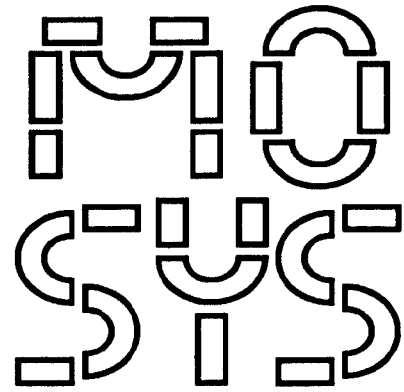


**Universität Bielefeld  
Technische Fakultät (i.A.)  
Postfach 8640  
D-4800 Bielefeld 1**



## **Modularisierung Wissensbasierter Systeme**

**Ipke Wachsmuth**

# MODULARISIERUNG WISSENSBASIERTER SYSTEME

- MOSYS -

## Rahmenentwurf für ein Forschungsprogramm im Bereich Wissensbasierte Systeme/Künstliche Intelligenz\*

Ipke Wachsmuth

### Übergeordnetes Ziel

Zukünftige Wissensbasierte Systeme sollen dem Menschen bei der Bewältigung und Nutzung zunehmend unüberschaubarer Informationsressourcen helfen. Der vorhersehbare und sich bereits abzeichnende Umfang solcher Systeme erfordert besondere Maßnahmen bei der Systementwicklung, die gleichermaßen die *technische* wie die *arbeitsmäßige* Realisierung betreffen.

Übergeordnetes Ziel aller Projekte ist daher die Entwicklung von Konzepten und Methoden für die Akquisition, Verwaltung und Verarbeitung komplexer Bereichswissensbestände (komplex = groß und divers). Im Hinblick auf zukünftige Entwicklungen soll die Möglichkeit natürlichsprachlicher Informationsaufnahme und -verarbeitung von vornherein einbezogen werden.

In Analogie zu Techniken, die heute bei der Bewältigung großer Softwareprojekte Einsatz finden, soll zur Erreichung dieses Ziels eine *Modularisierung Wissensbasierter Systeme* als programmatischer Rahmen und Leitthema für Forschungsaktivitäten verfolgt werden.

---

\*Das vorliegende Papier skizziert einen Rahmen für konkrete Projekte, die in den Bereichen fortgeschrittener Expertensystem-Technologie sowie wissensbasierter Sprachverarbeitung angesiedelt werden sollen. Es ist eine überarbeitete, aktualisierte Fassung der ersten Niederschrift eines Forschungsprogramms vom November 1988.

## 1. Gegenstand

Als *wissensverarbeitendes System* wird im folgenden ein System bezeichnet, das über manipulierbare Modelle bestimmter Aspekte der Realität verfügt. Solche Modelle liegen in Form symbolischer Repräsentationen von Objekten, Fakten und Regeln vor, auf denen ein symbolverarbeitender Interpret entsprechend übertragener Aufträge zweckgerichtet operiert.

Von einem *natürlichsprachlichen System* soll gesprochen werden, wenn eine Teilmenge der vom System akzeptierten bzw. generierten Nachrichten natürlichsprachlich kodiert ist und zur Verarbeitung der Nachrichten syntaktische und semantisch-pragmatische Verfahren eingesetzt werden, z.B. um Bedeutung aus Texten zu entnehmen, Fragen zu beantworten oder aufgabenorientierte Dialoge zu führen.

Während man in konventionellen Programmen üblicherweise Algorithmen und Daten unterscheidet, folgt die programmiertechnische Realisierung wissensverarbeitender und natürlichsprachlicher Systeme dem *wissensbasierten Ansatz*. Hier wird die Trennung von (Bereichs-)wissen, Verarbeitungsstrategie (Problemlösungsstrategie, Kontrollstrategie) und Daten vorgesehen. Als prinzipielle Vorgehensweise wird ein Anwendungsbereich in einer formalen Sprache axiomatisiert und dann "intelligentes" Verhalten eines Systems dadurch produziert, daß mit maschinenausführbaren Schlußmechanismen Konsequenzen des in der formalen Sprache repräsentierten Wissens abgeleitet werden. Die Leistung eines solchen Systems geht also über ein schlichtes Auffinden abgespeicherter Daten hinaus. Aus diesem Grund ist der Ausdruck *Wissensbasis* für die gespeicherte Information und für ein derart aufgebautes System der Ausdruck *Wissensbasiertes System* gängig.

Eine Wissensbasis kann grundsätzlich als Modell eines Weltausschnitts (z.B. des Wissens eines Experten über ein Fachgebiet) aufgefaßt werden. Sie besteht aus einer Menge persistent gespeicherter Ausdrücke einer formalen (Repräsentations-)sprache, die von der Verarbeitungsinstanz eines Wissensbasierten Systems manipuliert werden können. Ein Problem, mit dem sich Systementwickler zunehmend konfrontiert sehen, besteht darin, daß die Menge der Ausdrücke einer Wissensbasis verschiedenste Teilaspekte betreffen kann und sehr schnell groß und unüberschaubar wird. Zudem verändert sich diese Menge während der Lebensdauer eines Wissensbasierten Systems ständig. Deshalb sind besondere Maßnahmen bei der Systementwicklung erforderlich, die gleichermaßen die *technische* wie die *arbeitsmäßige* Realisierung betreffen. Solche Maßnahmen werden im vorliegenden Papier unter dem Schlagwort "Modularisierung" diskutiert.

