

Wissensrepräsentation und kognitive Modelle

Ipke Wachsmuth
Technische Fakultät
Universität Bielefeld
D-4800 Bielefeld 1
ipke@techfak.uni-bielefeld.de

Zusammenfassung.

Unter einer Wissensrepräsentation versteht man ein symbolisch dargestelltes Modell eines Wissensbereichs aus Objekten, Fakten und Regeln in operationaler Form für einen Handlungsträger mit symbolverarbeitender Kompetenz. Seit den Aufbrüchen der Künstlichen Intelligenz gilt als wichtigstes Prinzip der Wissensrepräsentation, daß Wissen über Fakten und Relationen in einem Weltausschnitt explizit (deklarativ) kodiert werden muß, so daß interpretative Prozeduren es verarbeiten können. Dabei geht es nicht nur um die Frage, wie sich Wissen formal repräsentieren und manipulieren läßt, sondern auch darum, wie ein Formalismus zu benutzen ist, um bestimmte Wissensinhalte und Verarbeitungsstrategien auszudrücken.

Die Aufgabenstellung für das Gebiet Wissensrepräsentation umfaßt demnach zweierlei: (1) Die Entwicklung von Repräsentationsformalimen, die Beschreibungsmittel für die symbolische Darstellung des Wissens bestimmter Fachgebiete und Weltbereiche bereitstellen und die automatische aufgabengesteuerte Berechnung von Inferenzen über der Menge kodierten Wissens ermöglichen; (2) die Modellierung von Wissensbereichen, bei der insbesondere folgende Fragen zu beantworten sind: *Welche* Kategorien sind zur Einordnung von Objekten und Ereignissen eines Wissensbereichs zu wählen? *Welche* Eigenschaften und Annahmen sind ihnen zuzuordnen? *Wie* werden sie untereinander in Beziehung gesetzt? *Welche* Folgerungen sollen im Kontext bestimmter Annahmen möglich sein?

Der erste Gesichtspunkt, die Entwicklung von Repräsentationsformalimen, wird mittlerweile sehr weitgehend und grundsätzlich untersucht, und es zeichnen sich Bestrebungen zur Standardisierung ab. Die Fragen, mit denen sich die Entwerfer von Repräsentationsformalimen beschäftigen, beziehen sich u.a. auf die Ausdrucksfähigkeit und die Komplexität von Repräsentationen, aber auch auf ihre prädikatenlogische Rekonstruktion und Spezifikation. Im Rahmen einer DARPA-Initiative hat sich u.a. eine "Knowledge Representation System Specification Group" formiert, die sich mit der Vereinheitlichung von Wissensrepräsentationssystemen der KL-ONE-Familie, den sogenannten terminologischen Repräsentationssystemen, befaßt.

Die Aufgabenstellung der Modellierung, die im Mittelpunkt dieses Beitrages stehen soll, hat sich in der Wissensrepräsentation als ausgesprochen schwierig erwiesen. Das hat auf der einen Seite sicherlich damit zu tun, daß beim Entwurf eines wissensbasierten Systems nicht nur betreffs des Gesamtsystems, sondern vor allem hinsichtlich der Wissensbasis ein kreativer Akt zu vollziehen ist, der sich auf die individuelle Behandlung der oben unter (2) genannten Fragen richtet. Vor allem aber läßt sich diese Aufgabe der Wissensmodellierung nicht grundsätzlich (etwa im Sinne eines GPS) lösen, da die erwähnten Tätigkeiten bei jedem neuen System und für jede weitere Domäne zu erbringen sind.

Deshalb stellt sich, vor allem angesichts des wachsenden Umfangs projektierter wissensbasierter Systeme, zunehmend stark die Frage nach einer Wiederverwendbarkeit schon existierender Wissensbasen bzw. nach einer Aggregation großer Wissensbasen aus bibliotheksmäßig gesammelten oder inkrementell entwickelten Teilen. Am pointiertesten ist dieser Anspruch in jüngster Zeit von Neches, Fikes et al. formuliert worden ("Knowledge Sharing"), auf der anderen Seite mit Ansätzen zur Modularisierung wissensbasierter Systeme, wobei es sich in beiden Fällen eher um Forschungsprogramme als um schon existierende und elaborierte Vorgehensweisen handelt.

