker sind, aber Domänenwissen besitzen. Und wer anderer kann das letztendlich sein, als Terminologie- oder sonstige Expertinnen der verschiedenen Fachdisziplinen?

Vom MeSH zum UMLS – Wege zu mehr semantischer Mächtigkeit.

Für gelernte Dokumentarinnen oder Bibliothekarinnen entbehrt es nicht eines gewissen Amusements, wenn die

informatischen Ontologie-Adepten, erzwungen durch die Nichtwahrnehmung anderer Disziplinen, das Rad der Klassifikation de novo erfinden. Ein besonderes Gustostückerl bieten hier Natalya Noy und Deborah McGuinness [5], die unter Verwendung von aus Programmierung und Datenbankentwicklung bekanntem Vokabular eine Thesaurus-Terminologie faktisch neu strickten. Einen gewissen Vorgeschmack auf die daraus resultierende Sprachverwirrung bietet etwa die Arbeit von Sigel [6], in der versucht wird, der ISKO-Community die Ontologie-Terminologie auseinander zu setzen.

Im Grunde bestehen ja drei Möglichkeiten zu Ontologien, also maschinenlesbaren kontrollierten Vokabularen mit größtmöglicher semantischer Mächtigkeit, zu kommen:

1. Bestehende Vokabularen unterschiedlicher Qualität und semantischer Mächtigkeit durch semantische Beziehungen weiter anreichern.
2. Verschiedene Vokabularen zusammenführen und ggf. durch ein semantisches Meta-Vokabular verknüpfen.
3. Ontologien neu verfassen.

Alle drei Wege werden in der Praxis beschränkt, wobei in der Medizin besonders der zweite Weg auf Grund historischer Bedingungen Bedeutung hat. In anderen Sachbereichen, wo etwa gut entwickelte Thesauri bestehen, wird der erste Weg oder auch eine Kombination aus Weg 1 und 2 beschränkt. Den dritten Weg habe ich bisher nur bei experimentellen Projekten gesehen.

In der Medizin haben Fachterminologien eine lange Geschichte. So ist es nicht verwunderlich, dass die National Library of Medicine mit ihrem gut gepflegten und kontinuierlich weiterentwickeltem Thesaurus „Medical Subject Headings“ schon früh nach Möglichkeiten gesucht hat, Einschränkungen zugunsten einer in der Medizin global und breit anwendbaren Terminologie zu überwinden.

Auch die Produzenten der Datenbank EMBASE begannen in den 90er Jahren, zwar imho aus den falschen Motiven und mit begrenztem Erfolg, einen vergleichbaren Weg zu beschreiten, indem sie kontinuierlich MeSH Begriffe als Synonyme in den EMBASE-Thesaurus Emtree einarbeiteten.

Aber NLM war, soweit ich das feststellen kann, die erste, die sich auf diese Spur setzte. Es wäre sicher nicht uninteressant der Frage nachzugehen, was oder wer die NLM schon in den 80er Jahren dazu brachte auf das Thema der semantischen Anreicherung zu setzen.

Im Jahre 1986 wurde das sog. UMLS Project initiiert „as a long-term NLM research and development effort to facilitate the retrieval and integration of information from multiple machine-readable biomedical information sources.“ [7].

Das Ziel des auch als „Semantisches Netz“ bezeichneten UMLS (=Unified Medical Language System) liegt in der semantisch konsistenten Zusammenführung verschiedener internationaler, in der Medizin etablierter Klassifikationen zu einem Instrument, dessen Wert und Mächtigkeit höher ist als die Summe seiner Einzelteile (m.E. könnte man auch von einer Topic Map sprechen, wobei diese

einen eigenen Standard repräsentiert und für den Autor nicht ganz klar ist, ob dieser Standard hier voll erfüllt wird).

