

LDV — FORUM

Gesellschaft für Linguistische Datenverarbeitung

GLDV

INHALT:

Nr. 2 Dezember 1985

(3. Jg.)

ISSN 0172-9926

Editorial	1
Forschungsförderung	
<i>Ch Hauenschild:</i> Kontextfreie Forschungsförderung	3
<i>Wippa-Meldung:</i> Forschungspolitik der Bundesregierung verfehlt	5
<i>Schacker:</i> Antwort des Bundesministeriums für Bildung und Wissenschaft	6
Schwerpunkt Textverarbeitung	
<i>B Böhm; J. Krause:</i> Exemplarisches zur Textverarbeitung an wissenschaftlichen Institutionen	11
<i>K Wothke:</i> Text-to-Speech-Systems for German — a Short Survey	17
<i>R v. Ammon:</i> Tools für die fortgeschrittene Textverarbeitung auf Mikrocomputern	25
Fachbeiträge	
<i>Ch Kemke:</i> SC — Ein intelligentes Hilfesystem für SINIX	43
<i>G Salton:</i> Remarks Concerning the Use of the Expert System Approach in Information Retrieval	61
<i>A Eisele:</i> A Lexical Functional Grammar System in Prolog	67
Thesen zu: "Ansätze der sprachorientierten KI-Forschung"	77
(Lutz S. 77; Fauser S. 80; Scheffé S. 81; Hahn S. 82; Morik S. 88; Endres-Niggemeyer S. 90)	
Nachrichten	91
(u.a. von: Universität Tübingen, TFH-Berlin, TU Berlin, Universität Mainz, Universität/Gesamthochschule Duisburg; u.a. über: FuE-Projekt für Elektronisches Publizieren, PC-Software)	
Tagungsberichte	
Klassifikation als Werkzeug der Forschung	101
Eindrücke von der IJCAI-85	103
A.U.T.-Symposium Trier	106
Jahrestagung der Kanadischen Semiotischen Gesellschaft	107
Natürliche Sprache und Wissensbasierte Systeme	108
BWAI-Tagungsbericht	112
Arbeitskreise	
<i>J Klenk:</i> "Maschinelle Sprachübersetzung"	115
<i>G Frackenpohl:</i> "Maschinelle Lexikographie"	116
Projekte	
Neue Analyse und Syntheseverfahren zur maschinellen Übersetzung	117
Sonderforschungsbereich 100 (TP A2)	119
Die Ulmer Textdatenbank	121
Aufbau der Transferlexika in EUROTRA-D	121
Veranstaltungen	123
Publikationen	137
Mitteilungen aus der GLDV	157
<i>M Lutz-Hensel:</i> Liebe neue GLDV-Mitglieder S. 157; Protokoll der Beiratssitzung S. 162)	

LDV-FORUM

Nr.2 , Dezember 1985 (3. Jg)

Editorial

Mitteilungsblatt der Gesellschaft für Linguistische Datenverarbeitung e.V. (GLDV)

Die Beiträge vertreten ausschließlich die Meinung der Autoren

Redaktion:
Dr. Gerhard Knorz,
TH Darmstadt, FB Informatik,
Karolinenplatz 5
D-6100 Darmstadt
(tel. 06151/16-3952 , -2859)

Erscheinungsweise:
halbjährlich: Juni /
Dezember

Nächster Redaktionsschluß: 2. Mai 1986

Bezugsbedingungen:
Einzelheftpreis: DM 15,-
Bestellung an die Redaktion). Für Mitglieder der GLDV im Jahresbeitrag enthalten

Anzeigenwünsche sind an die Redaktion zu richten.

Beiträge:
Bitte beachten Sie die "Hinweise für Autoren" auf dem rückwärtigen Umschlagsblatt.

Bankverbindung:
Sparkasse Bonn,
BLZ 380 500 00
Kto.-Nr. 120 825 633

Herstellung: Gesellschaft für Information und Dokumentation (GID), Abteilung für Reprotechnik

Auflage:
300 Exemplare

Noch fehlen einige technische Arbeitsgänge, so daß sich beim Schreiben dieser Zeilen das große "Aufatmen" noch nicht so recht einstellen mag. Aber immerhin scheint eines der von mir als unverzichtbar eingeschätzten Ziele (fast) erreicht: Der Offsetdruck wird noch vor Weihnachten die Druckvorlage erhalten und somit kann mit dem Versand der LDV-FORUM-Ausgabe 2/85 Mitte Januar gerechnet werden: Ein halber Monat Verspätung mag als verzeihbar angesehen werden. Ein wesentlicher Grund der Verzögerung war gewesen, daß die Hardware-Situation, die als Infrastruktur für die Heferstellung genutzt werden kann, sich gerade im Umbruch befindet. Die Hoffnung, zusätzliche Ressourcen bereits diesmal nutzen zu können, war leider nur ein Störfaktor für einen zeitigeren Beginn der "Abschlußarbeiten".

Ein Teil der inhaltlichen Konzeption des LDV-FORUM sollte bereits am Inhaltsverzeichnis auszumachen sein. Als regelmäßig sind geplant: Thematische Schwerpunkte wie diesmal "Textverarbeitung", breit gestreute Fachbeiträge, die das Heft in seiner Substanz tragen, fachliche und wissenschaftspolitische Stellungnahmen (wie diesmal in Form von Thesen als Vorbereitung des nächsten Themenschwerpunktes) und Berichte, aktuelle Informationen aus dem Hochschulbereich und darüber hinaus, umfassender Überblick über Veranstaltungen, Veröffentlichungen und GLDV-Aktivitäten.

Das fachliche Niveau und die thematische Orientierung der Beiträge bestimmt sicherlich zu einem wesentlichen Teil die Attraktivität des LDV-FORUM. Daß es sich mit der vorliegenden Ausgabe auf dem richtigen Weg befindet, möchte ich mir gerne von den Lesern bestätigen lassen. Ich selbst empfinde es als eine positive Entwicklung, daß die redaktionelle Planung vom bisher praktizierten "von-der-Hand-in-den-Mund-leben" abrücken kann: Die Zusagen für die nächste Ausgabe stellen deren Substanz bereits sicher, ohne Aktualität und Offenheit zu gefährden.

Gestiegenen Ansprüchen an den Inhalt des Heftes sollte das äußere Erscheinungsbild folgen. "Sensiblen" Lesern mag vielleicht schon bei der vorliegenden Ausgabe der dezente Einfluß eines "Designer-Teams" auffallen.



Tatsächlich haben Studenten der Darmstädter Fachhochschule vom Fachbereich Gestaltung sich gerade darangesetzt, das Layout für eine spätere Heftproduktion mit Hilfe des TeX-Systems zu entwerfen. Wenn diese Arbeit gut voranschreitet, könnte man auf der GLDV-Jahrestagung in Göttingen bereits über erste Ausarbeitungen diskutieren.

Noch nicht endgültig geklärt ist leider die optimale Schnittstelle, über die Autoren ihre Beiträge maschinenlesbar an die Redaktion übermitteln können. Hier werden sich jedoch die Randbedingungen sehr schnell klären, so daß spätestens im nächsten Heft genaue Spezifikationen angegeben sein werden (siehe dazu auch die "Hinweise für Autoren").

Zwei Fragen waren von mir für diese Ausgabe vorläufig zu beantworten. Wenngleich meine Antworten durchaus eine Begründung haben, will ich sie doch zunächst einmal an den Leser weitergeben:

- in welchen Fällen sollen Beiträge in englischer Sprache aufgenommen werden?
- wie steht es mit Beiträgen (z.B. Tagungsberichten), die auch in "benachbarten" Publikationen abgedruckt sind.

Nachdem nun auch der allerletzte Redaktionsschluß unwiderruflich überschritten ist, eine kleine Zusammenstellung "Was fehlt":

- Eine ganze Reihe zugesagter Tagungsberichte (leider)
- Einige "Thesen zu Ansätzen der sprachorientierten KI" (dieses Thema ist allerdings beim nächsten Mal nicht überholt - im Gegenteil!).
- Mehr eigenes Engagement in inhaltlicher Hinsicht - hier habe ich einige Vorsätze leider nicht umsetzen können.
- (weitgehend) Nachrichten aus dem "Nicht-Hochschulbereich". Nach anfänglich positiven Reaktionen leider Funkstille.
- Ein geplanter "Blick zum großen Bruder ACL", anlässlich des 10-jährigen Bestehens der GLDV (aufgehoben ist nicht aufgeschoben).
- Rezensionen aktueller oder auch einfach wichtiger Bücher: Hier sollte sich der Hochschulbereich doch wohl ansprechbereit zeigen! (Vorschläge liegen bereits auf dem Tisch)

Was noch bleibt ist, allen Autoren und Helfern zu danken (von letzteren insbesondere J. Reichardt, Andrea Trotter und L. Rostek), dem Arbeitskreis für Semiotik, Berlin und dem GID-IZ für die Zusammenarbeit und auch den Firmen, deren Anzeigen einen erfreulichen Beitrag zur Finanzierung der Herstellung leisten.

... und noch ein Wink mit dem Zaunpfahl: Redaktionsschluß für 1/86 ist der 2. Mai 1986. Themenschwerpunkt wird sein: "Ansätze der sprachorientierten KI". Für die Diskussion des Themenschwerpunktes im übernächsten Heft werden noch Anregungen gesucht!

G. K.

KONTEXTFREIE FORSCHUNGSFÖRDERUNG

Christa Hauenschild

Technische Universität Berlin
 Institut für Angewandte Informatik
 Projektgruppe KIT / FR 5-8
 Franklinstr. 28/29
 D-1000 Berlin 10

Eine Podiumsdiskussion und was daraus (nicht) geworden ist

Eigentlich war fuer diese Ausgabe des LDV-Forums ein Bericht von mir ueber die Podiumsdiskussion angekuendigt, die am 7.3.1985 im Rahmen der 10. Jahrestagung der GLDV in Hannover zum Thema "Verbundforschung" stattgefunden hat. Da das betreffende Ereignis inzwischen aber schon so weit zurueckliegt, werde ich den Bericht selbst relativ kurz fassen und stattdessen einiges dazu sagen, welche Folgen sich aus der Podiumsdiskussion ergeben haben bzw. nicht ergeben haben.

