

User Modelling and Adapted Interaction:

Ankündigung für eine Internationale Zeitschrift

Eine neue Zeitschrift wurde gegründet, die ein interdisziplinäres Forum für die Verbreitung neuer Forschungsergebnisse unter allen Aspekten der Benutzergestaltung und benutzer-angepaßten Interaktion bei Mensch-Maschine-Schnittstellen, natürlich-sprachlichen Dialogsystemen, intelligenten Tutorssystemen und intelligenten Schnittstellen darstellt. Die Zeitschrift veröffentlicht hoch qualifizierte Originalabhandlungen mit Beiträgen aus den genannten Bereichen, einschließlich der folgenden Gebiete:

- . Akquisition von Benutzer- und Studentenmodellen
- . konzeptuelle Modelle, mentale Modelle
- . Dialogplanung und Antwortzuschnitt
- . Ebenen der Benutzerexpertisen
- . Planerkennung und -entwicklung
- . Planung der Präsentation
- . Erkennen und Korrigieren von Mißverständnissen
- . Benutzerstereotype
- . formale Repräsentation von Benutzer- und Studentenmodellen
- . Shells für Benutzergestaltung
- . Beratungsstrategien

Relevante Abhandlungen aus den Bereichen der Psychologie, Linguistik und der Erziehungswissenschaften werden ebenfalls berücksichtigt.

Der hauptsächliche Leserkreis der Zeitschrift besteht aus Wissenschaftlern, Studenten und Wirtschaftspraktikern aus folgenden Bereichen: Künstliche Intelligenz (Schwerpunkt auf "knowledge-based systems"), Mensch-Computer-Interaktion (Schwerpunkt auf "cognitive engineering and intelligent interfaces"), Linguistik (Schwerpunkt auf "pragmatics and dialog models") und Erziehungswissenschaften (Schwerpunkt auf "computer-based tutoring systems").

Editor-in-Chief: *Alfred Kobsa*, Dept. of Computer Science, University of Saarbrücken, D-6600 Saarbrücken 11, WEST GERMANY

Editorial Board: *David Benyon*, Open University, England * *Sandra Carberry*, Univ. of Delaware, U.S.A. * *David Chin*, Univ. of Hawaii at Manoa, U.S.A. * *Robin Cohen*, Univ. of Waterloo, Canada * *Timothy Finin*, Unisys, Paoli, U.S.A. * *Gerd Fischer*, Univ. of Colorado, U.S.A. * *Gordon McCalla*, Univ. of Saskatchewan, Canada * *Dianne Murray*, City University, England * *Cecile Paris*, Information Science Institute, U.S.A. * *Elaine Rich*, MCC, Austin, U.S.A. * *Hans Spada*, Univ. of Freiburg, West Germany * *Karen Sparck-Jones*, Univ. of Cambridge, England * *Michael Tauber*, Univ. of Paderborn, West Germany * *Wolfgang Wahlster*, Univ. of Saarbrücken, West Germany * *Richard Young*, MCR, Cambridge, England * Das Editorial Board wird weiter ergänzt werden.

UMUAI wird vierteljährlich herausgegeben von Kluwer Academic Publishers, Dordrecht, Niederlande. Die erste Ausgabe soll im Januar 1991 erscheinen.

Autoren, die einen Beitrag einreichen möchten, sollten vier Kopien davon an den Editor schicken. Elektronische Einsendung von LaTeX-Dokumenten ist ebenfalls möglich, wenn der spezielle UMUAI-Stil verwendet wird. Für weitere Informationen (besonders über elektronische Vorlage) senden Sie bitte eine Nachricht an umuai-info%fb10vax.informatik.uni-saarland.dbp.de@relay.cs.net oder [unido!sbsvax!fb10vax!umuai-info](mailto:(uunet,mcvax!)unido!sbsvax!fb10vax!umuai-info). Diese Adressen werden möglicherweise 1990 geändert in umuai-info@cs.uni-sb.deo

Technikfolgenabschätzung in der Informationstechnik

BMFT-Projekt und Newsletter

Informationstechnik ist gegenwärtig zweifellos ein besonders dynamischer Bereich der Technikentwicklung. Wie jede andere hat auch diese Technik Folgen - beabsichtigte und unbeabsichtigte, erwünschte wie auch unerwünschte. Neue Technologien aus diesem Bereich sollten deshalb frühzeitig auf ihre Folgen hin untersucht werden. Risiken und Chancen können auf diese Weise rechtzeitig erkannt und Gestaltungsmöglichkeiten rechtzeitig

wahrgenommen werden.