1986 wurden folgende Klassifikationen als Quellmaterial für das UMLS herangezogen und durch klassifikatorische Arbeit semantisch angereichert und zusammengefasst:

- MeSH – NLMs Medical Subject Headings
- DSM – Diagnostic and Statistic Manual of Mental Disorders
- ICD CM – International Classification of Diseases – Clinical Modification
- SNOMED – The College of American Pathologists Systematized Nomenclature of Medicine
- CPT – American Medical Association´s Current Procedural Terminology
- LCSH – Library of Congress Subject Headings
- COSTAR - Clinical terms

Heute sind im UMLS mehr als 130 verschiedene Vokabularen ganz oder teilweise integriert (z.T. sogar mit mehreren Sprachversionen), wie z.B. das ICPC93 - The International Classification of Primary Care mit 13 Europäischen Sprachen (<http://www.nlm.nih.gov/research/umls/metaa1.html>).

Auch wenn es in den Quellen zum UMLS nicht explizit genannt wird, lässt sich an der heutigen Eigendefinition des UMLS (http://www.nlm.nih.gov/research/umls/about_umls.html) leicht erkennen, dass das UMLS eine, auch im informatischen Sinne und vor allem auch in der Praxis voll eingesetzte Ontologie ist:

The purpose of NLM's Unified Medical Language System® (UMLS) is to facilitate the development of computer systems that behave as if they "understand" the meaning of the language of biomedicine and health. To that end, NLM produces and distributes the UMLS Knowledge Sources (databases) and associated software tools (programs) for use by system developers in building or enhancing electronic information systems that create, process, retrieve, integrate, and/or aggregate biomedical and health data and information, as well as in informatics research.

Die Ontologie des UMLS besteht heute aus 4 Teilen:

- Metathesaurus, dessen Quellen bereits besprochen wurden
- sog. Semantic Network
- Specialist Lexicon
- MetamorphoSys

Der **Metathesaurus** bildet dabei die Grundmenge aller, aus allen berücksichtigten Vokabularen einbezogenen Vorzugsbegriffe, mit all ihren Relationen, Synonymen und sonstigen Querbeziehungen ab, durchaus ähnlich dem klassischen Thesaurus. Der markanteste Unterschied gegenüber dem klassischen Thesaurus liegt in der Ersetzung der in der alten Thesauruslehre als „Vorzugsbegriff“ definierten Einheit durch den Begriff des Konzepts (engl. concept). Dieser Begriff kommt aus dem informatischen Ontologie-Ansatz (computational ontology) bei dem Ontologie eine Datenstruktur (von Informatikern auch gerne

als Graph bezeichnet) mit Knoten (nodes) bezeichnet. Knoten repräsentieren ein einzelnes Konzept, das von James Geller als „a unit that one can think about“ umschrieben wird, bzw. mit „Concepts correspond to words or short phrases. Typically, concepts correspond to nouns or noun phrases, but they don't have to“ (http://web.njit.edu/~geller/what_is_an_ontology.html) definiert wird. Der informatische Konzept-Begriff wird von Barry Smith [8] stark kritisiert.

Das UMLS beinhaltet heute ca. 1 Million biomedizinische Konzepte mit etwa 5 Millionen Konzept-Namen (Quelle: <http://de.wikipedia.org/wiki/UMLS>, letzter Zugriff am 17.8.2006).

Die Tatsache, dass das Wort "Konzept" im Sinne der „computational ontology“ nicht in Umstätters semiotischem Thesaurus (<http://www.ib.hu-berlin.de/~wumsta/infopub/textbook/definitions/thesauindex.html>) vorkommt, sondern nur im umgangssprachlichen Sinn, ist vielleicht ein weiterer, ganz guter Beleg für die Nichtwahrnehmung wichtiger Entwicklungen in den eigenen Kreisen, wie in der Vorbemerkung angesprochen.

Das „**Semantic Network**“ repräsentiert imho die dazugehörige DTD (Document Type Definition) in dem Sinne, als dieser Teil der Ontologie alle möglichen Relationstypen (in Summe 54) und alle „semantischen Typen“ (in Summe 135) definiert und auflistet. In der informatischen Terminologie wird hier von Links zwischen den Knoten gesprochen.