Unter einer *Modularisierung eines Wissensbasierten Systems* soll im folgenden die Segmentierung einer Wissensbasis in Teile verstanden werden,

- die für fest definierte Aufgabenbereiche zuständig sind
- die über spezifische Wissensrepräsentationen und Kontrollstrategien verfügen können
- deren Inhalt gegenüber anderen Teilen verborgen sein kann
- die unabhängig voneinander entwickelt werden können
- auf die Änderungsmaßnahmen, Fehlersuche, Konsistenzprüfungen beschränkt werden
- die für sich genommen überschaubar sind und die Aggregation komplexer Wissenssysteme ermöglichen.

Hierbei sei der Fall eingeschlossen, daß spezielle Teile einer Wissensbasis *Kontrollwissen* zur Ablaufsteuerung und Ausführung generischer Problemlösungsaufgaben enthalten.

## 2. Zum Begriff der Modularisierung

Der Begriff der *Modularisierung* ist in der Informatik ursprünglich im Kontext des Software Engineering eingeführt worden. Unter Software Engineering versteht man "die Anwendung von Prinzipien, Fähigkeiten und Kunstfertigkeiten auf den Entwurf und die Erstellung von Programmen und Systemen von Programmen" [Dennis75]. Ansätze einer systematischen, ingenieurmäßigen Herstellung von großer Software entstanden Ende der 60er Jahre als Antwort auf die Softwarekrise und leiteten eine Wende vom Individualprogrammieren mit ad-hoc-Lösungen zu einem planvollen, kooperativen Teamprogrammieren ein. Zentralen Stellenwert erhielt die Spezifikation und Dokumentation von Programmen.

Parnas [Parnas75] beschreibt Software Engineering als Programmierung unter mindestens einer der folgenden beiden Bedingungen: (1) mehr als eine Person ist befaßt mit der Erstellung und/oder dem Gebrauch des Programms und (2) mehr als eine Version des Programms wird erstellt werden. Es ist das Ziel, "ökonomisch Software zu erhalten, die zuverlässig ist und effizient auf realen Maschinen arbeitet" [Bauer75]. Die Herstellung großer Software unterscheidet sich qualitativ von der Herstellung kleiner Software: Zentraler Punkt der Programmierung auf allen Ebenen vom Entwurf bis zur Implementierung ist die *Bewältigung von Komplexität*. Ein integraler Bestandteil der Programmierung im Großen ist die geregelte Zusammenarbeit zwischen den am Softwareherstellungs-

prozeß Beteiligten [Kimm/Koch/Simonsmeier/Tontsch79]. Bestrebungen, Regeln der Kunst für die Erstellung großer Software anzugeben, schlugen sich in verschiedenen Phasenmodellen der Softwareentwicklung und -verwendung nieder ("Software Life Cycle").

Die an ein Softwareprodukt gestellten qualitativen Anforderungen betreffen:

- Korrektheit, Lesbarkeit
- Adaptabilität, Portabilität
- Robustheit, Ausfallsicherheit, Benutzungsfreundlichkeit
- Effizienz

Der hauptsächliche Sinn der Modularisierung großer Programmsysteme besteht also darin, daß die von einem Bearbeiter zu jedem Zeitpunkt zu bewältigende Komplexität möglichst gering gehalten werden soll (Arbeitsaufteilung). Da keine allgemein verbindlichen einschränkenden Programmiernormen existieren, läßt sich der Begriff des Moduls nicht genau definieren. Als *Moduln* bezeichnet man bestimmte Programmteile, die unter ganz bestimmten Gesichtspunkten eine logische Einheit bilden. Im allgemeinen bestehen Moduln aus Daten und aus Funktionen, die auf diesen Daten operieren (intuitiv: "Baustein"; insbesondere ist jedes vollständige Programm ein Modul.)

Modularisierungen werden dementsprechend nach den folgenden Kriterien beurteilt [Kimm et al.79]:

- *Unabhängige Entwickelbarkeit/Änderbarkeit:* Moduln sollen weitestgehend unabhängig voneinander geschrieben werden können. Revisionen innerhalb eines Systems sollen auf voneinander unabhängige Teile beschränkt bleiben. Gefordert wird deshalb die getrennte Übersetzbarkeit von Moduln ("physikalische Modularität"), die einerseits die Unabhängigkeit der Entwickelbarkeit durch verschiedene Programmierer, andererseits die Austauschbarkeit von Moduln im fertigen Programm unterstützt.
- *Verständlichkeit:* Sowohl das Gesamtsystem wie auch die einzelnen Moduln sollen einfach erfaßbar sein. Eine geringe Größe der einzelnen Moduln soll die Übersichtlichkeit für den Ersteller gewährleisten. Eine weitere Forderung ist die Einfachheit und Übersichtlichkeit der Schnittstellen (Schnittstelle zweier Moduln = Menge aller Programmteile und Daten, die beiden Moduln bekannt sein sollen).