Teilnehmer(innen) an dem Podium waren Johannes Arz (Uni Saarbruecken), Tobias Brueckner (IdS Mannheim), Peter Canisius (GD Koeln), Christa Hauenschild (referenzidentisch mit der Autorin, damals arbeitslos in Heidelberg, inzwischen TU Berlin), Guenter Marx (BMFT), Claus-Rainer Rollinger (damals TU Berlin, jetzt IBM Stuttgart) und Nina-Jacqueline Schachter-Radig (SCS Hamburg). Moderiert wurde die Diskussion von Brigitte Endres-Niggemeyer (Fachhochschule Hannover). Diskutiert wurde auf der Grundlage des Positionspapiers "Verbundforschung - Chance fuer die Wissenschaft?" von Tobias Brueckner (abgedruckt im letzten LDV-Forum), und zwar hauptsaechlich zur Frage der Folgen, die sich aus dem Konzept der Verbundforschung (d.h. der Zusammenarbeit zwischen Industrie-Unternehmen und Hochschulen) fuer die Wissenschaft(ler(innen)) ergeben (es war tatsaechlich auch mal explizit von den -innen die Rede, allerdings in einer Weise, die das Publikum zu heftigen Reaktionen herausforderte - es ging naemlich um die mangelnde Bereitschaft der "weiblichen Kollegen", ihre Familienplanung offenzulegen).

Soweit die objektiven Fakten. Ueber die Inhalte der Diskussion kann und will ich nicht objektiv berichten, weil ich einfach zu betroffen war und bin. Vielmehr werde ich im folgenden versuchen, eine eher impressionistische Beschreibung zu geben, die aber meiner Ansicht nach subjektiv-sachlich begruendet ist.

1. Hauptindruck:

Politiker "loesen" Probleme, indem sie Zusammenhaenge leugnen.

Herr Marx vom Ministerium fuer Forschung und Technologie sagte z.B., dass sich aus der Verbundforschung ueberhaupt kaum Folgen fuer die Wissenschaft(ler) ergeben, weil fuer die erste Phase (84-88) im wesentlichen Zusammenarbeit zwischen verschiedenen Industrie-Unternehmen vorgesehen sei (tatsaechlich gibt es aber etliche Projekte mit Kooperation zwischen Hochschulen und Industrie, und die Folgen sind - soweit bisher abzusehen - ziemlich genau die, die in dem Positionspapier zur Verbundforschung und im Memorandum zur Situation der Kuenstlichen Intelligenz in der BRD (ebenfalls im letzten LDV-Forum abgedruckt) beschrieben wurden).

Bezüglich der Folgen fuer die Verbundforschung angesichts der Tatsache, dass im wesentlichen kurzfristige (um nicht zu sagen kurzatmige) Projekte gefoerdert werden, meinte Herr Marx, die Verbundforschung sei doch additiv und koenne schon deshalb keinen negativen Einfluss auf die Grundlagenforschung haben. Als daraufhin Christopher Habel aus dem Publikum bemerkte, dass die wesentlichen Leute im wesentlichen gerade anwesend seien, schien Herr Marx immerhin nachdenklich zu werden. Aehnlich war seine Reaktion (soweit feststellbar) auf die Bemerkung von Frau Schachter-Radig zur Ausbildung des Nachwuchses: ob das die Industrie machen sollte, wenn die Hochschulen wegen Mangels an qualifizierten Kraefte n nicht mehr dazu in der Lage seien?

Auch zwischen dem Konzept der Verbundforschung und der Zunahme von Zeitvertraegen fuer die wissenschaftlichen Mitarbeiter gibt es nach Ansicht von Herrn Marx und Herrn Canisius kaum Zusammenhang. Marx dazu: "Eine Drittmittelstelle ist doch keine Existenzgrundlage." Nur wenn man zufaellig von so einer Stelle lebt, ist es eben doch in einem gewissen Sinne eine Existenzgrundlage, wie Katharina Morik richtig bemerkte. Wenn es tatsaechlich keinen Zusammenhang zwischen kurzfristigen Verbundprojekten und der Zunahme von Zeitvertraegen gibt (die ja durch die neue Arbeits- und Hochschulrahmen-Gesetzgebung explizit angestrebt wird), dann ist das m.E. in erster Linie ein Zeichen fuer mangelnde Koordination zwischen den verschiedenen Bundesministerien.

Womit wir beim 2. Hauptindruck waeren: Politiker "loesen" Probleme, indem sie auf ihre Nicht-Zustaendigkeit verweisen. Aeusserungen wie "das ist Sache der Hochschulen" oder "da muessen Sie sich schon an Ihre Landtagsabgeordneten wenden" kamen in der Podiumsdiskussion haufig vor. Herr Marx versteht z.B. die Hochschulen nicht, dass sie ihre Leute nicht auf Dauer einstellen: "Die Industrie hat ihre Auftragsbuecher doch auch nicht bis zum Jahr 2000 voll." Wie wahr!

Allerdings scheint es nicht die Aufgabe der Verantwortlichen fuer einen Teilbereich zu sein, sich gelegentlich mit den Verantwortlichen fuer einen anderen Teilbereich zusammenzusetzen, um Massnahmen und Folgen aufeinander abzustimmen. Obwohl doch z.B. "die Bundesregierung in gewissem Sinne eine Einheit ist - oder nicht?" (K. Morik). Manchmal gewinnt man den Eindruck, dass die Politiker ein ausgekluegeltes zirkulaeres Verweissystem fuer Zustaendigkeiten entwickelt haben, das die Betroffenen im Kreis herumfuehrt (naetuerlich war es wirklich bedauerlich, dass kein Vertreter der Laenderparlamente anwesend war, wie Herr Canisius richtig bemerkte).

Kurz und boeswillig gesagt: die Struktur der Argumente vom BMFT war etwa die folgende: "Die von Ihnen befuerchteten Folgen koennen aus zwei Gruenden nicht eintreten: erstens sind sie im Konzept nicht vorgesehen, woraus folgt, dass sie mit den Massnahmen in keinem direkten Zusammenhang stehen; zweitens fallen sie nicht in unser Ressort."

Ebenso bedauerlich wie die Nicht-Anwesenheit von Landtagsabgeordneten bei der Podiumsdiskussion ist die Tatsache, dass keine Vertreter der Medien anwesend waren. Allerdings hat sich im Verlauf der Geschichte gezeigt, dass die Medien an der Problematik auch nicht besonders interessiert sind; sie berichten lieber ueber die tollen Erfolge, die die Bundesrepublik im Technologie-Wettlauf mit Japan und den USA erzielt. Eine Meldung, die im Zusammenhang mit dem German Workshop on Artificial Intelligence Ende September in Dassel (vgl. den Bericht von S. Busemann) an dpa gegeben wurde, wurde von der Presse anscheinend vollstaendig ignoriert. Auch auf das Memorandum zur Lage der KI, das inzwischen mit ca. 150 Unterschriften (hauptsaechlich von Vertreter(inne)n des Mittelbaus als den Haupt-Betroffenen) an das Bundesministerium fuer Forschung und Technologie sowie an das Bundesministerium fuer Bildung und Wissenschaft und an verschiedene Presseorgane geschickt wurde, hat bislang nur wenig Echo gefunden:

lediglich vom BBW kam eine Antwort, die wieder nach dem Strickmuster der Verweistechnik abgefasst ist (als neues Element kommt hinzu, dass Probleme dadurch geloest werden, dass man in besonders eklatanten Einzelfaellen von bestimmten Regelungen (wie z.B. der Gehaltsabsenkung bei Neueinstellungen) absieht); ausserdem zeigte sich das "Computer Magazin" interessiert (dort wurde fuer Dezember ein Beitrag angekuendigt). Der Hauptadressat BMFT hat sich bislang nicht geaeussert; schliesslich ist ja auch - wie wir bei der Podiumsdiskussion gesehen haben - alles in bester Ordnung (zumindest solange man von allen stoerenden Kontextfaktoren abstrahiert - dafuer haben alle, die sich z.B. mit den Komplexitaeten der natuerlichen Sprache herumschlagen, sicher vollstes Verstaendnis).

dpa-Meldung --- dpa-Meldung --- dpa-Meldung

FORSCHUNGSPOLITIK DER BUNDESREGIERUNG VERFEHLT

Harte Kritik uebten auf der Fachtagung fuer Kuenstliche Intelligenz (KI) der Gesellschaft fuer Informatik in Dassel vom 23.-27.9.85 weit ueber 100 hochqualifizierte Wissenschaftler an der Forschungspolitik der Bundesregierung. Die Regierungen der USA, Japans und Europa halten die KI fuer eine Schlüsseltechnologie der Zukunft (Computer der fuenften Generation). Also foerdert auch der Bundesminister fuer Forschung und Technologie (BMFT) Forschungsprojekte der KI im Rahmen seiner Konzeption zur Foerderung der Entwicklung der Mikroelektronik, der Informations- und der Kommunikationstechniken.

Die uebliche Laufzeit der gefoerderten Projekte betraegt drei bis vier Jahre. Wissenschaftler in einem derart komplexen Gebiet erreichen nach einer Einarbeitungsphase von drei bis vier Jahren die Hoehe ihrer Produktivitaet. Da sie nach Ablauf eben dieser Zeit aufgrund des Hochschulrahmengesetzes nicht in Anschlußprojekten an derselben Universitaet beschaeftigt werden duerfen, wird die muhsam aufgebaute Forschungskapazitaet planmaeßig wieder zerstoeert.