Die vielfältigen Wechselwirkungen zwischen Technik, Wirtschaft und Gesellschaft zu analysieren und zu bewerten, ist nicht von einzelnen Institutionen oder sogar einzelnen Personen zu leisten. Diese Herausforderung ist nur durch die Mitwirkung verschiedener Disziplinen zu bewältigen. In strukturierten Diskussionen der zusammengeführten vielfältigen Kenntnisse und Erfahrungen sollen Erkenntnisse und Ansichten ausgetauscht, unterschiedliche Auffassungen erkannt und benannt, konsensfähige Aussagen gewonnen werden. Diesen Zielen dient ein Vorhaben des Bundesministeriums für Forschung und Technologie (BMFT), wobei in einer Presseerklärung des BMFT die Entwicklung und die Methoden der künstlichen Intelligenz, speziell der Expertensysteme, als die Spitze der Entwicklung bezeichnet werden. In dieser Presseerklärung wird auch ausdrücklich die Bereitschaft des BMFT unterstrichen, Aktivitäten zur Technologieabschätzung zu unterstützen. Als Forum für Informationen und Meinungen in dem angestrebten Prozess liegt nun mehr die erste Ausgabe eines Informationsblattes vor, das bei der Redaktion, *Reinhard Stransfeld, VDI/VDE - Technologiezentrum Informationstechnik GmbH, Budapester Straße 40, D-1000 Berlin 30* bestellt werden kann.

VDI/VDE - Technologiezentrum, Berlin

Maschinelles Lernen und Neuronale Netze

Brainware Seminar mit E. von Goldammer

Die internationale Forschung im Bereich "Maschinelles Lernen" erlebt seit geraumer Zeit eine wachsende Renaissance des Paradigmas "Neuronale Netze". Auch in Deutschland ist mittlerweile ein stark gestiegenes Interesse an diesen Arbeiten zu beobachten. Der Themenkreis "Neural Networks" war im November Gegenstand eines High Level Seminars der Brainware GmbH, welches eine nüchterne Einschätzung der realen Möglichkeiten dieses mit viel Euphorie verfolgten Forschungszweiges ermöglichen sollte.

Professor E. von Goldammer und Mitarbeiter des Institutes für Biophysik und Biokybernetik in Neumünster/Lübeck trugen drei Tage lang die Ergebnisse ihrer mehrjährigen Forschungs- und Entwicklungsarbeit auf dem Gebiet der kognitiven Wissenschaften vor. Nach einer Klärung von Grundlagen wurden typische Netzwerkmodelle im Detail besprochen und durch anschließende Praktika mit Neural Nets Software am Computer vertieft.

Hard- und Software wurden eingehend behandelt. Abgerundet wurde das Veranstaltungsprogramm durch Videos von den Neural Nets "Päpsten" Widrow und Rumelhart sowie durch den Vortrag "Trends und Perspektiven" (Goldammer).

Über das eigentliche Programm hinaus wurde durch abendliche Kaminrunden die richtige Atmosphäre für Diskussionen geschaffen, die immer bis weit in die Nacht hinein reichten.

Aufgrund des aktuellen Interesses und der großen Nachfrage nach diesem Seminar stehen bereits die Termine für eine ganze Veranstaltungsreihe "Neurocomputing und Neokonnektionismus" und andere Themen aus dem Bereich "Maschinelles Lernen" im Jahr 1990 fest. Interessenten können sich an Brainware Berlin (Tel.: 0304633048) wenden.

Pressemitteilung Brainware

Projektvorstellung wit:

Automatischer Aufbau von Wissensbasen durch Wissensakquisition aus Texten

Das wit-Projekt (REIMER et al. 89) will einen Beitrag dazu leisten, den für wissensbasierte Systeme bestehenden Engpaß beim Aufbau und der Aktualisierung ihrer Wissensbasen zu überwinden. Es werden Verfahren entwickelt und in einem prototypischen System realisiert, die es erlauben, Wissen zu dem Diskursgebiet *Informations- und Kommunikationstechnologie* aus nicht präeditierten, deutschsprachigen Fachartikeln verschiedener Verfasser abzuleiten und in eine Wissensbasis zu integrieren.