Ein wesentlicher Gesichtspunkt bei der Zerlegung eines Systems in Moduln ist es deshalb, die Realisierung wesentlicher Datenstrukturen des Programms von ihrer Benutzung zu isolieren (abstrakte Datenstrukturen bzw. -typen). Zentral ist der Gedanke des *information hiding*, d.h. des Verbergens von Entwurfsentscheidungen vor anderen Programmteilen. Zerlegungen sollen so gewählt werden, daß Entwurfsentscheidungen möglichst in einzelnen Moduln isoliert werden, besonders solche, die Informationsdarstellungen betreffen. Bei der Zerlegung in Moduln spielt daher nicht die Festlegung von Abarbeitungsreihenfolgen, sondern die Aufteilung nach *Zuständigkeiten* eine tragende Rolle [Parnas72].

Eine geringe Komplexität der intermodularen Beziehungen wird als notwendige Voraussetzung dafür gesehen, das wesentliche Entwurfsziel zu erreichen, nämlich die Komplexität jedes Lösungsschritts so klein wie möglich zu halten. Die Beziehungen zwischen Moduln sollten daher so klar und einfach wie möglich gehalten werden. Die wichtigste Beziehung ist die Benutzt-Beziehung, die im allgemeinen eine Hierarchie bilden sollte.

### 3. Bezug zum Gegenstand des Forschungsprogramms

Wissensbasierte Systeme sind - auf der Ebene ihrer Kodierung betrachtet - tendenziell große Softwarekonstrukte. Bei ihrer Realisierung spielen Gesichtspunkte des Software Engineering eine zunehmend große Rolle. Es ist mittlerweile ein Anliegen der Anbieter von Softwaresystemen, die Durchführung größerer Implementierungsvorhaben durch geeignete Entwicklungswerkzeuge zu unterstützen. Dies ist jedoch nicht die Ebene, auf der Modularisierungen Wissensbasierter Systeme betrachtet werden sollen, denn implementierungsnahe Entwürfe erweisen sich hier als von zu geringem Abstraktionsgrad.

Grundlagenforschung wird deswegen für erforderlich gehalten, um Zielsetzungen der Modularisierung auf eine höhere Abstraktionsebene, nämlich des *Designs* von Wissensbasierten Systemen zu übertragen. Es handelt sich also nicht um die Ebene der symbolischen Verarbeitung von Information in Programmen, sondern zentral um die *Wissensebene* als die der Denkweise von Entwerfern und Nutzern solcher Systeme angepaßte Ebene [Newell82]. Wissen wird dabei als eine *Verhaltenskapazität* aufgefaßt, die nicht davon abhängt, mit welchen Datenstrukturen das Verhalten eines intelligenten Systems auf Symbolebene repräsentiert wird. Es muß deshalb das grundsätzliche Ziel sein, die Repräsentation von Fachwissensbeständen bereits in der Entwurfsspezifikation

von der Ablaufsteuerung zum zweckgerichteten Einsatz des Wissens zu trennen. Die software-technische Realisierung ist dann vorgezeichnet.

Die Leitziele des Modularisierungsgedankens beim Software Engineering: (1) unabhängige Entwickelbarkeit und Änderbarkeit und (2) Verständlichkeit können grundsätzlich ebenfalls als *Leitziele bei der Modularisierung Wissensbasierter Systeme* gelten. Wie deutlich geworden sein sollte, erfordern es diese Ziele zuallererst, den modularen Aufbau eines Wissenssystems, nicht eine software-technische Realisierung, im Auge zu haben. Einen zusätzlichen Stellenwert erhalten diese Ziele dadurch, daß nur auf dem Weg über eine Modularisierung die graduelle Überwindung der *Bereichsspezifität* (Beschränkung der Kompetenz auf enge Expertisegebiete bzw. Weltausschnitte) von Wissensbasierten Systemen - gegenwärtig als unüberwindliche Hürde akzeptiert - überhaupt denkbar erscheint.

#### 4. Abgrenzung des Forschungsumfeldes

Die Modularisierung Wissensbasierter Systeme berührt verschiedene Themenkomplexe und Forschungsfragen, die sich wechselseitig beeinflussen und nicht unabhängig voneinander betrachtet werden können. Sie betreffen einerseits die Entwicklung geeigneter Repräsentationssysteme, andererseits Maßnahmen auf der Seite der Bereichsmodellierung. Schließlich sind Ansätze zur Verarbeitung natürlicher Sprache und zur Wissensakquisition von vornherein einzubeziehen, da sie auf eine modulare Wissensauffassung zugeschnitten sein müssen.

---

### MODULARISIERUNG WISSENSBASIERTER SYSTEME

---

Repräsen-  
tionssysteme

Bereichs-  
modellierung

Modulari-  
sierung

Natürliche  
Sprache

Wissens-  
erwerb

---

## 4.1 Repräsentationssysteme

Die Aufgaben eines Repräsentationssystems lassen sich in diesem Kontext wie folgt umreißen:

- (1a) Verwahrungsort für symbolisch repräsentierte Bestände des Wissens spezieller Fachgebiete und Weltbereiche
- (1b) Verwahrungsort für kontextbezogene Wortbedeutungen (näheres hierzu siehe unter 4.3)
- (2) automatische aufgabengesteuerte Berechnung von Inferenzen über der Menge kodierten Wissens

Entsprechend den Ausführungen in Abschnitt 1 erfordert die Konzeption eines Repräsentationssystems die Spezifizierung einer formalen Repräsentationssprache, deren Konstrukte den darzustellenden Wissensentitäten angemessen sind (Ausdrucksfähigkeit) und die Anwendung maschinenausführbarer Schlußmechanismen erlauben.