Zu beachten ist dabei, dass, wie jetzt vielleicht schon erwartet, dieses „Semantic Network“ wesentliche Quelle für die informatische, oder besser gesagt mathematische Umsetzung ist und somit die Basis für eine maschinelle Umsetzung bildet, ganz im Sinne des Berners-Lee'schen Postulats (s. oben).

Das sog. "**Spezialist Lexicon**" ist quasi ein grammatisches Wörterbuch, vermehrt um medizinisches Vokabular. Es hilft dem sog. Natural Language Processing (NLP) System zum Beispiel grammatikalische Beugungen, wie sie in Deklinationen und Konjugationen vorkommen, korrekt zu interpretieren. Weiters werden Komposita und Variationen der Wortfolge zusammengeführt.

Beim letzten Teil „**MetaMorphoSys**“ handelt es sich um eine Software-Oberfläche, mit der die anderen drei Teile konfiguriert bzw. bedient werden können, entsprechend der verwendeten UMLS-Lizenz. Für jene, die mit dem Braunschweiger Allegro vertraut sind: Es handelt es sich hier um eine Art Cockpit-Funktionalität.

Das UMLS bildet heute das lexikalische und terminologische Rückgrat des Internetportals PubMed. Dabei möge die geschätzte Leserin beachten, dass es sich hier nicht allein um die Datenbank MedLine handelt, sondern um ca. 30 verschiedene Datenbanken, die alle durch dieses Portal gemeinsam zugänglich und durchsuchbar sind.

Das UMLS ist darüber hinaus ein frei lizenzierbares Werkzeug, das verschiedenen Communities und Entwicklerinnen offen steht. Das wird genutzt, um andere medizinische Informationssysteme mit einer ständig weiterentwickelten Ontologie zu unterstützen.

Mit „MetaMorphoSys“ sind Entwicklerinnen in der Lage, gezielt auf jene Teile des UMLS zuzugreifen, die für das jeweilige Projekt relevant sind, unter Auslassung jener Teile, die nicht benötigt oder nicht lizenziert wurden.

Wer sich ausführlicher mit dem UMLS beschäftigen möchte - und dies sei jeder geraten die noch fünf oder mehr Jahre im Metier tätig sein wird - hat dazu jederzeit die Möglichkeit unter <http://www.nlm.nih.gov/research/umls/>.

Eine gute Übersicht über den aktuellen Status von UMLS bietet auch <http://www.openclinical.org/medTermUmls.html> (letzter Zugriff 7. 12. 2006).

Weitere Beispiele für medizinische Ontologien:

- MGT - Medical Guideline Tool: Auf Basis von Texten in der Patientendokumentation werden dazupassende Guideline-Textteile angezeigt.
- PATMAN: ontologiegestützte semantische Abfrage von Guidelines.
- OncoTerm: ontologiegestützte Übersetzung.

Zusammenfassung

Ontologien und andere semantische Technologien haben bei PubMed, der weltgrößten medizinischen Literaturdatenbank, den Thesaurus als terminologisches Kontrollwerkzeug abgelöst. Nur mehr die Oberfläche erscheint in thesaurusartiger Form, während unter dem Deckmantel des MeSH bereits eine voll funktional arbeitende Ontologie arbeitet.

Wie sich aus der Fülle der mehr als 100 verwendeten Quellen leicht ablesen lässt, bietet eine solch mächtige Ontologie ein breites Instrumentarium zur Erleichterung und Verbesserung der Literatursuche (sog. selbstlernende Begriffsaugmentierung): Größerer Recall, höhere Precision und, beinahe greifbar, die natürlichsprachige Abfrage von Retrieval-Systemen.

Diese neuen Möglichkeiten werden, zusammen mit etablierten Retrievaltechniken, in den verschiedensten Ausprägungen und in verschiedenen proprietären Lösungen in der Industrie und in anderen Wirtschaftszweigen bereits mit Erfolg eingesetzt.

Ein öffentlicher und weitgehend kostenfreier Einsatz ist absehbar. Ferner ist zu erwarten, dass Ontologien und andere Klassifikationen mit höherer semantischer Mächtigkeit zunehmend in der medizinischen Informationspraxis eingesetzt werden. Es ist auch anzunehmen, dass Medizinbibliothekarinnen und Medizindokumentarinnen als sog. Domänenexpertinnen, beim weiteren Aufbau von Domänen-Ontologien bzw. der semantischen Anreicherung bestehender Klassifikationen eine wesentliche Rolle spielen werden.