Auch die Ausbildung kuenftiger Fachleute ist keine Angelegenheit, die sich in Drei- bis Vierjahresplaenen erledigen laeßt. Die an Projekten beteiligten Wissenschaftler tragen naemlich, auch wenn sie nicht Professoren sind, einen erheblichen Teil der Studentenausbildung.

In einem Memorandum, das unmittelbar nach der Fachtagung an den BMFT geschickt wurde, warnen die unterzeichnenden Wissenschaftler/innen davor, ueberproportional anwendungsnahe Forschung zu foerdern und die langfristig wichtige Grundlagenforschung an den Universitaeten auszutrocknen. Auf das Fehlen der Benutzer- und Wirkungsforschung im Foerderprogramm sowie auf die Nachteile, die auch fuer die KI aus den Kuerzungen der Gesellschafts- und Geisteswissenschaften folgen, wird hingewiesen. Neue Technologien ohne Beruecksichtigung der von ihnen betroffenen Benutzer zu entwickeln ist genauso falsch wie Forschung ohne Beruecksichtigung der Arbeitsvoraussetzungen zu foerdern.

Demgegenueber fordern die Wissenschaftler normale, unbefristete und tariflich abgesicherte Arbeitsvertraege ohne Altersbegrenzung, die es auch an Universitaeten ermoeglichen, mehrere Projekte mit demselben erfolgreichen Team durchzufuehren, kontinuierlich mit anderen Fachbereichen zusammenzuarbeiten und Studenten zu betreuen.

**ANTWORT
DES BUNDESMINISTERIUMS FÜR BILDUNG UND WISSENSCHAFT**

DER BUNDESMINISTER
FÜR BILDUNG UND WISSENSCHAFT

IV B 5 - 4810

Bei Antwortschreiben bitte dieses Geschäftszeichen angeben

Tel. (0228) 57-2234 Bonn, 4. 11. 1985
III/42/h

Der Bundesminister für Bildung und Wissenschaft Postfach 200108 5300 Bonn 2

Herrn
Stephan Busemann
Technische Universität Berlin
Skr. FR 5 - 8
Franklinstraße 28/29

1000 Berlin 10

Sehr geehrter Herr Busemann!

Haben Sie vielen Dank für Ihr Schreiben vom 8. Oktober, mit dem Sie ein Memorandum zur Situation der Künstlichen Intelligenz in der Bundesrepublik Deutschland übersenden. Frau Bundesminister Dr. Wilms hat mich gebeten, Ihnen zu antworten.

Zu den im Memorandum angesprochenen Themenkomplexen möchte ich im folgenden Stellung nehmen.

Soweit für bestimmte von Ihnen angesprochene Fragen - wie z. B. Fachprogramm zur Förderung der Informations- und Kommunikationstechniken, Förderung der Verbundforschung u. ä. - das Bundesministerium für Forschung und Technologie innerhalb der Bundesregierung federführend zuständig ist, gehe ich davon aus, daß Sie sich auch an dieses Bundesressort gewandt haben.

Aus der Sicht des Bundesministeriums für Bildung und Wissenschaft finden Sie im folgenden eine Stellungnahme, die Ihnen vielleicht zur Information der Unterzeichner von Nutzen ist.

Dem prinzipiell unbegrenzten Bedarf der Forschung und Wissenschaft an materiellen Ressourcen stehen leider nur begrenzte Mittel des Staates

gegenüber. Diese Diskrepanz besteht auch in Bezug auf den wissenschaftlichen Nachwuchs, wobei hier als zusätzliches Moment das Bedürfnis nach sozialer Sicherheit und langfristigen Berufsperspektiven der Nachwuchswissenschaftler einerseits und das Interesse der Wissenschaft - in ihrer organisierten Form - nach nur zeitlich begrenzter Mitarbeit jüngerer Wissenschaftler in konkreten Forschungsprojekten andererseits hinzukommt. Dabei verkenne ich nicht, daß einige ihrer Vorschläge hier durchaus geeignet sein können, die Qualität der Forschung zu verbessern. Sie zu verwirklichen, ist aber bezüglich der Stellenausstattung der Hochschulen ausschließlich Sache der Länder, bezüglich der konkreten Organisation von Forschung Sache der Hochschulen und der Wissenschaft selbst. Bei der Förderung des wissenschaftlichen Nachwuchses, die ich durchaus als eine gesamtstaatliche Aufgabe ansehe, leistet der Bund teils allein, teils zusammen mit den Ländern und anderen Organisationen Erhebliches. Beispielhaft möchte ich auf das neue Postdoktoranden-Programm verweisen, das von der Deutschen Forschungsgemeinschaft abgewickelt wird und für das das Bundesministerium für Bildung und Wissenschaft nach 5 Mio. DM im laufenden Jahr ab 1986 jährlich 15 Mio. DM bereitstellt.

Arbeitsrechtliche Risiken haben den Abschluß von befristeten Arbeitsverträgen in der Vergangenheit zunehmend erschwert. Die Folge dieser Schwierigkeiten war eine wachsende Zurückhaltung bei den Hochschulen, befristete Arbeitsverträge mit wissenschaftlichen Mitarbeitern abzuschließen. Die Alternative für den wissenschaftlichen Nachwuchs heißt in vielen Fällen nicht "Zeitvertrag oder Dauerstelle", sondern "Zeitvertrag oder überhaupt keine Beschäftigung in der Wissenschaft". Die in das Hochschulrahmengesetz (HRG) eingefügten Vorschriften des neuen Zeitvertragsgesetzes schaffen für die Hochschulen die für ihre wissenschaftliche Arbeit erforderliche personelle Flexibilität und mehr Rechtssicherheit beim Abschluß von befristeten Verträgen mit wissenschaftlichen Mitarbeitern. Das Zeitvertragsgesetz schafft zusätzliche Möglichkeiten zum Abschluß von Zeitverträgen, indem mehrere Sondertatbestände vorgesehen sind, die eine Befristung von Arbeitsverträgen im Bereich der Hochschulen und außeruniversitärer Forschung begründen. Eine Verpflichtung zum Abschluß von befristeten Arbeitsverträgen ist diesem Gesetz aber nicht zu entnehmen. Zum Schutz der Arbeitnehmer gegen eine sozial unvertretbare Ausdehnung der befristeten Beschäftigung legt § 57 c HRG bestimmte zeitliche Höchstgrenzen für befristete Arbeitsverträge fest. Diese Höchstgrenzenregelung soll sicherstellen, daß

die befristete Beschäftigung eines wissenschaftlichen Mitarbeiters an derselben Hochschule grundsätzlich nur so lange erfolgt, bis eine abschließende Beurteilung seiner Qualifikation und seiner Eignung für das Aufrücken in eine höherwertige Funktion möglich ist. Um Schwierigkeiten zu begegnen, die sich aus diesen Höchstgrenzen bei Forschungsvorhaben ergeben könnten, die von vornherein auf einen längeren Zeitraum angelegt sind, sieht das HRG in bestimmten Sonderfällen eine Auflockerung dieser Höchstgrenzenregelung vor (z. B. Nichtanrechnung des sog. Eingangsvertrages und von Zeiten zur Vorbereitung einer Promotion). Diese in das HRG eingefügten Vorschriften des Zeitvertragsgesetzes enthalten somit eine flexible Regelung, die ausreichende Möglichkeiten zur Durchführung von Anschlußprojekten mit denselben Mitarbeitern oder mit demselben Forschungsteam einräumen.

Im Memorandum wird auch die sogenannte Absenkung der Eingangsbezahlung im Hochschulbereich angesprochen. Dem Bundesministerium für Bildung und Wissenschaft sind die damit zusammenhängenden Probleme durch zahlreiche Stellungnahmen sowie Eingaben bekannt. Deshalb hat sich das Bundesministerium für Bildung und Wissenschaft mit Nachdruck dafür eingesetzt, Ausnahmen zu ermöglichen. Der Gesetzgeber hat im Dezember 1984 dieser besonderen Problematik im Hochschulbereich durch Einfügung eines Absatzes in § 19 a Bundesbesoldungsgesetz Rechnung getragen. Danach können im Einzelfall Hochschulbeamte auf Zeit von der Absenkung der Eingangsbezahlung ausgenommen werden, sofern dies zur Gewinnung geeigneter Bewerber dringend erforderlich ist. Ausnahmemöglichkeiten bestehen auch für Auslandsstipendiaten. Für den Tarifbereich des Bundes und der Länder sind entsprechende Regelungen getroffen worden.

Einer der Schwerpunkte der Novellierung des HRG ist die Erleichterung der Forschung mit Drittmitteln. Denn Drittmittel sind für die Leistungsfähigkeit der Hochschulforschung von wesentlicher Bedeutung. Sie müssen eingeworben werden und fördern damit den wissenschaftlichen Wettbewerb. Durch diese Neuregelung im HRG soll insbesondere die Initiative und Bereitschaft von Hochschulen und Hochschullehrern gestärkt werden, Drittmittel einzuwerben. Zugleich soll aber auch die Bereitschaft der Wirtschaft gestärkt werden, das Forschungspotential der Hochschulen im Hinblick auf ihren gesteigerten Bedarf an Forschung zur Lösung wissenschaftlicher Aufgaben zu nutzen. Dabei geht es auf seiten der Wirtschaft aber nicht nur um die großen, sondern vor allem auch um die kleinen und mittleren Unternehmen. Ausweislich der Angaben im Bundes-

bericht Forschung 1984 und im Drittmittelbericht der Bundesregierung erhielten z. B. in 1982 die Hochschulen rd. die Hälfte aller Drittmittel von der Deutschen Forschungsgemeinschaft. Um die Größenordnung der Drittmittel aus der Wirtschaft zu verdeutlichen, ist darauf hinzuweisen, daß von den rd. 1,6 Mrd. DM, die die Hochschulen als Drittmittel zur Förderung von Forschung und Entwicklung z. B. in 1982 erhalten haben, nur rd. 120 Mio. DM unmittelbar aus der Wirtschaft stammen.