Wegen der Verschiedenheit derzeit verwendeter KI-Sprachen, Implementationen und Werkzeuge scheint es gegenwärtig - und möglicherweise grundsätzlich - am aussichtsreichsten, die formulierten Aufgaben auf einer Ebene von abstrakten Wissensmodellen anzugehen. Solche Modelle können als "high level" Spezifikationen von wissensbasierten Systemen verstanden werden und bieten für die modulare Wissensbasisentwicklung und Wiederverwendung von Wissensbasen (durch Reimplementation) nach derzeitiger Einschätzung bessere Ausgangsbedingungen als der Austausch konkreter Wissensbasen über ein sog. "Knowledge Interchange Format", wie es etwa in der Interlingua-Gruppe der DARPA-Initiative versucht wird.

Ansätze allgemeiner Vorgehensweisen des Entwurfs wissensbasierter Systeme richten sich zunehmend stark auf die von Newell Anfang der 80er Jahre proklamierte Wissensebene, die von der Ebene der symbolischen Verarbeitung unterschieden wird. Hiermit verbindet sich der Anspruch, die Wissensinhalte und ihre Funktion für einen Systemzweck ins Zentrum der Modellierungstätigkeit zu stellen und zu abstrahieren von der Form der symbolischen Darstellung des Wissens und den symbolverarbeitenden Prozeduren, die die Systemfunktionalität zustande bringt. Dieser Gedanke entspricht in etwa der Erstellung konzeptueller Modelle in der Datenbank-Entwicklung.

Vorangetrieben wurde diese Entwicklung vornehmlich im Kontext des Entwurfs von Expertensystemen. So wird etwa im KADS-Ansatz im Prozeß der Wissensanalyse und -interpretation vom Knowledge Engineer ein mentales konzeptuelles Modell formuliert, welches sowohl den Problemlöseprozeß als auch Strukturen des Bereichswissens rekonstruiert und das - zusammen mit Anforderungen an ein zu erstellendes wissensbasiertes System - Ausgangspunkt für das Design des Systems ist. Die Konstruktion von Wissensmodellen wird durch formal beschriebene, katalogisierte Interpretationsmodelle unterstützt. Sofern sich eine Problemlöseaufgabe mit Hilfe eines Interpretationsmodells charakterisieren läßt, kann beim Systementwurf auf abstrakte Designelemente (ähnlich den "generic tasks" von Chandrasekaran) zurückgegriffen werden.

Diese Vorgehensweise des Expertensystementwurfs richtet sich vornehmlich auf die Erfassung und Modellierung der abstrakten Problemlöseaufgabe. Die Repräsentation des Bereichswissens wird stark von der Natur des Problems und der auf das Wissen angewandten Inferenzstrategie beeinflusst (wie von Bylander und Chandrasekaran als "Interaktionsthese" formuliert wurde); das modellierte Wissen ist auf die Lösung der Fachaufgaben zugeschnitten. Das entworfene Domänenmodell spiegelt mehr oder weniger stark spezifische Wissensinhalte und individuelle Vorgehensweisen des (in der Regel einen) Experten wider, mit dem das Modell in Wissensakquisitionssitzungen expliziert und ausgehandelt wird. Typischerweise ist der Gegenstandsbereich des Systems ein eng umrissenes Spezialgebiet, in dem hohes Potential an spezifischer Problemlösefähigkeit in einem weitgehend vorab festgelegten Verwendungsrahmen verlangt ist.