Der Autor hofft, dass die Ausführungen in diesem Artikel vielen Kolleginnen helfen werden einen Einstieg in dieses wichtige Thema zu finden, und sie anregt, sich weiter mit diesem Bereich zu beschäftigen.

Weiterführende Literatur

Es gibt mittlerweile eine enorme Fülle an Fachartikeln zum Thema Ontologien, von denen allerdings die meisten durch die informatische Terminologie an Verständlichkeit leiden. Ich habe mich daher im folgenden bemüht, einige Quellen aufzulisten, die den BID-Kolleginnen einen geeigneten Einstieg bieten.

Zur Erstinformation bieten sich die Artikel in der deutschen und englischen Wikipedia an.

http://de.wikipedia.org/wiki/Semantisches_Netz (letzter Zugriff 11.9.2006)

http://de.wikipedia.org/wiki/Topic_Maps (letzter Zugriff 25.9.2006)

http://de.wikipedia.org/wiki/Semantic_Web (letzter Zugriff 20.6.2006)

<http://de.wikipedia.org/wiki/Semantik> (letzter Zugriff 20.6.2006)

http://de.wikipedia.org/wiki/Ontologie_%28Informatik%29 (letzter Zugriff 18.5.2006)

[http://en.wikipedia.org/wiki/Ontology_\(computer_science\)](http://en.wikipedia.org/wiki/Ontology_(computer_science)) (letzter Zugriff 23.6.2006)

<http://de.wikipedia.org/wiki/Thesaurus> (letzter Zugriff 20.6.2006)

Die Beschäftigung mit Ontologien aus der bibliothekarisch-dokumentarischen Praxis heraus führt zwangsläufig zur Auseinandersetzung mit den aktuellen Entwicklungen in der Thesaurusforschung. Der Thesaurus war ja ursprünglich nicht für die maschinelle Prozessierung entwickelt worden, sondern als Dokumentationsprache mit relativ schwachen semantischen Verknüpfungen (vgl. dazu Mazzocchi F., Plini P.: Refining thesaurus relational structure: implications and opportunities. Abstract available from <http://www.bonn.iz-soz.de/wiss-org/wissorg06/Mazzocchi.htm> (letzter Zugriff 20.6.2006)). Eine, in verschiedenen Arbeiten aus dem BibDok-Bereich diskutierte Möglichkeit den Thesaurus in ontologische Richtung weiterzuentwickeln, bietet sich durch die sog. Aspektierung an (vgl. dazu: [9])

ad Ontologien:

Die von Harald Sack (sack@minet.uni-jena.de) von der Uni-Jena stammende Präsentation zum Thema Ontologien vom Januar 2006 ist zwar sehr ausführlich, aber dafür eine recht umfassende und instruktive Einführung in das Thema Ontologie:

Was sind und zu welchem Ende studieren wir Ontologien. Grundlagen und Anwendungen in der Informatik. Available from <http://www.informatik.uni-jena.de/~sack/Material/Ontologien.pdf> (letzter Zugriff 7.12.2006).

Ein für BibDok-Kolleginnen gut lesbarer Text ist:

Garshol LM.: Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all. Proceedings by deepX.Ltd. Available from <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html> (letzter Zugriff 7.12.2006)

ad UMLS: [10], [11].

BID bezogene Aufsätze: [12], [13], [14], [15].

Schlussbemerkung

Die größte, am Anfang zu bewältigende Schwierigkeit für Mitglieder der BID-Community beim Einstieg in das Thema semantische Technologien/Ontologien besteht darin, sich mit einer rein informatischen Terminologie herumschlagen zu müssen, die die Begriffswelt der klassischen Klassifikation weitgehend dupliziert und die Einsteigerin zwingt, beide Begriffswelten für sich zu synonymisieren. Dazu kommt, dass die Informatikerinnen, die das semantische Problem der Sprache und des Web lösen wollen, in ihren Textformaten und in ihrer eigenen Wortwahl wenig diszipliniert sind. Hat man diese Schwierigkeit einmal hinter sich gebracht, lüftet sich schnell der Schleier des Obskuren zugunsten der Verständlichkeit.