Leistungs- und Wettbewerbsfähigkeit unserer Wirtschaft lassen sich nur erhalten, wenn die Möglichkeiten, praxisbezogene Fragestellungen an Lehre und Forschung heranzutragen, intensiviert werden. Ebenso wichtig ist die schnelle Verwendbarkeit der Ergebnisse von Forschungs- und Entwicklungsvorhaben außerhalb der Hochschulen. Dabei muß vor allem kleinen und mittleren Unternehmen, die nur in sehr begrenztem Umfang selbst Forschung betreiben können, ein verbesserter Zugang zu neuen wissenschaftlich-technischen Entwicklungen eröffnet werden. Die Verstärkung der Drittmittelforschung und die Kooperation zwischen Hochschule und Wirtschaft kommen der Wissenschaft in gleichem Maße zugute. Ohne Drittmittel kann namentlich in den experimentell und empirisch arbeitenden Disziplinen heute kaum noch anspruchsvolle Forschung stattfinden. Außerdem ist technologischer Fortschritt in Praxis und Forschung nicht denkbar ohne einen kontinuierlichen Informations- und Wissensaustausch zwischen diesen beiden Bereichen. Eine Überbetonung der Anwendungsforschung bei gleichzeitiger Vernachlässigung der Grundlagenforschung ist dabei aus der Sicht der Bundesregierung nicht zu befürchten. Sie befürwortet und unterstützt es zwar, daß der Anwendungsbezug der Hochschulforschung künftig stärkere Berücksichtigung findet. Damit wird aber lediglich ein Defizit ausgefüllt, wie die oben dargelegten Zahlen zeigen. Es geht also nicht um Gewichtsverschiebungen zu Lasten der Grundlagenforschung, sondern nur darum, der angewandten Forschung den ihr zukommenden Platz einzuräumen. Die Grundlagenforschung bleibt nach wie vor eindeutiger Schwerpunkt der Hochschulforschung.

Mit freundlichen Grüßen

Im Auftrag



Schacker

Künstliche Intelligenz bei Nixdorf

Als einer der international führenden Computer-Hersteller haben wir frühzeitig unsere Organisation auf die Bedürfnisse und Anforderungen unserer Kunden und Interessenten ausgerichtet. Der Erfolg gibt uns recht. Unser Umsatz wächst jährlich um ca. 20 Prozent. Wir bauen unsere Marktposition ständig weiter aus und werden auch in Zukunft systematisch neue Märkte erschließen. Einer dieser neuen Märkte ist der Markt der **Expertensysteme** (Künstliche Intelligenz). Dafür suchen wir weitere Mitarbeiter für unsere Entwicklungsmannschaft in Paderborn, die als

Computer-Linguist Expertensysteme

eingesetzt werden sollen.

Ihre Schwerpunktaufgaben:

- Entwurf und Implementierung von Programmen, um den Ausbau der Erklärungskomponente von Expertensystemen entscheidend voranzutreiben.
- Mitarbeit am vom BMFT geförderten WISBER-Projekt

Ihre Qualifikationen:

- Abgeschlossenes Hochschulstudium bzw. Promotion im Bereich der Linguistik (idealerweise mit Nebenfach Informatik)
- Kenntnisse in höheren Programmiersprachen (PROLOG, C, UNIX)
- Gutes Englisch in Wort und Schrift
- Möglichst Erfahrung als Teacher und Kenntnis in intelligenten, computerunterstützten Unterweisungen
- Spezifische Kenntnisse über Definite Clause Grammars und kaskadierte Augmented Transition Networks
- Erfahrung in KI-Methoden und Implementierungen

Eine spezielle mehrmonatige Ausbildung, u. a. durch unsere Entwicklungsgruppe Expertensysteme, geht Ihrem Einsatz voraus. Wenn Sie dieses neuartige und zukunftsreiche Arbeitsgebiet mit seiner faszinierenden Technologie reizt, freuen wir uns auf Ihre Zuschrift.

Für telefonische Vorabinformationen steht Ihnen Herr Dr. Savory bzw. Frau Dr. Rösner unter Tel. 0 52 51 / 14 - 61 08 zur Verfügung.

Nixdorf Computer AG
Personalbereich Entwicklung
z. H. Herrn Kurt Pape
Pontanusstraße 55, 4790 Paderborn

NIXDORF
COMPUTER

EXEMPLARISCHES ZUR PRAXIS DER TEXTVERARBEITUNG AN WISSENSCHAFTLICHEN INSTITUTIONEN

B. Böhm, J. Krause

Universität Regensburg
Fachgebiet Linguistische Informations-
wissenschaft

Universitätsstraße 31
D8400 Regensburg 1

1 Ausgangssituation und Rahmenbe- dingungen

Neue Bürotechnologien werden in Zukunft auch die Arbeitssituation und die Arbeitsgestaltung in den Sekretariaten wissenschaftlicher Hochschulen und vergleichbarer wissenschaftlicher Einrichtungen verändern. Rahmenbedingungen und Ausgangslage unterscheiden sich dabei deutlich vom kommerziellen Bereich, dem bisher das Hauptaugenmerk galt.

Lehrstühle an Hochschulen haben drei Aufgaben: Lehre, Forschung und Verwaltung. Die Organisationsstruktur ist in der Regel streng hierarchisch: die Professoren werden jeweils durch wissenschaftliche Mitarbeiter, die Sekretärin und studentische Hilfskräfte unterstützt. Die arbeitsteilige Zuordnung der Mitarbeiter eines Lehrstuhls untereinander und zum Professor, und damit die von ihnen ausgeführten Arbeitsinhalte differieren zum Teil von Lehrstuhl zu Lehrstuhl sehr stark. Gemeinsam ist allen, daß im Rahmen der Aufgabenverteilung zwischen Wissenschaftlern und Sekretärin die Sekretärin neben der Schreibfähigkeit in mehr oder weniger

großem Umfang Verwaltungsaufgaben und Assistenzfähigkeiten für Lehre und Forschung übernimmt. Letztere (z.B. Folien vorbereiten, Statistiken anlegen, Bücher suchen) treten allerdings heute in der Regel deutlich vor den Schreib- und Verwaltungstätigkeiten zurück. Diese Entwicklung verstärkt sich noch dort, wo die Ganztags-Sekretärin im Zuge der Sparmaßnahmen der letzten Jahre durch eine Halbtagskraft ersetzt bzw. eine Sekretärin mehreren Professoren zugeordnet wurde. Besonders im geisteswissenschaftlichen Bereich steigt dadurch der Anteil der Schreibtätigkeiten stark an, da Professoren die Schriftguterstellung durch die Sekretärin als unverzichtbar ansehen.

In diesem Umfeld gibt es eine Reihe von Schwachstellen, bei denen der Einsatz neuer Büro- bzw. Informationstechnologien eine deutliche Leistungssteigerung verspricht, und zwar im wesentlichen in bezug auf

- a) die bessere Bewältigung des Schriftgutes mit Hilfe der Textverarbeitung (TV) und
- b) die Ausweitung und "effizientere" Gestaltung der bisher stark einge-

schränkten Informationsbeschaffung v.a. für Lehre und Forschung (z.B. Literaturrecherche) durch den Einsatz von Datenbanken.

Sicher scheint, daß zumindest der Einsatz von Textverarbeitungs- und Datenbanksoftware in naher Zukunft eine große Rolle spielen wird.

Am direktesten werden von der Einführung der neuen Büro- bzw. Informationstechnologien die Sekretariatsarbeitsplätze beeinflusst, bei denen sich die Schriftguterstellung konzentriert. Der Einsatz von TV und Datenbanken betrifft aber potentiell alle Arbeitsplätze eines wissenschaftlichen Institutes. Veränderungen auf der inhaltlichen wie auf der organisatorischen Ebene sind zu erwarten.

Da die Einführung der TV bisher am weitesten fortgeschritten ist, konzentrieren wir uns auf diesen Bereich.

2 Textverarbeitung

Textverarbeitungsanlagen, als Ersatz für Schreibmaschinen begegnen einem zusehends häufiger in den wissenschaftlichen Sekretariaten von Hochschulen.

Diese Entwicklung ist nicht unproblematisch. Zu den allgemeinen Schwierigkeiten, die die Vorteile der TV begleiten, kommen spezifische Probleme, die bei der Einführung der TV (als einer wesentlichen Untergruppe von Bildschirmarbeitsplätzen) in Industrie und Verwaltung nicht oder doch sehr verschieden auftreten. Sie beruhen zu einem nicht unerheblichen Teil auf den weitgehend inhomogenen Strukturen, wie man sie typischerweise an wissenschaftlichen Einrichtungen vorfindet, und die geprägt sind von einer Vielfalt persönlicher Konstellationen in einem ausgeprägt hierarchischen Gefüge. Von daher stehen umfassende Innovationsmaßnahmen, wie sie etwa die Einführung der sog. "organi-

sierten Textverarbeitung" darstellt, kaum zur Diskussion. Vielmehr ist zu erwarten, daß der Innovationsprozeß am erfolgversprechendsten über eine Vielzahl kleinerer Maßnahmen schrittweise vorangetrieben werden kann.

TV an wissenschaftlichen Einrichtungen hat sowohl unter funktionalen als auch unter organisatorisch-psychologischen Aspekten eigenständige Anforderungen, die zumindest teilweise ungelöst erscheinen und für deren Humanisierungspotential in den gewachsenen und institutionalisierten EDV-Strukturen an wissenschaftlichen Einrichtungen (z.B. Universitäten) bisher noch wenig Verständnis aufgebracht wird. Hier werden bei der Einführung neuer Technologien z.T. weiterhin einseitige Wirtschaftlichkeitsdefinitionen vorgenommen, die lediglich quantitative Einsparungseffekte in Teilbereichen berücksichtigen, ohne dabei einer menschengerechten Arbeitsgestaltung in gleichem Umfang Aufmerksamkeit zu widmen. Dies kann aber nur dadurch geschehen, daß zu der jeweiligen Technik auch das entsprechende Umfeld in bezug auf Arbeitsorganisation, Ergonomie des Arbeitsmittels, Qualifikationschancen sowie persönliche Motivation der Benutzer geschaffen wird.