Entsprechend den Bemerkungen im 1. und 3. Abschnitt muß die Repräsentationssprache ferner Konstrukte einschließen, die die Grundlage für eine Modularisierung bieten und die es insbesondere erlauben, den Geltungsbereich von Inferenzmöglichkeiten auf Teilwissensbestände einzuschränken.

Mit der Möglichkeit der Abschottung von Teilen einer Wissensbasis gegenüber der Zugriffsmöglichkeit durch die Verarbeitungsinstanz eines Wissensbasierten Systems wird schließlich eine *Steuerung des Wissenszugangs* erforderlich. Ein Repräsentationssystem muß daher Möglichkeiten zur Beschreibung und Regulierung von (statischen und dynamischen) Zugangsbedingungen enthalten.

## 4.2 Bereichsmodellierung

Neben der strukturellen Erfordernis eines Repräsentationssystems, das Wissen in symbolischer Form bereitstellen und zum Lösen von Aufgaben einsetzen kann, ergeben sich Anforderungen, wie zur Modellierung von Weltbereichen mit einem solchen Formalismus umzugehen ist ("Modellierungskunst"): *Welche* Kategorien werden zur Einordnung von Objekten und Ereignissen eines Expertise- bzw. Diskursbereichs



herangezogen und *welche* Eigenschaften/Annahmen werden ihnen zugeordnet; wie werden sie untereinander in Beziehung gesetzt, *welche* Folgerungen sollen im Kontext bestimmter Annahmen möglich sein etc.

Bereits die Modellierung enger Ausschnitte eines Weltbereichs stellt eine nicht zu unterschätzende Aufgabe dar, die neben guter Kenntnis des Bereichs auch Erkenntnisse aus Disziplinen, die sich mit dem Phänomen *Wissen* befassen (Sprachwissenschaft, Philosophie, Kognitionswissenschaft), einzubringen erfordert. Sollen verschiedene Weltbereiche in eine Wissensrepräsentation einbezogen werden, gelangt der Systementwickler schnell an Grenzen der Überschaubarkeit, die ein modulares Vorgehen geradezu zwingend erscheinen lassen. Ein noch weitgehend ungelöstes Problem ist ferner die Modellierung des von Menschen auch bei der Lösung fachspezifischer Aufgaben implizit herangezogenen *Alltagswissens* (Wissen über Raum und Zeit, über physikalische Zusammenhänge in erfahrbaren Phänomenbereichen, über Motive und Überzeugungen anderer etc.).

#### 4.3 Rolle der natürlichen Sprache

Die zukünftige Entwicklung wissensverarbeitender und natürlichsprachlicher Systeme kann nicht getrennt voneinander gesehen werden, und zwar aus folgenden Gründen:

1. Die maschinengerechte Aufbereitung textuellen Wissens kann nicht in beliebig großem Umfang durch den Menschen geleistet werden; hieraus folgt, daß Überlegungen für eine *natürlichsprachliche Wissensaufnahme* bei Konzeptionen für solche Systeme einzubeziehen sind.
2. Der Mensch als Nutzer verkehrt zu einem großen Teil in *natürlichsprachlichen Dialogen* mit solchen Systemen; das heißt, der Einbezug natürlicher Sprache dient auch einer menschengerechten Kommunikationsform mit Wissensbasierten Systemen und wird eine entscheidende Voraussetzung für ihre breite Akzeptanz sein.
3. Der Aufbau großer bereichsspezifischer Wissensbestände unabhängig vom Aufbau entsprechender "Bereichslexika" ist nicht realistisch; der Beitrag von Wortbedeutungen zu situations- und diskursgebundenen Inferenzprozessen und die entscheidende Rolle bestimmter (einzeln oder kumuliert auftretender) Inhaltswörter für den Wissenszugang erfordert insbesondere den Einbezug der lexikalischen Semantik.

Das Problem der Wissensrepräsentation stellt sich bei natürlichsprachlichen Systemen verschärft, da sie auf einen umfangreichen Wissensfundus zurückgreifen müssen, der aus vielen heterogen strukturierten Einheiten bestehen kann. Nicht nur bei der Wissensverarbeitung, sondern insbesondere bei der Verarbeitung natürlicher Sprache spielen partielle und kontextgebundene Spezifikationen eine wesentliche Rolle. Demnach muß ein Repräsentationssystem Wortbedeutungen unterdefinierend und nach Kontexten differenziert bereitstellen können.