Aus den Texten soll primär definitives Wissen gelernt werden, also die (möglichen) Eigenschaften von Konzepten und ihre semantischen Beziehungen zu anderen Konzepten. Soweit möglich, werden auch kontingente Aussagen (z.B. in Form von Empfehlungen, Meinungen, Stellungnahmen) über Konzepte gelernt.

Um eine größtmögliche Diskursbereichsunabhängigkeit zu erzielen, soll das System mit nur geringem domänenspezifischem Vorwissen ausgestattet werden. Dieses Wissen wird mit wachsender Zahl analysierter Texte durch die Lernergebnisse zunehmend ausgebaut. Die sonstigen vorgegebenen Wissensquellen sind domänenunabhängig und werden deshalb extensiv eingesetzt. Dies sind domänenübergreifendes Wissen sowie Wissen über sprachliche Phänomene auf der Wort-, Nominalgruppen-, Satz- und Textebene.

Als längerfristiges Forschungsziel wird angestrebt, mit einer durch weit aufgebauten Wissensbasis, in Verbindung mit geeigneten "Front-Ends", neuartige Informationsprodukte zu entwickeln, beispielsweise eine flexible, benutzerspezifische interaktive Präsentation von Fachwissen (vgl. KUHLEN et al. 89) oder die Generierung von Berichten über den aktuellen Stand der Technik.

Zur Lösung der skizzierten Aufgabenstellung wird ein den Zielen von weit angemessener *partieller semantischer Textparser* zur Behandlung von Textkohäsions- und Textkohärenzphänomenen entwickelt, um die in Texten mitgeteilten Aussagen über Konzepte erschließen zu können. Zusätzlich wird an einer *Lernkomponente* gearbeitet, die auf der Basis der vom Textparser erschlossenen Aussagen das Lernen von Konzepten über mehrere Texte hinweg leistet und dabei die notwendigen Erweiterungen, Generalisierungen und Homogenisierungen von Konzeptbeschreibungen vornimmt. Die Textverstehens- und Lernkomponente setzen geeignete *Wissensrepräsentationskonstrukte* voraus, die die Darstellung des vorgegebenen wie auch des neu erworbenen Wissens ermöglichen. Dabei wird auf das in Konstanz entwickelte FRM (vgl. REIMER 89) zurückgegriffen, das um neue Möglichkeiten der Darstellung kontingenten Wissens erweitert wird.

Das Projekt läuft seit Frühjahr 89 und wird zur Zeit mit Eigenmitteln des Lehrstuhls Informationswissenschaft durchgeführt. Ein erster wit-Prototyp zur Bestimmung der Tragfähigkeit der gewählten Ansätze wird auf einer Sun-4 in Prolog und C++ implementiert.

Projektmitglieder:

Rainer Kuhlen / Ulrich Reimer / Gabi Sonnenberger / Klaus Mußnug (Fachgruppe Informationswissenschaft an der Universität Konstanz, Postfach 5560, D-7750 Konstanz)

Literaturhinweise:

Kuhlen, R./ Hammwöhner, R./ Sonnenberger, G./ Thiel, U.: TWRM-TOPOGRAPHIC: Ein Wissensbasiertes System zur situationsgerechten Aufbereitung und Präsentation von Textinformation in graphischen Retrievaldialogen. In: *Informatik Forschung und Entwicklung*, Band 4, 1989. S. 89-107.
Reimer, U.: FRM: Ein Frame-Repräsentationsmodell und seine formale Semantik. Zur Integration von Datenbank- und Wissensrepräsentationsansätzen. Berlin: Springer-Verlag, 1989.
Reimer, U./ Sonnenberger, G./ Kuhlen, R./ Mußnug, K./ Senf, K.: wit: Automatischer Aufbau von Wissensbasen durch Wissensakquisition aus Texten. Universität Konstanz, Informationswissenschaft, 1989.

„Veränderung der Wissensproduktion und -verteilung durch Expertensysteme“

Verbundprojekt zur Einbettung von Technikfolgenforschung in den Prozess der KI-Entwicklung

Ab Januar 1990 fördert das BMFT ein interdisziplinäres Verbundvorhaben, in dem der Versuch unternommen wird, Technikfolgenforschung in den Prozeß der KI-Entwicklung einzubetten. Sprecher des Projektes, das vom Nordrhein- Westfälischen KI-Forschungsverbund durchgeführt wird, ist Prof. Cremers, Dortmund, der auch als Obmann den Diskurs "Menschenbild der KI" beim VDI betreut (Info 1/90).