Nicht zu leugnen ist, daß die Einführung der TV v.a. im geisteswissenschaftlichen Bereich potentiell enorme Effizienzfortschritte bietet; Wissenschaftsfelder, deren Arbeitsstil durch eine hohe Textproduktion gekennzeichnet sind, haben erstmals in größerem Umfang die Möglichkeit, die EDV als arbeitsförderndes Hilfsmittel einzusetzen.

Für die Ausführenden der anfallenden Tätigkeiten (insbesondere Sekretärinnen) kann das u.U. als weiteren, ebenso wichtigen Effekt, eine Abnahme einseitig belastender und routinisierten Arbeiten bedeuten.

Wird allerdings die Problematik des Einsatzes der TV an wissenschaftlichen

Einrichtungen nicht gesehen und nicht gelöst, besteht die Gefahr, daß Wissenschaftler in den nächsten Jahren inhaltlich suboptimal, zu überhöhten Kosten, und Sekretärinnen bei sich verschlechternden Arbeitsbedingungen und unter zunehmenden Arbeitsbelastungen mit der TV arbeiten.

Diesen Tendenzen kann nur dadurch begegnet werden, daß die Einführung neuer Büro- und Informationstechnologien als ganzheitliche Aufgabenstellung verstanden wird, die Organisation, Technik und Bedürfnisse der Anwender als gleichwertige Komponenten betrachtet.

Die Erfahrungen, die bei den ersten - noch sehr zufälligen - Installationen von Mikrorechnern für die Textverarbeitung an der Universität Regensburg gemacht wurden, haben gezeigt, daß hier generelle Probleme auftreten, die funktionale Aspekte genauso betreffen wie ergonomische und organisatorische.

Die Beobachtungen der Pilotstudie TV an der Universität Regensburg sind in Böhm/Krause/Womser 1985 beschrieben. Hier sollen nur einige generelle Punkte herausgegriffen und ein zusammenfassendes *Résumé* gezogen werden.

3 Pilotstudie TV

3.1 Ausgangslage

Die Pilotstudie umfaßt den Zeitraum Herbst 1983 bis Sommer 1985. Zu Beginn wurde ein Sekretärinnen-TV-Arbeitsplatz im Fachbereich Sprach- und Literaturwissenschaften für die Allgemeine Sprachwissenschaft, die Anglistik und die Romanistik eingerichtet (8-bit Rechner KBS mit VISUAL 200 Bildschirmen). Als TV-Software diente WordStar 3.0. Sowohl die Konfiguration als auch die Software waren die gleichen wie der bisherigen EDV-Nutzer. Der Bildschirm wurde im November 1984 durch den ergonomisch

besseren TE 77 der Firma Schoembs ersetzt. Das Rechenzentrum der Universität hatte hierfür Zusatzsoftware entwickelt, die es ermöglichte, durch vier programmierbare Zeichensätze beliebige Sonderzeichen auf dem Bildschirm darzustellen und über den Matrix-Schönschreibdrucker NDK wiederzugeben. Darüberhinaus kamen drei weitere TV-Arbeitsplätze hinzu, an denen nun auch die Slavisten, Germanisten und Klassischen Philologen arbeiteten (insgesamt 10 Lehrstühle).

Die in der Pilotstudie gemachten Beobachtungen beziehen sich v.a. auf diese Arbeitsplätze. Hinzu kamen Vergleiche mit einem Siemens-Textsystem und einem von IBM, die hier jedoch ausgeklammert bleiben.

3.2 Schwerpunktverlagerung

Mitte der 70er Jahre wurde an der Universität Regensburg das ROBUR-Konzept entwickelt. Es sieht eine zentrale Beschaffung, Verwaltung und Betreuung aller Rechnerkapazitäten der Universität vor. Die bisherige Ausstattung mit Mikrorechnern kam vorwiegend Bereichen zugute, die bereits EDV-Kapazitäten für ihre traditionellen EDV-Aktivitäten einsetzten (v.a. naturwissenschaftliche Einrichtungen). Dies erschien zumindest zunächst dadurch gerechtfertigt, daß der EDV-Einsatz z.B. im geisteswissenschaftlichen Bereich sich bis vor kurzem in engen Grenzen hielt.

Die Einführung der TV an Sekretariatsarbeitsplätzen führt jedoch zu einem erheblichen Nachholbedarf an EDV-Kapazitäten v.a. im geisteswissenschaftlichen Bereich. Die Vorteile der TV gegenüber der herkömmlichen Schriftguterstellung sind hier, u.a. bedingt durch eine umfangreiche Textproduktion, deutlich höher als in den Naturwissenschaften. Deshalb muß mit einer Schwerpunktverlagerung gerechnet werden; zumindest bei

den Mikrorechnern dürften in Zukunft Geisteswissenschaftler (besser die Gruppe der Endbenutzer) den Anspruch erheben, in gleichem Umfang wie Naturwissenschaftler an den EDV-Kapazitäten zu partizipieren.

Neben unbestreitbaren Vorzügen bringt diese Entwicklung substantiell neue Probleme mit sich. Von zentraler Wichtigkeit ist, daß damit verstärkt EDV-unkundige Personen Benutzer von TV werden (sog. Endbenutzer). Die bisherigen Anwender der TV besaßen in der Regel weitergehende EDV-Kenntnisse oder waren zumindest über Kollegen in ein EDV-kundiges Umfeld eingebunden. Erst wenn anfallende Fragen mit diesem Wissen nicht mehr gelöst werden konnten, wandte man sich an das Rechenzentrum. Da der Benutzerkreis zudem überwiegend aus Wissenschaftlern, Diplomanden oder Studenten bestand, entwickelte sich zwischen dem Rechenzentrum, dessen Beratung ebenfalls von Akademikern durchgeführt wird, und den Benutzern eine relativ homogene Kommunikationsstruktur. Diese Homogenität wurde noch dadurch unterstützt, daß die fast ausschließlich naturwissenschaftlich ausgebildeten Mitarbeiter des Rechenzentrums mit den Naturwissenschaftlern als dominierender Benutzergruppe keine schwerwiegenden Verständnisprobleme hatten. Diese Situation hat sich spätestens mit der Einführung der TV in den Verwaltungs- und Sekretariatsbereichen grundlegend gewandelt. Der neue Benutzerkreis verfügt in den wenigsten Fällen über EDV- oder TV-Kenntnisse. Die Sekretärinnen sind zwar bereit, sich intensiv mit der Textverarbeitung zu befassen, besitzen jedoch in der Regel nicht die Absicht oder den Wunsch, sich darüberhinaus mit der EDV auseinanderzusetzen. Sekretärinnen sehen die TV als vielseitiges und arbeitserleichterndes Hilfsmittel zur Erledigung ihrer Arbeit, mit dem sie eine lange Zeit des Tages (durchschnittlich 3-4 Stunden) beschäftigt sind. Dies prägt deutlich

ihre Einstellung zur TV und hebt sie vom bisherigen Benutzerkreis ab. Der geisteswissenschaftliche Bereich unterscheidet sich dabei in der Regel dadurch vom naturwissenschaftlichen, daß in der näheren Umgebung der Sekretärinnen zumindest heute noch selten Personen mit EDV-Erfahrung anzutreffen sind. Dabei wäre gerade in diesem Einsatzfeld eine beratende Unterstützung der Anwender notwendig. So hat man es in den Sprach- und Literaturwissenschaften aufgrund des schwierigen Textmaterials mit einer äußerst komplizierten Form der TV zu tun (zahlreiche Sonderzeichen, mehrsprachige Texte und Fußnotenverarbeitung u.ä.). Das Rechenzentrum stellt in dieser Situation bei den bisher vorliegenden Strukturen keinen geeigneten Ansprechpartner dar. Eine auf EDV-Vorkenntnisse hin konzipierte Kommunikationsstruktur bringt es u.a. mit sich, daß EDV-unkundige Benutzer auf ihre oft recht grundlegenden Fragen keine zufriedenstellende Antwort erhalten. V.a. zu Beginn ihrer Tätigkeit am Textgerät sind ihre Anliegen von andersartigen Voraussetzungen geprägt: die Einarbeitungszeit ist länger als bei Benutzergruppen mit Vorkenntnissen, und eine fundamentale Unsicherheit dem neuen Arbeitsmittel gegenüber bleibt über größere Zeiträume hinweg bestehen.

Es kann nun nicht darum gehen, die durch die Hereinnahme des Problemfeldes Textverarbeitung in das EDV-Angebot der Universität Regensburg aufgetretenen Schwierigkeiten durchweg der Trägerstruktur als "Fehler" oder "Versagen" einzelner Personen oder Gruppen anzulasten. Vielmehr muß der TV-Bereich und die dazugehörige neue Benutzergruppe in die bestehenden EDV-Strukturen integriert werden. Dabei ist zu gewährleisten, daß einerseits der bisherige Benutzerkreis effektiv weiterarbeiten kann, andererseits den spezifischen Bedürfnissen der neuen Anwendergruppe entsprochen wird. Veränderungen - darunter

auch kostensteigernde - sind bei dieser Konstellation nicht vermeidbar. So stießen z.B. die bis November 1984 verwendeten Bildschirme (Visual 200), die an den bisherigen EDV-Arbeitsplätzen von wissenschaftlichen Mitarbeitern oder Studenten klaglos akzeptiert wurden, bei den TV-Arbeitsplätzen wegen der ungenügenden Qualität der Zeichendarstellung auf Ablehnung. In ähnlicher Weise werden einzelne Programmpakete oder TV-Funktionen, die sich beim Einsatz durch EDV-sachkundige Anwender durchaus bewährt hatten, in der neuen Benutzergruppe abgelehnt, weil ihre Handhabung zu kompliziert erscheint.