Beim Entwurf von Wissenskomponenten für die Sprachverarbeitung sind insbesondere folgende Problempunkte zu berücksichtigen, vor deren Hintergrund Modularisierungsansätze diskutiert werden müssen:

- der Reichtum des Begriffsuniversums von Diskursbereichen
- der Umfang und die Diversität des Rahmen- und Hintergrundwissens
- die Nichtzielgerichtetheit des Wissens (Erfordernis einer kontextbezogenen Auswahl von Vorwärtsinferenzen)

#### 4.4 Wissensakquisition

Unter *Wissensakquisition* wird im Kontext Wissensbasierter Systeme die Erhebung des für einen Aufgabenbereich maßgeblichen Wissens von einer Wissensquelle (Experte, Fachtext) und die Abbildung auf eine dem Bereich angemessene Wissensrepräsentationssprache verstanden. Bei der Entwicklung wissensverarbeitender Systeme hat sich dieser Teil als der aufwendigste und schwierigste herausgestellt. Die zentrale Aufgabe liegt dabei derzeit bei sog. Wissensingenieuren, die zwischen dem Wissen, wie es Menschen (zum Teil implizit) zur Verfügung steht, und computergeeigneten Formalismen vermitteln müssen. Forschungs- und Entwicklungsansätze im Gebiet Wissensakquisition richten sich zunehmend auf die Bereitstellung von Werkzeugen, mit denen Experten Wissensbasen weitgehend selbständig aufbauen können, und auf die (Teil-)Automatisierung des Wissenserwerbs.

Bei jedem Entwurf eines Wissensakquisitionssystems muß ein aufgabenbereichsbezogenes Wissensmodell formuliert werden, das es gestattet, Wissen in "modulare" Bestandteile zu zergliedern, die mit der Verarbeitungsinstanz eines Systems in Beziehung gesetzt werden können. Häufig werden z.B. Fakten und Regeln (bedingte Aussagen, die Objekte oder Klassen von Objekten eines Weltausschnitts in Beziehung setzen), als modulare Bestandteile gewertet, deren Hinzufügung oder Wegnahme ohne weitere

Änderungen einer Wissensbasis möglich ist. Die einem (Teil-) Aufgabenbereich entsprechende Expertise schlägt sich jedoch i.a. in einer *Menge* von Fakten und Regeln nieder (vergleichbar den nichtlogischen Axiomen einer Theorie). Deshalb scheint es erforderlich, die Modularisierung eines Wissensbasierten Systems nicht auf der Ebene von Fakten und Regeln, sondern auf der Ebene von *Teilaufgabenbereichen* (reflektiert in Teilwissensbeständen und Aufgabenlösungstypen) anzugehen und Wissensakquisitionssysteme auf derartige Moduln zuzuschneiden. Erwünscht ist auch eine Modularisierung nach *Weltbereichen*, z.B. um weltbereichsspezifische Folgerungen ableiten zu können.

Ein grundsätzliches Problem bei der Erstellung großer Wissensbasen ist die Vermeidung von *Inkonsistenzen*. Die Konsistenz einer Wissensbasis ist aufgrund der Übernahme von Wissen aus externen, möglicherweise heterogenen Quellen keinesfalls gesichert. Entsprechende Prüfungen verursachen mit zunehmendem Umfang von Wissensbasen erheblichen Aufwand und werden daher häufig außer acht gelassen. Diese Tatsache ist insofern kritisch, als die Wissensverarbeitung in vielen Fällen auf logischen Schlußmethoden beruht, die Konsistenz voraussetzen müssen. Allerdings könnten - wie im Wissen des Menschen - Inkonsistenzen dort toleriert werden, wo Teile einer Wissensbasis voneinander unabhängige Fakten oder Vorgehensweisen betreffen, die nie im gleichen Kontext zum Tragen kommen. Zielsetzungen einer Modularisierung Wissensbasierter Systeme sollten daher "lokale" Konsistenzprüfungen im notwendigen Umfang einschließen; zu untersuchen ist die Frage, inwieweit sich der Aufwand für die Konsistenzüberwachung durch Modularisierungen reduzieren läßt.

Im Kontext der Wissensakquisition kommt Modularisierungen auch deshalb besondere Bedeutung zu, da sich nur auf diesem Wege die Erstellung von Wissensbasen durch mehr als eine Person kontrolliert angehen läßt. Ebenfalls dient ein modulares Vorgehen der Änderbarkeit, Handhabung von Versionen und der Verständlichkeit der Wissensarchitektur eines Gesamtsystems.

## 5. Vorarbeiten

Grundsätzliche Vorarbeiten wurden in den Jahren 1984-86 am Fachbereich Mathematik/Informatik der Universität Osnabrück in dem hausinternen Projekt LAKOS (*Logische Analyse kognitiver Organisations-Strukturen*) geleistet, welches in Kooperation des Verfassers mit der Forschungsgruppe Computerlinguistik am Fachbereich Sprachwissenschaft durchgeführt wurde. Diese Arbeiten waren von dem Ziel geleitet,

kognitiv realistische Theorien über Wissensverarbeitungsprozesse zu entwerfen, die bereits bei der Theoriebildung der Forderung nach effektiver Implementierbarkeit Rechnung zu tragen versuchen [Wachsmuth85]. Dabei wurden Konzeptionen für ein Modell der Wissensrepräsentation entwickelt, dessen wesentliche Aspekte im folgenden zusammengefaßt sind:

- Erklärungsmodell für hypostasierte kognitive Prozesse
- Rekonstruktion von Wissensstrukturen als parzellierte Axiomenmengen
- reproduktive Simulation von Dialogen aus diagnostischen Interviews
- kontextabhängige Analyse und Synthese von Ein- bzw. Ausgaben
- Erklärung von Aspekten der operationalen und der linguistischen Kompetenz des "rational handelnden" Individuums
- Forderung effektiver Implementierbarkeit

Hervorzuheben ist die empirische Basis dieser Arbeiten, die sich auf eine in den Jahren 1981-83 in den USA vom Verfasser et al. durchgeführte Langzeituntersuchung zum Wissenserwerb und -einsatz im Bereich Mathematik stützen. Diese Untersuchung wurde von der National Science Foundation im RISE-Programm (Research In Science Education) unter dem Grant No. SED 81-12643 gefördert.