Mit Expertensystemen wird versucht, in eng begrenzten Wissensgebieten die Vorgehensweise von Experten zu simulieren. Expertensysteme sind in der Geschichte der KI-Forschung die Konsequenz aus dem Scheitern der Versuche, Computer als allgemeine Problemlöser zu programmieren. Die Konzentration auf eingegrenzte Problembereiche ist ein entscheidender Schritt, um auf dem Gebiet der Wissensverarbeitung zu anwendungsreifen Systemen zu gelangen.

Expertensysteme können eingesetzt werden, um die orts- und zeitgebundene Leistung von Spezialisten allgemein verfügbar zu halten, um Experten von Routineaufgaben zu befreien oder um sie bei der Lösung besonders komplexer Aufgaben zu unterstützen. Zu bedenken sind mögliche Veränderungen, die den Status von Experten, die Qualität und Verantwortbarkeit der Expertisen sowie die Entwicklung von Wissensgebieten betreffen.

Der derzeit erreichte Entwicklungsstand der Expertensystemtechnik erlaubt es

- Technikfolgenabschätzung im Rahmen der Technikgenese zu betreiben, d. h. Entscheidungsgrundlagen der Entwickler zu analysieren und Ansatzpunkte für frühzeitige Korrekturen herauszufinden,
- die gewonnenen Erkenntnisse präventiv in Entwicklungsmethoden und Gestaltungskonzepte umzusetzen, um unerwünschten Folgen vorzubeugen.

In den Teilprojekten des Verbundes werden folgende Fragen bearbeitet:

- * Wie verändert sich durch den Einsatz und die Entwicklung von Expertensystemen das Wissen von Experten und die Wissensstruktur von Fachgebieten? Hierbei spielen die Vorstellungen, die Systementwickler vom menschlichen Wissen haben, eine wesentliche Rolle,

ebenso wie die verwendeten Methoden und Instrumente bei der Wissenserhebung. Ähnlich wie die Erfindung der Schrift und des Buchdruckes nachhaltigen Einfluß auf die Entwicklung von Wissensgebieten nahmen, kann dies auch für den Einsatz computergestützter Medien unterstellt werden. Dies betrifft den Wissensstand des am Entstehungsprozeß eines Systems beteiligten "Produzenten" (vergleichbar dem Buchautor) ebenso wie den Nutzer (vergleichbar dem Leser von Fachbüchern).

* Wie wird das Wissen eines Fachgebietes zwischen Mensch und Maschine verteilt, welche Handlungsmöglichkeiten haben Benutzer gegenüber Expertensystemen und was kann man aus bisherigen Fachdialogen für die Gestaltung lernen? Wenn man sich auf den Standpunkt stellt, daß Expertensysteme den menschlichen Experten nicht ersetzen, sondern unterstützen sollen, dann muß es dem Nutzer möglich sein, Erklärungen abzufordern, Einblick ins System zu nehmen und selbstgefundene Lösungen zu überprüfen. Es ist zu untersuchen, ob durch solche Handlungsmöglichkeiten Anwendungsprobleme angegangen, die Qualität und Verantwortbarkeit von Ergebnissen gesichert und Wissensgebiete weiterentwickelt werden können.

* Wie wirken sich die Kommunikationsprozesse bei der Entwicklung und beim Einsatz wissensbasierter Systeme auf deren Qualität und Anwendungsmöglichkeiten aus? Experten müssen befragt werden, Ihre Aussagen sind zu interpretieren, Nutzer geben ihre Arbeitsergebnisse an andere weiter und so fort. Dieser "Kommunikationsweg" über mehrere Stationen ist störanfällig und kann zu Verständigungsproblemen führen.

* Welche Organisationsprobleme beeinflussen die Leistung und Benutzbarkeit von Expertensystemen, und wie verändert die neue Technik bestehende Organisationsstrukturen? Bei der Erhebung und Vermittlung von Wissen über eingesetzte Expertensysteme werden die Rollen von Experten und die Kooperationsbeziehungen in Betrieben ggf. neu zu organisieren sein. Zu untersuchen ist "wie der Einsatz von Expertensystemen Arbeitsabläufe und Qualitätsanforderungen verändert."