3.3 Funktionale, ergonomische und organisatorische Schwierigkeiten

Die Probleme, die in allen diesen Bereichen auftraten, waren überraschend vielfältig und störten in ihrer Gesamtheit den Einsatz der TV erheblich (cf. Böhm/Krause/Womser 1985).

Im funktionalen Bereich irritierte v.a. die Kumulation kleinerer Fehler (z.B. eine Unterstreichung, die zu knapp unter den Buchstaben saß, so daß u.a. "g" und "q" nicht mehr unterschieden werden konnten), die Unmöglichkeit, Zusatzsoftware z.B. zur Fußnotenverarbeitung effektiv zu integrieren und das Fehlen einer praktikablen Ausdrucksmöglichkeit über Laser-Drucker, der für Publikationen unerlässlich ist.

Die Ergonomieverstöße reichten zu Beginn der Studie von zu hohen Tischen bis zu Druckern, die einen Lärmpegel von 82,3 db(A) erzeugten. Hier konnten am wirkungsvollsten Verbesserungen erreicht werden.

Auf der organisatorischen Seite erwiesen sich Schulungs-, Wartungs- und Betreuungprobleme als schwerwiegend:

a) Relativ unabhängig von der Qualität der einführenden Schulung treten vie-

le Probleme mit der TV erst in der Einarbeitungsphase nach dem Schulungskurs auf. Je nach dem Schwierigkeitsgrad des Textmaterials kann diese Einarbeitungsphase recht lange Zeit in Anspruch nehmen (etwa bis zu drei Monate), zumal wenn später komplexere TV-Funktionen und evtl. die Benutzung von Zusatzsoftware erlernt werden müssen. Von daher ist wichtig, daß eine Nachbetreuung angeboten wird. Sind diese Voraussetzungen erfüllt, kann davon ausgegangen werden, daß die Motivation der Benutzer durchwegs hoch ist. Befürchtungen, daß Sekretärinnen den TV-Arbeitsplatz ablehnen, bzw. ihre Ablehnung mit überzogenen Forderungen an die TV kaschieren oder generell eine Verweigerungshaltung einnehmen, haben sich nicht bestätigt.

b) Die Beratung für TV direkt im Rechenzentrum erwies sich als problematisch. Aufgrund negativer Erfahrungen wird sie von den meisten Sekretärinnen im geisteswissenschaftlichen Bereich als Anlaufstelle bisher nicht akzeptiert. Typische Stellungnahmen und Erfahrungen sind: Auf Probleme der Sekretärinnen wird nicht in der für sie erforderlichen Weise eingegangen; sie spüren deutlich, daß ihre Anliegen als unwichtig gegenüber denen von EDV-Anwendern betrachtet werden.

Zurückzuführen ist dies u.a. auf die in Abschnitt 2 angeführte Ausgangssituation: der Beratungsdienst des Rechenzentrums orientierte sich bisher an den Bedürfnissen und Erfordernissen der traditionellen EDV-Nutzer und kommt mit der neuen Benutzergruppe vorerst noch schlecht zurecht. Untypisch sind Probleme des Rechenzentrums-personals im Umgang mit Endbenutzern nicht. Vielerorts wird deshalb empfohlen, wenigstens zum Teil auf externe Stellen auszuweichen (cf.

z.B. Grochla et al. 1981, 375).

- c) Ein weiteres schwerwiegendes Problem ist die Wartung bei Gerätedefekten. Die Geräte sind - gemessen an den Bedürfnissen der Benutzer - durchwegs zu lange in der Wartungsstelle: unter drei Tagen kann unserer Erfahrung nach mit einer Reparatur nicht gerechnet werden, der Durchschnitt liegt eher höher. Gerade bei Mehrbenutzerarbeitsplätzen in der TV hat aber fast immer ein Benutzer unaufschiebbare Schreibearbeiten zu erledigen. Ein Ausfall der Maschine für mehrere Tage wird deshalb nicht toleriert; er kann nicht toleriert werden, weil Terminarbeiten liegen bleiben (Reaktion: "mit der Schreibmaschine wäre mir so etwas nicht passiert"). Wegen der geringen Anzahl der Geräte in den Geisteswissenschaften ist derzeit ein Ausweichen auf einen anderen Arbeitsplatz praktisch nicht möglich.

4 Fazit

Die Pilotstudie TV zeigte, daß die mit der TV verbundenen Probleme in der Praxis - v.a. in Bereichen der Geisteswissenschaften, in denen Standard-TV nicht ausreicht - dazu neigen, zu einer Gesamtsituation zu führen, in der die TV deutlich suboptimal eingesetzt wird. In Teilbereichen des TV-Einsatzes an der Universität Regensburg überwogen die Nachteile so stark, daß von einer sinnvollen Nutzung der TV nicht mehr gesprochen werden konnte. Vor allem zeigte sich, daß bei der Einführung psychologische Aspekte eine wichtige Rolle spielen; man muß die Ängste und die Wünsche der Betroffenen ernst nehmen (auch wenn sie nicht immer problemadäquat vorgebracht werden), will man die bei der Einführung des neuen Arbeitsmittels auf-

tauchenden Probleme bewältigen. Richtig ist, daß neuere Entwicklungen, v.a. auf dem Ergonomiesektor, durchaus Lösungsmöglichkeiten für einen Großteil der bestehenden Probleme bieten. Der Kostenzuwachs gegenüber einer geplanten Standardlösung kann jedoch erheblich werden. Ähnliches gilt für organisatorische Maßnahmen; diese müssen an bestehende Strukturen anknüpfen, der finanzielle Aufwand muß tragbar bleiben. Eine unkritische Übernahme von Maximalpositionen löst TV-Probleme an wissenschaftlichen Einrichtungen mit ihren beschränkten Möglichkeiten in der Regel nicht, verhindert vielmehr die Einführung der TV auf breiter Ebene. In gleicher Weise negativ wirken andererseits vordergründig "billige" Lösungen bzw. das Festhalten an vorgegebenen Organisationsstrukturen, die an den Bedürfnissen der Betroffenen vorbeigehen. Hier einen sinnvollen Ausgleich zu finden, scheint prinzipiell möglich, wenn auch beim gegenwärtigen Bewußtseinsstand aller Beteiligten schwierig zu erreichen.

LITERATUR

- Böhm, B., Krause, J., Womser-Hacker, C. (1985): Pilotstudie Textverarbeitung an wissenschaftlichen Einrichtungen.
- Grochla, F. et al. (1981): Handbuch der Textverarbeitung. Landsberg.

TEXT-TO-SPEECH-SYSTEMS FOR GERMAN - A SHORT SURVEY

K. Wothke

Department of Computational Linguistics
Institut für Deutsche Sprache

D-6800 Mannheim

Abstract: A text-to-speech-system can usually be divided into two main components: The first component generates for an orthographic input text its phonetic transcription. The second component maps this phonetic transcription on a sequence of control symbols for a speech-synthesizer and transmits these symbols to the synthesizer, where they trigger the production of synthetic speech. This paper gives a short survey of the first component in several text-to-speech-systems for German which were developed during the last 10 years. First the linguistic problems are outlined. Then the different types of solutions for the realization of the first component are presented and compared. Subsequently some existing systems which transcribe German text are discussed. The last part of the paper deals with the most important unsolved linguistic problems in these systems.

Outline of the problem

The first main task to be performed in a text-to-speech-system is the fully automatic conversion of an orthographic input into its phonetic transcription. Suppose the input-sentence

"The vocal cords are vibrating at about 100 Hz now."

is fed into a text-to-speech-system for English. Then this system at first has to generate the transcription:

"ðə 'vəukəl kɔ:dz ə: vai 'breitiŋ ət ə'baut 'hʌndrəd 'hɜ:ts nau"

This example demonstrates some of the problems to be solved during the automatic transcription: The system has to transcribe not only orthographic words but also abbreviations and (sequences of) digits. It must also recognize the word stress and mark it in the transcription. Furthermore there are some problems not illustrated by the example:

- transcription of special characters (% , \$, \$ etc.),
- determination of pauses of different length,
- determination of sentence stress.

Types of solution

Most of the existing solutions for the problem of automatic transcription can be divided into three main steps. Especially the second step is realized in very different ways.

The first step is a preprocessing of the orthographic input. This preprocessing scans the text for the occurrence of digits, abbreviations and special characters, and it replaces them by the corresponding orthographic words. In the systems for German these orthographic words are usually determined

- with a rule based method for the (sequences of) digits
- with a dictionary look-up for the abbreviations and the special characters.

The second step transcribes the preprocessed text on the segmental level (suprasegmentals are usually not considered in this step). Three alternative solutions exist for the transcription on the segmental level: The dictionary based method, the rule based method and a combination of these two methods.

Dictionary based systems use a dictionary whose entries have the structure

orthographic word	phonetic transcription
-------------------	------------------------

In order to transcribe an orthographic text these systems look up its words in the dictionary and replace them by the phonetic transcriptions found in the dictionary. This method guarantees that the system can always decide if a word is correctly transcribed, i.e. found in the dictionary. This is an advantage the rule based systems do not possess. A substantial drawback of the dictionary based method is that it can only transcribe those words of the text which are also entries in the dictionary. Thus this method does not take into account the fact that each natural language is an open system. The dictionary will never be complete. There are two further disadvantages of the dictionary based method:

- The creation of the dictionary requires a lot of work.
- The storage requirement does not permit implementation on a micro-computer.