Insgesamt wurden diese Forschungsarbeiten zur Konzeption eines formalen Modells intelligenter Wissensorganisation verwertet, das als Vorlage für den Entwurf künstlicher wissensverarbeitender Systeme dienen kann [Wachsmuth89]. Insbesondere wird damit auf die Strukturierung assertionaler Wissensbasen, den index- und inhaltsgesteuerten Zugang zu gespeichertem Wissen und die Behandlung von inkonsistentem Wissen eingegangen. Das Modell gestattet die formale Untersuchung von Eigenschaften gegebener Wissensstrukturen wie Konsistenzfragen, Zugänglichkeit von Wissen etc. auf Basis eines mengensprachlich formulierten Modells von Wissensstrukturen.

Durchgeführt wurden ferner explorative Untersuchungen der praktischen Nutzbarkeit der Ergebnisse im Bereich wissensverarbeitender und natürlichsprachlicher Systeme. Im Rahmen einer Gastforschertätigkeit des Verfassers bei IBM Deutschland (1986-88) wurde das Konzept in einer logikbasierten Wissensrepräsentationssprache integriert und in ersten Ansätzen implementiert (Projekt LILOG [Rollinger/Studer/Uszkoreit/Wachsmuth87]). Die Arbeiten zur *Wissensstrukturierung* wurden in dem an der Universität Osnabrück angesiedelten LILOG-Teilprojekt "Wissensstrukturen und lexikalische Strukturen" fortgesetzt und in eine erweiterte Version des bei IBM entwickelten Prototypen eines textverstehenden Systems integriert. Das Konzept erlaubt eine topic-bezogene Organisation von

bereichsspezifischem Wissen, welche selektive Benutzung zuläßt. Hiermit läßt sich die Grundidee für einen *bedarfsgesteuerten Wissenszugang* skizzieren:

- Jeder Wissensbasiseintrag (Formeln, Sortendeklarationen, Prädikate, Referenzobjekte) wird einem oder mehreren *Wissenspaketen* zugeordnet. Wissenspakete können mittels eines Fokuskonzepts der Verarbeitung durch das System zugänglich gemacht werden.
- Zur Vorpositionierung des Fokus auf relevante Wissenspakete dienen *Topics*. Einem Wissenspaket kann eine *Topic-Liste* zugeordnet sein.
- Für die von einer Eingabe indizierte Topic-Menge wird jeweils der größtmögliche Grad der Übereinstimmung mit der Topic-Liste eines Wissenspakets festgestellt (Topics müssen nicht eindeutig sein).
- Wenn bei der linguistischen Vorverarbeitung eines neuen Textabschnitts (einzeln oder kumuliert auftretende) Inhaltswörter als relevant für einen möglichen Themenwechsel erkannt werden, wird der Fokus neu positioniert, sonst persistiert er.

Das hier nur kurz umrissene Konzept der *Wissenspaketstrukturen* ist Ausgangspunkt für Überlegungen zur Modularisierung Wissensbasierter Systeme. Es wurde zusammenfassend gemäß folgender Vorgehensweise entwickelt:

- kognitive Grundlegung
- empirische Absicherung
- prototypische Implementierung
- Beschreibung durch Prinzipien
- mathematische Formalisierung
- Integration in eine Wissensrepräsentationssprache
- explorative Evaluation

Augenblicklich vorangetriebene Arbeiten richten sich auf die formale Ausarbeitung des Modularisierungskonzepts auf Spezifikationsebene und die Organisation lexikalischen Wissens auf Basis bereichsbezogener (modularer) Ontologien. Ferner sind Vorarbeiten für mehrere Projekte angelaufen, in denen insbesondere die Team-Entwicklung von Wissensbasen angegangen werden soll.

## 6. Stoßrichtungen für Projekte

### a) Expertensysteme in der Medizin

Hier sind wissensbasierte Computerverfahren gefragt, die den Arzt bei komplexen Diagnose- und Therapieentscheidungen unterstützen (Experten- und Konsultationssysteme). Die Kluft zwischen dem Stand medizinischer Forschungserkenntnisse und dem Umfang ihrer Anwendung beim Einsatz in der Patientenversorgung gründet sich hauptsächlich auf die Tatsache, daß praktizierende Ärzte häufig nicht über die jüngsten Daten der Forschung verfügen und zudem Entscheidungen unter beschränkten zeitlichen Ressourcen treffen müssen. Eine Chance für Anwendungsmöglichkeiten der Künstlichen Intelligenz läge darin, die Expertise und den Informationsstand von Ärzten durch künstliche Systeme zu ergänzen.