Das Vorhaben soll die Realisierungsmöglichkeiten einer eingebetteten Technikfolgenforschung kritisch reflektieren, wie auch Hinweise zur Entwicklung und Anwendung von Expertensystemen liefern.

Kontakt:

Prof. Dr. Armin B. Cremers (Sprecher)
Universität Dortmund, FB Informatik 6

Postfach 50 05 00, 4600 Dortmund 50
Tel.: (0231) 755-2116

weitere beteiligte Institutionen:

- Prof. Dr. R. Jünemann, Fraunhofer-Institut für Materialfluß und Logistik (IML), Dortmund,
- Prof. Dr. D. Metzger; Dr. R. Weingarten, Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld,
- Prof. Dr. D. Krallmann, Fachbereich 3, Sprachwissenschaften, Universität-GH Essen,
- Prof. Dr. F. Rapp, Fachbereich Gesellschaftswissenschaften, Philosophie und Theologie, Universität Dortmund,
- Dr. P. Hoschka, Institut für Angewandte Informationstechnik (F3), Gesellschaft für Mathematik und Datenverarbeitung, Schloß Birlinghoven,
- Dr. W. Rammert, Fakultät für Soziologie, Universität Bielefeld,

Informationen zur Technikfolgenabschätzung der Informationstechnik, Jahrgang 2-1990, Nr. 2

LEXWERK - ein Computerarbeitsplatz für Lexikographen

Natürlich gehören Geistes-, Sozial- und Humanwissenschaften und Computer zusammen: zum einen, weil sie uns mit gesellschaftlichem Orientierungswissen und Wertgerüsten für Wahrnehmungen, Denken, ästhetisches Empfinden und Handeln, also auch für technisches Handeln versorgen. Zum anderen ergibt sich der Zusammenhang aus der Tatsache, daß Sprache als Gegenstand und Erkenntnisinstrument dieses Wissenschaftsbereiches fungiert. Als wahrhaft interdisziplinärer Kern nicht naturwissenschaftlicher Fächer steht denn auch die Linguistik, - ihre Erkenntnisinteressen überlappen mit den Sozialwissenschaften, der Philosophie, der Literaturwissenschaft, den historischen Wissenschaften aber auch mit Psychologie und Medizin, in der jahrhundertealten Frage nach der Rechenbarkeit von Sprache. Heute existiert mit dem Computer als symbolverarbeitende Maschine ein Funktionsmodell, das wichtige Eigenschaften des Erkenntnisgegenstandes und des Erkenntnismittels Sprache simulieren kann. Ob er jemals alle Eigenschaften, selbst alle definierenden Eigenschaften von Sprache als menschlicher Fähigkeit zu intelligentem, kommunikativem Handeln wird simulieren können, mag man mit Fug und Recht bezweifeln.

Grob gesagt sind an Sprache erstens zeichenketten-spezifische Eigenschaften der Signaloberfläche und zweitens Struktureigenschaften zu unterscheiden. So ist ein sprachliches Zeichen eine visuell wahrnehmbare Kette zulässiger Buchstaben, die man zählen, vorwärts und rückwärts sortieren kann, die man als ASCII Code speichern und suchen, bearbeiten und drucken kann. Die vielen Werkzeuge zur Textverarbeitung, zum Erstellen von Konkordanzen und Indizes, zur Bibliographieverwaltung und als Zettelkastenersatz beweisen, wie nützlich und allgegenwärtig der Computer in geisteswissenschaftlichen Anwendungen inzwischen geworden ist. Die Struktureigenschaften betreffen die für die Informationsübermittlung erforderlichen Merkmale, also tieferen Schichten von Sprache wie Syntax, Semantik und Pragmatik. Sie sind meist mit dem Wissen von Sprache identisch, das Menschen haben und das sie befähigt, auf intelligente Weise Bedeutungen zu kommunizieren.

Inzwischen sind viele Struktureigenschaften von Sprache als formale Theorien in Rechnern implementiert, so daß von Werkzeugen, die grammatische Fehler oder stilistische Verstöße in Texten erkennen sollen, bis zu Systemen der maschinellen Übersetzung eine weite Palette möglicher und intelligenter Sprachprodukte entsteht. Damit definiert der Rechner geradezu die Gütekriterien für sprachwissenschaftliche Theoriebildung: was sich an strukturellen Eigenschaften im Rechner darstellen läßt, und damit zur Simulation von Sprachfähigkeit führt, muß auch in einem eingeschränkten Sinn als gute Sprachtheorie gelten. Zunehmend setzt sich dabei die Einsicht durch, daß existierende Grammatikmodelle in einem Bereich der empirischen Unterfütterung bedürfen, der lange Zeit als außerhalb des theoretischen Interesses liegend angesehen wurde - im Bereich der Lexikographie. Damit wird nicht etwa die Wichtigkeit der Syntax geleugnet, sie wird nur dadurch konkreter, daß sie nun im Lexikon als regelhafte Valenz- und Bindungspotenz von Wörtern erscheint.