A rule based system applies a set of transcription rules to the letters or letter sequences of the input text in order to transcribe them. Usually the rules are context sensitive, i.e. they transcribe the letters of the orthographic text dependent on their graphemic context. It is important to note that the term "rule" does not denote any linguistic regularities which exist between the orthographic representation and the phonetic transcription in a language. "Rule" denotes an instruction formulated in a formal language as to how to replace (strings of) letters by (strings of) phonetic symbols. The type of rule mostly used in the rule based systems for German is

u <v> w -> x.

Read: If the letter (or string of letters) v occurs in a word on the right side of the letter/string of letters u and on the left side of the letter/string of letters w, then replace v by the phonetic symbol(s) x. u and w can also be missing.

Usually a set of rules is ordered: The system tries at first to apply the more specific rules to a given letter(-sequence). If these rules cannot be applied, the more general rules are applied.

- In contrast to the dictionary based method the rule based method takes into account that each natural language is an open system. The rule based method can potentially transcribe each word of an

input text. Further advantages are:

- A set of transcription rules can usually be created with less effort than a dictionary.
- Rule based systems usually need less storage than dictionary based systems. Therefore they can also be run on micro-computers.

Rule based systems have the drawback that they cannot decide whether a word was correctly transcribed. - An important subclassification of rule based systems can be made on the level of their software architecture. The differences in the software architecture, which lead to this subclassification, have strong effects on the course of the development of a rule based system and on its transparency and changeability. Rule based systems can be subclassified into so-called programmed systems, interpreters, and compilers. In a programmed system the rules are formulated as statements of a programming language. The rules are incorporated into a source code. Interpreters read the rules from an external data set and apply them to the orthographic text. The rules have been written to the data set by a programmer or linguist in advance. For the notation of the rules the linguist can employ a problem oriented formal language (the usual representation of the rules is: $u \langle v \rangle w \rightarrow x$). Compilers first read the rules from an external data set where they were written by the user. Then they generate a load module, which can execute the transcription process described by the rules. When this load module is started, it reads the orthographic text from a data set and generates the corresponding phonetic transcription.

- Programmed systems are generally faster than interpreters. Since the rules are formulated as statements of a programming language they are often not very transparent and it is quite difficult to alter them. Furthermore, each change of the rules requires a new compilation of the source code. If an interpreter is used, it is much easier to alter the rules because they are formulated in a problem oriented notation and therefore more transparent. A change of the rules does not require a change and a new compilation of the interpreter program. An important advantage of an interpreter (and also of a compiler) is that it can be used for the automatic transcription of texts from different languages. The linguist must only provide sets of rules for the particular languages in an external data set. A compiler has the advantage that the rules can easily be changed and that the generated load module usually is faster than an interpreter. It may be considered as a drawback that each change of the rules requires their new compilation.

The third type of solution for the transcription of orthographic texts on the segmental level consists in a combination of the dictionary based method and the rule based method. Some of these systems try to compensate the incompleteness of a dictionary based system by adding a rule based component: Each orthographic word which cannot be found in the dictionary is transcribed by the application of the rules. Thus especially neologisms, which are often not in the dictionary, can also be transcribed. On the other hand, there are primarily rule based systems which try to prevent the set of rules from being overloaded with descriptions of exceptions by using a dictionary which contains words with irregular phonetic transcriptions.

The third step in the automatic transcription of orthographic text deals with the suprasegmentals. This step tries to insert into the text marks for pauses of different length, marks for the word stress and marks for the sentence stress. In the existing text-to-speech-systems for German this step shows the most problems. In particular, the automatic identification of the sentence stress is

still very defective. An established paradigm of possible solutions does not exist for this step. A remarkable study on the treatment of suprasegmentals in text-to-speech-systems for German was carried out by Zinglé (10).

Existing systems for the transcription of German text

In the following paragraphs some selected systems for the automatic transcription of German text will be presented.

A dictionary based system was developed by Chisholm (4). (This system was not used as a part of a text-to-speech-system, but for the transcription of large corpora of German verse as a preparation for a computer-assisted phonological analysis of German verse.) The dictionary comprises about 25.000 entries, i. e. orthographic words with their transcriptions (including word stress). The system correctly transcribes about 95% of the words (= tokens) of a text. Sentence stress and pauses are not taken into consideration. Further noteworthy dictionary based systems for German do not exist.

One of the first rule based systems for the automatic transcription of German text was developed in 1975/76 by Berry-Rogghe (1) at the Institut für deutsche Sprache/Mannheim. Later, this system, named PHONOL, was improved by Breuer/Brustkern/Thyssen/Willée (3) and it became part of a text-to-speech-system. PHONOL is a programmed system. The system comprises

- a preprocessing of the input (substitution of digits by the corresponding written words),
- the phonetic transcription on the segmental level,
- and the determination of the word stress and of pauses of different length.

Sentence stress is not considered by the system. PHONOL correctly transcribes 98% of the running words (= tokens) of a German text.

A rule based interpreter system was developed in the years 1977-1980 by Slaby and Spellmann (8, 9) at the University of Münster. The system consists of two steps: In the first step all digits in the text are replaced by the corresponding orthographic words. The second step transcribes the preprocessed text on the segmental level. For this process a very voluminous set of 13.370 rules is used. Suprasegmentals are not considered by the system. A statistical evaluation of the linguistic adequacy of the system was not carried out.

There exist several noteworthy compiler-systems which have been developed during the last years. Two of them are presented here.

At the Gesellschaft für Mathematik und Datenverarbeitung/Bonn B. S. Müller (6) developed a system called REDE. Compared to the other existing systems REDE takes into account most of the linguistic problems which must be considered during the automatic transcription of German text. The transcription is performed in eight steps:

1. Preprocessing of the orthographic text (abbreviations, digits, and special characters are replaced by the corresponding orthographic words).
2. Identification of unstressed words. This step is performed by looking up the words of the text in a list of words which are normally unstressed.
3. Identification of morpheme boundaries. The morpheme boundaries are recognized by scanning the words for sequences of letters which usually occur only at the end or at the beginning of a morpheme. Of all systems presented here, Müller's system is the only rule based system which tries to recognize morpheme boundaries. - The identification of morphemes should be obligatory in a text-to-speech-system for German since in German words

many sequences of letters must be replaced by other phonetic symbols if they are part of one morpheme as opposed to occurring on the boundary of two morphemes (cf. 'Busch' vs 'Werkschor')

4. Identification of prefixes and suffixes which exert an influence on the stress of the stem of a word (e. g.: 'einsetzen vs ver'setzen).
5. Marking of the word stress dependent on the results of steps 2 and 4.
6. and 7. Transcription of the vocalic letters.
8. Transcription of the consonantal letters.

As this sequence of steps illustrates, the processes involved in the automatic transcription of German text are not always executed in the order which was described under "Types of solution": In the system REDE the suprasegmentals are treated before the transcription on the segmental level is carried out. An experimental version of Müller's system also included an additional component for the identification of sentence intonation. - Müller does not present any statistical evaluations of the linguistic capacity of his system.

Another compiler-system, named RULBOL, is being developed by Bierfert (2) and Stock (5) at the University of Bonn. Up to now the main interest in the development of this system has been the creation of a comfortable formal language for the notation of the transcription rules and the construction of a compiler for sets of rules written in this formal language. The existing set of rules for the transcription of German text is only a provisional one. In the past this set of rules has mainly been used to test the compiler during its development. The most interesting aspect of RULBOL is the formal language for the formulation of the rules. Some features of this language may be illustrated with examples from the provisional set of rules for German:

- [$\langle V\emptyset \rangle$, l, r] G [$\langle V1 \rangle$, L, R] -> g

Read: If the letter 'G' occurs on the right side of a phonetic vowel symbol ($\langle V\emptyset \rangle$) or the phonetic symbols [l] or [r] and if immediately to the right of 'G' is a vocalic letter ($\langle V1 \rangle$), the letter 'L' or the letter 'R' then substitute 'G' by the phonetic symbol [g]. - Note that this rule expects the left context of the letter 'G' already in a phonetic transcription.

- Signifiers for the sets of letters or sets of phonetic symbols (cf. $\langle V\emptyset \rangle$ and $\langle V1 \rangle$ in the rule illustrated above) are defined by the user at the beginning of a set of rules. The signifier $\langle V1 \rangle$ used in the rule above is defined with the instruction

$\langle V1 \rangle = \text{AEIOUY}$

- G [$\langle K1 \rangle$; R, L], #] -> k

Read: Replace the letter 'G' by the phonetic symbol [k] if 'G' occurs on the left side of a consonantal letter ($\langle K1 \rangle$), except (;) 'R' or 'L', or if 'G' is immediately followed by a word boundary (#).

- [#] AB [-(ER)] -> '+ap

Read: Replace the letter sequence 'AB' by the sequence of phonetic symbols ['+ap] (+ symbolizes the glottal stop), if 'AB' is at the beginning of a word (#) and is not (-) followed by the

letter sequence 'ER'.
RULBOL seems to be an excellent tool for the development and automatic transcription of sets of rules for various transformations of strings.

Unsolved linguistic problems

A process which requires many improvements is the identification of word stress and of sentence stress. Solutions exist only for standard stress. Emphases which deviate from the standard-stress cannot at present be recognized automatically. For their identification at least a rudimentary semantic and pragmatic analysis of the input text would be necessary.

- Problems also occur when (sequences of) letters must be transcribed which occur at morpheme boundaries. For their correct transcription it is often necessary to know where the morpheme boundaries are and therefore to segment the words of the orthographic text into their morphemes (cf. 'Buchstabe' vs 'Wachstum'). None of the existing text-to-speech-systems for German contains a component which performs the necessary segmentations in a sufficient way.

- Since none of the existing systems does an adequate syntactical and semantic analysis of sentences, words which have the same spelling but are differently pronounced cannot be differentiated (c.f. 'Montage' with the pronounciations [mɔn'ta:zə], which means mounting, and ['mo:nta:gə], which means mondays).