Einen Anwendungskontext für Projekte im Bereich medizinischer Expertensysteme bietet das gegenwärtig innerhalb der medizinischen Einrichtungen in Bad Oeynhausen aufgebaute *Zentrum Technologietransfer Biomedizin*, das mit der Universität Bielefeld kooperieren wird. Ein Bereich, der nach Vorgesprächen mit Beteiligten von Interesse für Systementwicklungen sein könnte, ist der Bereich Toxikologie, speziell toxische Einflüsse durch Umweltbelastung. Medizinische Diagnostik ist im Bereich Expertensysteme ein mittlerweile recht gut verstandenes Gebiet, soweit es die Problemlösungskomponente solcher Systeme betrifft. Andererseits ist der Bereich Toxikologie als ausufernder Bereich bekannt, schon allein, was sog. Faktendatenbanken anbelangt. Soll er sowohl im Rahmen eines Informations- bzw. Konsultationssystems als auch integriert mit Diagnostiksystemen erschlossen werden, liegt die Frage einer modularen Vorgehensweise auf der Hand.

Neben Bestandsaufnahmen - Was können medizinische Expertensysteme bisher leisten? In welchen Kontexten werden sie eingesetzt? - wären zentrale Forschungsfragen u.a. die folgenden:

- (Wie) läßt sich ein Diagnostik-System im Bereich Toxikologie realisieren?
- Wie soll die Ergebnisdarstellung und -erklärung aussehen?
- Wie soll der Wissenserwerb gestaltet werden?
- Wie lassen sich Eigenschaften eines Informations- und eines Diagnostik-Systems in der Toxikologie verbinden?
- Inwieweit lassen sich Zugriffsmöglichkeiten auf Fachliteratur einbeziehen? Wie erfolgt die Aktualisierung?

Einbezogen werden sollen von Beginn an Fragen des sozial- und individualverträglichen Einsatzes solcher Systeme; Kooperationsgespräche wurden mit Bielefelder Sozialwissenschaftlern aufgenommen. Hier wären Fragen des verantwortlichen Einsatzes solcher Systeme von Interesse: Wie soll die Kompetenzabgrenzung Arzt - System aussehen? Können Systeme in Ansätzen selbst zur Kompetenzeinschätzung befähigt werden? Diese letzte Frage wirft wiederum Probleme auf, die im Bereich Künstliche Intelligenz der Grundlagenforschung zuzurechnen sind.

#### b) Modularisierung von Wissensbasen für sprachverarbeitende Systeme

Hier sind wissensbasierte Computerverfahren gefragt, die vor dem Hintergrund des repräsentierten Wissens über einen Weltausschnitt *Bedeutungsrepräsentationen* für natürlichsprachliche Eingaben (Aufträge, Anfragen, Texte) erstellen, auf die eine Wissensverarbeitungskomponente aufsetzt und entsprechend dem Zweck eines sprachverarbeitenden Systems agiert. Dies kann die Erzeugung natürlichsprachlicher Ausgaben aus Bedeutungsrepräsentationen einschließen. Als Beispiele sind sprachverstehende Systeme, Frage-Antwort-Systeme und Systeme zur Wissensakquisition aus natürlichsprachlichen Wissensquellen zu nennen.

Während derzeitige Erfolge der Forschung und -entwicklung im Gebiet Expertensysteme ganz wesentlich darauf beruhen, daß sich der modellierte Weltausschnitt auf eng umrissene Gebiete beschränkt (Bereichsspezifität), stößt die Übernahme dort etablierter Wissensrepräsentations- und -verarbeitungsmethoden bei natürlichsprachlichen Systemen an Grenzen, wenn das betrachtete Sprachfragment nicht auf enge Bereiche festgelegt werden kann. Bereits zum Verstehen sprachlicher Eingaben muß auf umfangreiches Rahmen- oder Hintergrundwissen ("Weltwissen") zurückgegriffen werden. Bereitet die Unterscheidung von terminologischem und assertionalem Wissen in Expertisebereichen normalerweise wenig Probleme, liegt bei natürlichsprachlichen Systemen eine der Schwierigkeiten darin, daß die Grenzen zwischen lexikalischer Information und allgemeinem Wissen über die Welt fließend sind. Abhängig vom Kontext können zudem unterschiedliche Begriffsattribuierungen zum Tragen kommen. Auch soll ein System nur die im jeweiligen Kontext angebrachten Inferenzen machen können, um implizites Wissen zu explizieren.

Stoßrichtung für konkrete Projekte ist daher die Entwicklung von Methoden und Techniken für wissensbasierte Sprachverarbeitungssysteme, die strukturierte Bestände

von Hintergrundwissen auf modularer Basis bereithalten und "bedarfsgesteuert" einsetzen können. Im Zusammenhang damit sollen insbesondere auch ingenieurmäßige Methoden zur Entwicklung, Dokumentation, Wartung und - im notwendigen Umfang - Konsistent-erhaltung großer Lexika und Wissensbasen angegangen werden.

Eine wesentliche Voraussetzung für die Definition und Durchführung konkreter Projekte ist die Kooperation mit Partnern auf Seiten der Computerlinguistik, die an der Bearbeitung größerer Weltausschnitte Interesse und Bedarf haben. Die Voraussetzungen sind in Bielefeld hierfür günstig; gegenwärtig geplante Forschungsvorhaben können an vorhandene Schwerpunkte in der Fakultät für Linguistik und Literaturwissenschaft anknüpfen.