Diese Kritik an der informatischen Ontology-Community sollte nicht den Eindruck einer Geringschätzung erwecken und die geschätzte Leserin dazu verleiten auf die Informatik herunterzuschauen. Dafür schafft die Informatik in der realen Welt zu viele Fakten, mit denen wir uns weiter werden auseinandersetzen müssen.

Eine kritische, aber durchaus wissbegierige Neugier ist aber imho ein guter Weg sich diesen Fakten produktiv zu stellen.

Literatur

1. Beats Biblionetz: Begriffe Ontologie: <http://beat.doebe.li/bibliothek/w00085.html>
2. Gruber TR. A translation approach to portable ontologies. Knowledge Acquisition. 1993;5(2):199-220.
3. Neches R et al. Enabling technology for knowledge sharing. AI Magazine 1991;12(3):36-56.
4. Durnbill E. Berners-Lee and the Semantic Web Vision. Published on XML.com. Available from <http://www.xml.com/lpt/a/2000/12/xml2000/timbl.html> (Letzter Zugriff am 20.6.2006).
5. Noy NF, McGuinness D. Ontology Development 101: A guide to creating your first ontology. 2001. Available from <http://www.ksi.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html>
6. Schmitz-Esser W, Sigel A. Introducing Terminology-based Ontologies. Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" at the 9th International Conference of the International Society for Knowledge Organization (ISKO). Vienna, Austria, July 6th, 2006. 130 pages. Published electronically on E-LIS (E-prints in Library and Information Science, <http://eprints.rclis.org>), 2006-07-14.
7. Humphreys BL. UMLS Progress Report. First Versions of Metathesaurus™ and Semantic network Announced. National Library of Medicine News, ISSN 0027-965X 1990;45(9-10):10-1.
8. Smith B. Beyond Concepts: Ontology as Reality Representation in: Varzi A, Vieu L (eds.). Proceedings of FOIS 2004 International Conference on Formal Ontology and Information Systems. Turin, 4-6 Nov. 2004.
9. Zimmermann H. Aspektierung von Thesaurus Relationen, Öffnung in universale Anwendbarkeit? In: Wolfram Neubauer (Hrsg.). Qualität und Information. Jena: Deutscher Dokumentartag; 1993. p. 275-90.
10. McCray AT. An upper level ontology for the biomedical domain. Comparative and Functional Genomics. 2003;4:80-4.

11. Hunter L, Cohen BK. Biomedical Language Processing. What's beyond PubMed? *Molecular Cell*. 2006;21:589-94.
12. Knorsz G, Rein B. Semantische Suche in einer Hochschulontologie *Information. Wissenschaft & Praxis (NfD)*. 2005;56(5-6):281-90.
13. Beier H. Vom Wort zum Wissen. Semantische Netze als Mittel gegen die Informationsflut. *Information Wissenschaft & Praxis (NfD)*. 2004;55(3):133-8.
14. Fischer DH. Ein Lehrbeispiel für eine Ontologie: OpenCyc. *Information Wissenschaft & Praxis (NfD)*. 2004;55(3):139-42.
15. Smith B, Siebert D, Ceusters W. Was die philosophische Ontologie zur biomedizinischen Informatik beitragen kann. *Information Wissenschaft & Praxis (NfD)*. 2004;55(3):143-6.

Bitte zitieren als

Cazan C. *Medizinische Ontologien: das Ende des MeSH*. *GMS Med Bibl Inf*. 2006;6(3):Doc31.

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/mbi/2006-6/mbi000049.shtml>

Veröffentlicht: 28.12.2006

Copyright

©2006 Cazan. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.

Korrespondenzadresse:

Constantin Cazan
Postfach 104, A-1238 Wien, Österreich
cazanc@gmx.at