Für Wörterbücher und Lexikographen bricht also ein neues Zeitalter an. Waren wir seit der Aufklärung gewohnt, unter (Sprach)-wörterbuch ein für menschliche Benutzer gedachtes Verzeichnis von lexikalischen Einheiten einer Sprache zu verstehen (meist eben "Wörter"), die, systematisch angeordnet, den vorbildlichen Gebrauch unserer Nationalsprache kodifizieren und Informationen zum regelgerechten Gebrauch geben, stellen wir heute fest, daß von Orthographiehilfen, über natürlichsprachliche Schnittstellen für Datenbankabfragen bis zur maschinellen Übersetzung der Rechner nicht ohne "elektronisches" Wörterbuch auskommt: kein funktionales Modell menschlicher Sprachbeherrschung durch den Rechner kann auf eine Wörterbuchkomponente verzichten.

Aber für aussagekräftige Großversuche in der in-

telligenten Verarbeitung von Texten, z. B. in der maschinellen Übersetzung, der Inhaltserschließung und bei der automatischen Indexierung für Information Retrieval, erst recht aber für die Praxis fehlen entsprechend große, auf automatische Verarbeitung zugeschnittene elektronische Wörterbücher. Die für den menschlichen Benutzer konzipierten Papierwörterbücher sind natürlich nicht für Rechner geeignet und leider passen die für einen bestimmten Grammatikformalismus entworfenen elektronischen Wörterbücher nicht für andere Grammatiktheorien. Man kann die Größe des hier skizzierten Problems abschätzen durch einen Vergleich mit dem Prozeß des Lexikonaufbaus im kindlichen Spracherwerb. Bis zum sechsten Lebensjahr erwerben Kinder einen Wortschatz von ca. 15.000 Wörtern. Ohne die beiden ersten Lebensjahre, in denen nur wenige hundert Ausdrücke erworben werden, erzielt das Kind einen täglichen Lernzuwachs von ca. 10 Einträgen in sein mentales Lexikon. Dies entspricht ziemlich genau der Arbeitsgeschwindigkeit eines Lexikographen. Sollte er ein Wörterbuch für Zwecke der maschinellen Sprachverarbeitung in einer Größe von ca. 500.000 Einträgen erstellen - dabei sind noch nicht einmal fachsprachliche Termini berücksichtigt - so müßte er ca. 220 Mannjahre Entwicklungsarbeit leisten. Nichts liegt näher als nach Möglichkeiten zu suchen, den arbeitsintensiven lexikographischen Prozeß zu automatisieren.

Beim heutigen Stand der Linguistik und Computerlinguistik kann auch in der Computerlexikographie vom Rechner mehr verlangt werden, als nur zeichenkettenorientierte Werkzeuge zur Datenaufbereitung, z. B. in Konkordanzen, Listen, Indices, Register, zur Speicherung als ASCII Dateien und für die Editionstechnik zur Verfügung zu stellen. Obwohl der Linguist und Lexikograph bereits durch solche Werkzeuge von vielen repetitiven Arbeiten befreit wird und sich auf die morphologische, syntaktische und semantische Beschreibung der Einträge konzentrieren kann, liegt doch der eigentliche Fortschritt der Rechner- und Programmtechnologie im automatischen Parsen oder Analysieren großer Datenmengen. Wir können nicht erwarten, daß der Rechner fehlerfrei analysiert, und müssen deshalb den menschlichen Bearbeitern die Möglichkeit zur Bestätigung und Korrektur der Rechnerergebnisse einräumen. So korrigiertes Material sollte dann in eine lexikale Datenbank geschrieben werden, von wo aus sich ein Anwender ein Wörterbuch für genau seine Anforderungen zusammenstellen können müßte.