## **7. Kooperationen und Perspektiven**

Entscheidende Bedeutung für Forschungsaktivitäten im genannten Bereich ist der z.Tl. bereits begonnenen Kooperation mit Partnern im akademischen Bereich - in Bielefeld und an anderen Hochschulstandorten - und mit Industriepartnern zuzumessen:

Im Rahmen von Vorhaben im Bereich "Textverstehen" bestehen Kooperationen mit Arbeitsgruppen der Universitäten Bielefeld und Osnabrück (Prof. Rickheit, Dr. Gust und Prof. Kanngießer). Diese Aktivitäten beziehen insbesondere das bisherige engere Umfeld des Verfassers in Osnabrück ein und setzen dort begonnene Arbeiten an Modularisierungskonzeptionen fort.

Seitens der Arbeitsgruppe Computerlinguistik (Prof. Metzling) an der Universität Bielefeld besteht Interesse, Forschungsvorhaben im Bereich natürlichsprachlicher Systeme in Kooperation mit der Technischen Fakultät voranzutreiben. Das Umfeld für solche Vorhaben in Bielefeld wird ergänzt durch die DFG-Forschergruppe "Kohärenz" und in Bielefeld bereits aufgebaute Projektkooperationen mit weiteren bundesdeutschen Universitäten im DFG-Schwerpunkt "Kognitive Linguistik". Über Prof. Metzling besteht ferner Kontakt zu anwendungsorientierten Forschungsaktivitäten der Künstlichen Intelligenz im Rahmen des Forschungsverbundes "KI-Anwendungen in NW".

Vorgespräche über Kooperationsmöglichkeiten hinsichtlich von Aspekten des Software Engineering sind ferner mit dem Vertreter des Arbeitsgebiets Praktische Informatik an der Technischen Fakultät (Prof. Giegerich), aufgenommen worden.



Im Hinblick auf Industriepartnerschaft wird zunächst die begonnene Kooperation mit IBM Stuttgart, im Projekt LILOG fortgesetzt (Schwerpunkt Wissensstrukturierung). Diese Kooperation schließt spezifizierte Zusammenarbeit mit anderen universitären Vertragspartnern der IBM ein. Weitere Kooperationsmöglichkeiten mit dem neugegründeten "Institut für Wissensbasierte Systeme" am Wissenschaftlichen Zentrum der IBM Deutschland GmbH zeichnen sich auch im Bereich medizinischer Expertensysteme ab.

Langfristige Entwicklungsmöglichkeiten der hier vorgelegten Forschungsansätze laufen auf die Konzipierung Wissensbasierter Systeme der "2. Generation" hinaus, welche u.a. semantische Informationsverarbeitung mit Datenbankmethoden kombinieren sollen (externe Speicherung; wissensbasiertes Information Retrieval). Aus den Ausführungen dürfte klar geworden sein, daß dabei der Modularisierung Wissensbasierter Systeme zentrale Bedeutung zukommen wird.

## Literatur

- [Bauer75] F.L. Bauer, (ed.): *Software Engineering. An advanced course*. Berlin: Springer, 1975<sup>2</sup>.
- [Dennis75] J.B. Dennis: The design and construction of software systems. In F.L. Bauer (ed.): *Software Engineering. An advanced course*.(pp.12-28) Berlin: Springer, 1975<sup>2</sup>.
- [Kimm/Koch/Simonsmeier/Tontsch79] R. Kimm, W. Koch, W. Simonsmeier & F. Tontsch: *Einführung in Software Engineering*. Berlin: Walter de Gruyter, 1979.
- [Newell82] A. Newell: The knowledge level. *Artificial Intelligence* 18 (1), 1982, 1-20.
- [Parnas72] D.L. Parnas: On the criteria to be used in decomposing systems into modules. *Communications of the ACM* 15 (12), 1972, 1053-1058.
- [Parnas75] D.L. Parnas: Software engineering or methods for the multi-person construction of multi-version programs. In K.E. Hackl (ed.) *Programming methodology. 4th Informatik Symposium, IBM Germany, Wildbad 1974* (pp.225-235). Berlin: Springer, 1975.
- [Rollinger/Studer/Uszkoreit/Wachsmuth87] C.R. Rollinger, R. Studer, H. Uszkoreit & I. Wachsmuth: Textunderstanding in LILOG: Sorts and reference objects. In W. Brauer & W. Wahlster (Hg.): *Wissensbasierte Systeme, 2. Internationaler GI-Kongreß, München 1987* (S.246-259). Heidelberg: Springer, 1987.
- [Wachsmuth85] I. Wachsmuth: LAKOS - Ein Modell der Wissensrepräsentation zur Erklärung kognitiven Verhaltens. In H. Mandl. & P.M. Fischer (Hg.): *Lernen im Dialog mit dem Computer*. München: Urban & Schwarzenberg, 1985.
- [Wachsmuth89] I. Wachsmuth: *Zur intelligenten Organisation von Wissensbeständen in künstlichen Systemen*. Habilitationsschrift, Universität Osnabrück 1989 (erscheint als IWBS-Report).