Grundsätzlich anders verhält sich die Situation allerdings bei der Wissensrepräsentation für natürlich-sprachliche Systeme. Die Vorgehensweisen der Expertensystemtechnik beim Aufbau einer Wissensrepräsentation lassen sich dort schwerlich übernehmen. Beim Textverstehen etwa handelt es sich um die Modellierung einer allgemeinen menschlichen Fähigkeit mit einer von vornherein nicht festgelegten (Problemlöse-) Zielsetzung. Da bedingt durch die inferentielle, zielorientierte Verarbeitungsstrategie ein wissensbasiertes System nach Kriterien einer zu lösenden Aufgabe entworfen werden muß, erhält die Formulierung einer Rahmenbeschreibung ("Szenario"), in der nicht nur ein einschränkender Gegenstandsbereich, sondern auch eine funktionale Abgrenzung festgelegt wird, besonderes Gewicht; dabei können exemplarische Texte und Frage-Antwort-Dialoge, mit denen die Verstehensleistung demonstriert werden soll, vorgegeben werden.

Bei der Entwicklung von Systemen, die zur semantischen Verarbeitung von natürlicher Sprache fähig sind, geht es zentral um die Identifikation und Modellierung intersubjektivierbarer Bestände an Welt- oder Hintergrundwissen. Auch wenn mit dem System nicht allgemeine, sondern auf Aufgabe und Gegenstandsbereich eingeschränkte menschliche Sprachfähigkeit modelliert werden soll, erhält die Modellierung von Alltagswissen einen wesentlich höheren Stellenwert als bei der Entwicklung von Expertensystemen. Es muß davon ausgegangen werden, daß das modellierte Wissen nicht nur formalen Eigenschaften natürlicher Sprache, sondern auch den damit verbundenen psychologischen und kognitiven menschlichen Verhaltensweisen gerecht werden muß. Diese lassen sich nun nicht einfach durch eine Wissensakquisition mit einem "typischen" Vertreter einer Sprecher/Hörer-Gemeinschaft erschließen.

Der gegenwärtige, durchweg als unbefriedigend empfundene Zustand ist folgender: Dem Modellierungspersonal obliegt - in noch weitaus größerem Maße als bei der Expertensystementwicklung - die Gestaltung der Struktur und der Inhalte der Wissensbasen, etwa die Auswahl der ontologischen Kategorien, die in der Regel ad hoc und für Außenstehende schwer nachvollziehbar durchgeführt wird. Erschwert wird die Aufgabe durch den Reichtum des Begriffsuniversums, dem natürlich-sprachliche Systeme im Vergleich zu Expertensystemen Rechnung tragen müssen. Richten sich die systematischen Ansätze im Bereich Expertensysteme vornehmlich auf Strukturen und Typentaxonomien von Problemlösungsaufgaben ("Problemlöseontologien"), so erfuhren die Strukturen und die systematische Betrachtung von Ontologien für Domänenwissensbasen bislang wenig Aufmerksamkeit.

Die systematische Untersuchung repräsentierbarer kognitiver Modelle menschlicher Welt- und Aufgabenwahrnehmung, wie sie etwa in Arbeiten von Dahlgren und im CYC-Projekt versucht wird, steht erst am Anfang. Die Beziehung zwischen Domänenmodellen und Problemlösewissen von "Agenten" wird eine wichtige Aufgabenstellung werden. War es noch bis Ende der 80er Jahre ein überwiegend akzeptiertes Faktum, daß die Inferenzstrategie die Wissensrepräsentation diktiert, zeichnet sich zunehmend stark ab, daß es nicht nur darum geht, generische Problemlöseaufgaben zu betrachten, um den Entwurf wissensbasierter Systeme bei der Wissensrepräsentation zu systematisieren. Auch das Fakten- und Relationengefüge von diversen Domänen und die Untersuchung von Strukturen der Wissensmodelle des Menschen (wie etwa gestaffelte generische und bereichsbezogene Ontologien) bedürfen der Untersuchung. Die Unabhängigkeitsthese des Stanford Heuristic Programming Project ("Die Form der Darstellung von Bereichswissen ist unabhängig von seinem Gebrauch"), hat sich aus erkennbaren Gründen nicht halten lassen, jedoch wird es bei zukünftiger Forschung auch darum gehen, inwieweit sich applikationsunabhängige und -abhängige Wissensrepräsentationen unterscheiden und vereinbaren lassen.