Dies sind die Zielvorgaben, an denen Mitarbeiter des Arbeitsbereiches Linguistik der Westfälischen Wilhelm-Universität Münster sitzen, um LEXWERK, einen lexikographischen Arbeitsplatz zu entwickeln. Die Entwicklungsarbeiten laufen auf leistungsfähigen CADMUS Workstations und

im Rahmen einer Hochschulkooperation der Firma PCS.

Systemkomponenten

Als empirisches Ausgangsmaterial kommen für den Computerlexikographen große Textkorpora, maschinenlesbare Wörterbücher, Belegstellensammlungen, Grammatiken, Ergebnisse der Wörterbuchkritik, andere Wörterbücher und Wortlisten in Betracht. Idealerweise sollte ein lexikographischer Arbeitsplatz eine Vielzahl solcher Daten enthalten und in einer auf diesen Datentyp ausgelegten TEXTDATENBANK speichern. Da uns in der Bundesrepublik keine maschinenlesbaren Wörterbücher zur Verfügung stehen, aus denen man automatisch die bereits erarbeiteten lexikalischen Informationen entnehmen könnte, wird in Münster ein neues, großes Textkorpus von ca. 50 Millionen Textwörtern für das Deutsche aufgebaut; die ZEIT und die FAZ haben ihre Unterstützung für das Korpus zugesagt und sind bereit, maschinenlesbare Textmaterialien zur Verfügung zu stellen.

Der lexikographische Arbeitsplatz bietet nun Werkzeuge für Zeichenketten - wie strukturorientierte Verarbeitung - an: Zur manuellen Bearbeitung werden Konkordanzen und Indizes aus dem Korpus erzeugt, die auf verschiedene Arten organisiert, verwaltet und ausgegeben werden können. Für die strukturorientierte automatische Bearbeitung paßt ein Wortexpertenmodul das Korpus an und bietet die Ergebnisse seiner grammatischen Analyse zur Bestätigung oder Korrektur durch den menschlichen Benutzer an. Idealerweise "weiß" der Wortexperte alles über Wörter; er verfügt über sprachliches Wissen, mit dem er die Wörter des Textmaterials morphologisch und syntaktisch durch partielle Beschreibungen disambiguiert. Im Gegensatz zur linguistischen Analysepraxis, in der mehrere Beschreibungsebenen von der Phonologie bis zur Semantik modular voneinander getrennt sind, wird das Wortexpertenmodul linguistisches Wissen aller Beschreibungsebenen gleichzeitig einsetzen, um möglichst umfangreiche lexikale Analysen machen zu können. Man kann ihn als Modell einer Wortgrammatik denken.

Die korrigierten Informationen der automatischen Analyse werden in eine Datenbank geschrieben, in welcher die lexikalischen Einheiten möglichst neutral hinsichtlich ihrer Orientierung auf spezifische Grammatiktheorien durch abstrakte Schemata repräsentiert werden. Diese lexikale Datenbank soll wiederverwendbar und multifunktional sein. Die Kombination von Datenbank und Wortexpertenmodul bleibt auch dann noch sinnvoll, wenn einmal eine große Datenbank deutschen Wortgebrauches existiert. Gemein- und fachsprachliche Neuerungen werden immer wieder

die Aktualisierung der Datenbank erzwingen.

Damit der Wortexperte erfolgreich parsen kann, braucht er ein Basislexikon, mit dem er sich selbst aus dem Datensumpf ziehen kann. Man verstehe darunter ein präzises grammatisches Wörterbuch hochfrequenter Funktions- und Inhaltswörter und produktiver Wortbildungselemente im Maschinenformat.

Wir hoffen, diesen Arbeitsplatz auch mit Werkzeugen für die multilinguale Textverarbeitung ausstatten zu können, damit die lexikale Datenbank um zweisprachige Wörterbücher erweitert werden kann. Utopisch ist gegenwärtig allerdings die Vorstellung, die Datenbank durch Informationen zu ergänzen, wie sie gewöhnlich in Enzyklopädien enthalten sind. Gelänge dies, wäre das, was wir wissen und das, was wir ausdrücken können, im Computer so repräsentiert wie im menschlichen Kopf. Es wird Zeit, daß sich die Geisteswissenschaften wirklich mit dem Rechner beschäftigen.

Prof. Dr. Wolf Paprotté, Universität Münster, Arbeitsbereich Linguistik
(mit freundlicher Genehmigung nachgedruckt aus: CADMUS)