Despite all these problems the quality of some existing text-to-speech-systems for German seems to be sufficient for several applications (e.g. reading-aid for blind secretaries/typists who work at an electronic typewriter, accoustic output at computer terminals).

References

- 1: Berry-Rogghe, G. L. M. (1976): Ein Programm zur automatischen Phonemisierung deutscher Texte. Mannheim. (Internal paper of the department of Computational Linguistics at the Institut für deutsche Sprache).
- 2: Bierfert, H. (1985): Automatische Spracherkennung - von der Theorie zur Praxis. Die Konstruktion eines Spracherkennungssystems als thematischer, methodischer und pragmatischer Prozess. Tübingen.
- 3: Breuer, G./Brustkern, J./Thyssen, H./Willée, G. (1979): PHONTEXT - eine PHONOL-Anwendung zur Erzeugung synthetischer deutscher Sprache. In: Sprache und Datenverarbeitung. No. 1/2. P. 10-17.
- 4: (hisholm, D. (1980): A prosodic-phonological dictionary for automatic transcription of large corpora of German verse. In: Sprache und Datenverarbeitung. No. 1. P 10-13.
- 5: Lenders, W./Stock, D. (1984): Zur Entwicklung eines akustischen Terminals zur automatischen Spracherkennung und Sprachproduktion. In: Sprache und Datenverarbeitung. No. 1. P 44-53.
- 6: Müller, B. S. (1981): Regelgesteuerte Umsetzung von deutschen Texten in gesprochene Sprache für das Sprachausgabegerät VOIRAX. Bonn. (Published by the Gesellschaft für Mathematik und Datenverarbeitung/Bonn).
- 7: Müller, B. S. (Ed.) (1985): Sprachsynthese. Zur Synthese von natürlich gesprochener Sprache aus Texten und Konzepten. Hildesheim/Zürich/New York.
- 8: Slaby, W./Spellmann, F.-P. (1980): Automatische Umsetzung von Texten in gesprochene Sprache. In: Wossidlo, P. R. (Ed.) (1980): Textverarbeitung und Informatik. Berlin. P 64-72.

- 9: Spellmann, F.-P. (1981): Automatische Übersetzung von Texten des Deutschen in gesprochene Sprache. In: Schriftenreihe des Rechenzentrums der Universität Münster. No. 55.
- 10: Zinglé, H. (1982): Traitement de la prosodie allemande dans un système de synthèse de la parole. Straßburg. (Doctoral thesis).

Ankündigung für das LDV-FORUM 1/86
ERFAHRUNGSBERICHT ÜBER DAS KONKORDANZ-
PROGRAMM BYU-CONCORDANCE
FÜR ARBEITSPLATZRECHNER

J. Brustkern
Institut für Kommunikationsforschung
und Phonetik
D-5300 Bonn

Mit diesem Bericht soll ein Konkordanz-Programm vorgestellt werden, das Bestandteil eines Programmpaketes für die linguistische Datenverarbeitung auf einem Arbeitsplatzrechner ist. Am IKP wird dieses System vor allem zur Ermittlung von Lesarten in der Lexikonerstellung eingesetzt. Durch die Einbettung des Systems in die vielfältige Software, die für Personal Computer verfügbar ist, ist BYU-Concordance sehr leicht zu handhaben.

**TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA
TA**

High Tech bei TA!

Im Hause TRIUMPH-ADLER forscht und entwickelt der Bereich "Neue Technologien/Basisentwicklung" an neuen, innovativen Konzepten und Systemen im Bereich der Informationstechnik. Basis dieser Arbeit ist ein junges, hochqualifiziertes Team und eine ausgezeichnete Rechnerausstattung (5 Symbolics, 2 VAX, System M32). Eine enge Zusammenarbeit mit Hochschulen und Großforschungseinrichtungen hilft uns, in einem wissenschaftlich orientierten Umfeld neue Ideen rasch aufgreifen und umsetzen zu können. Einer unserer Arbeitsschwerpunkte liegt im Bereich

Büroinformationssysteme

Wir arbeiten insbesondere an der Entwicklung von Systemen im Bereich des Document Filing und Retrieval, unter Einschluß inhaltsorientierter Verfahren. Sollten Sie sich hierfür interessieren und eine fundierte Ausbildung in den Bereichen Informatik und/oder Informationswissenschaft besitzen, sollten wir uns über eine mögliche Zusammenarbeit unterhalten.

Für Personalfragen steht Ihnen Herr Reich (Tel.0911/322-345) zur Verfügung. Bei Fachfragen verbinden wir Sie gerne mit kompetenten Gesprächspartnern des Fachbereichs.

Ihre Bewerbung senden Sie bitte an unser Personalwesen
Gehalt, 8500 Nürnberg 80,
Fürther Str. 212

TA TRIUMPH-ADLER

TOOLS FÜR DIE FORTGESCHRITTENE TEXTVERARBEITUNG AUF MIKROCOMPUTERN

R. v. Ammon

ECO Institut für Electronische Communication
und Organisation
Eichelberg 8
D-8411 Pielmühle

Microcomputer/Personalcomputer befinden sich heute weltweit, auch in der BRD, bereits in sehr großem Ausmaß im Einsatz. Dementsprechend gewaltig ist deshalb auch die Palette der Software-Produkte, die auf solchen Geräten betrieben werden kann. Alleine für die Textbe- und verarbeitung gibt es weit über 100 verschiedene Systeme. Der Umfang ihrer grundsätzlichen Textbearbeitungs-Funktionen ist dabei weitgehend immer gleich; Funktionen wie Zeichen/Wort/Zeile/Absatz/Textblock-Löschen oder Textblock verschieben, einkopieren oder auslagern sind schon seit Jahren selbstverständlich. Davon soll hier nicht die Rede sein. Im folgenden geht es um Tools, die in der Textverarbeitung auf Microcomputern noch nicht zum Alltag gehören oder vereinzelt in Textsysteme zwar teilweise eingebaut sind, aber bisher noch eine verhältnismäßig schlechte Effektivität aufweisen. Das gilt ganz besonders für sprachabhängige Tools, z.B. für Rechtschreibhilfen und die Silbentrennung, da der überwiegende Teil der Textsysteme aus dem Ausland stammt und oftmals noch nicht an die Besonderheiten der deutschen Sprache angepaßt ist.

Rechtschreibfehlererkennung und - korrektur

Systeme zur Rechtschreibfehlererkennung und ggf. sogar zur (halb-)automatischen Korrektur sind für Microcomputer inzwischen verfügbar, und nach unseren Erfahrungen ist das Interesse für solche Zusatz-Tools nicht nur beim Endbenutzer mittlerweile stark gestiegen, sondern auch bei den Herstellern von Textverarbeitungsgen, die Rechtschreibprüfer als Zusatzoption anbieten wollen.

Wie später deutlich wird, können solche Systeme nur für leistungsfähige Microcomputer mit großer Plattenspeicherkapazität entwickelt werden. Da solche Computer - vor allem im deutschsprachigen Bereich - noch relativ neu sind, bestand bislang kein Markt für Mikrocomputer-geeignete Rechtschreibsysteme für die deutsche Sprache. Diese Situation hat sich nun mit der starken Verbreitung leistungsstarker 16-Bit-Rechner geändert.

Was ist computergestützte Rechtschreibfehlererkennung?

Im normalen Schreibbetrieb bestehen die Rechtschreibfehler hauptsächlich aus sog. "Buchstabendrehern" (z.B. Einhi~~e~~t), aus ausgelassenen oder mehrfach getasteten Zeichen und Falschanschlägen. Relativ selten sind dagegen zufällig grammatikalisch gültige Kombinationen von Silbenverbindungen, die nicht beabsichtigt waren und entweder wieder ein lexikalisch gültiges Wort erzeugen oder eine sinnlose Wortneuschöpfung darstellen (z.B. Widerspruch statt Widerspruch).

Ein Rechtschreibfehlersystem soll derartige Fehler erkennen, sie dem Benutzer im Text markieren und ggf. sogar automatisch korrigieren oder dem Benutzer einen Korrekturvorschlag anbieten.

Solche Systeme können in Textverarbeitungen verschieden eingebunden werden.

Fehlererkennung/korrektur während oder nach der Texterfassung?

Auf Microcomputern wird sich wahrscheinlich nur eine der beiden Implementationsmöglichkeiten realisieren lassen: die Fehlererkennung nach der Texterfassung. Der Grund ist einfach (und wird weiter unten noch offensichtlich): Rechtschreibfehlersysteme sind - vor allem, was ihre Wörterbücher angeht - sehr komplex. Eine simultane Fehlererkennung bei der Texterfassung würde die Texteingabe enorm behindern; die Arbeitsgeschwindigkeit und die Arbeitsspeicherkapazität von Microcomputern ist zu gering, so daß Schreibfehler markiert würden, nachdem der Benutzer schon längst über die zu markierende Textstelle hinausgeschrieben hat. Der Benutzer müßte also im Schreibprozeß innehalten, zum Fehlerort zurückkehren und könnte erst anschließend die Texterfassung fortführen. Eine simultane Fehlererkennung hätte nur einen Sinn, wenn der Fehler sofort bei seinem Auftreten markiert und dem Benutzer angezeigt werden kann.

Gerade bei Rechtschreibfehlersystemen spricht jedoch eigentlich nichts gegen ein nachträgliches Verfahren, das den vollständig erfaßten Text überprüft. Dies entspricht durchaus der Benutzererfahrung: der Benutzer überprüft normalerweise seinen Text nach der Erfassung auf Rechtschreibfehler, und genau dasselbe erledigt nun automatisch der Computer.

Die Prozedur der Rechtschreibfehlererkennung kann unabhängig von obiger Implementierungsstrategie ebenfalls unterschiedlich realisiert sein:

Interaktive Fehlerkorrektur oder Batchmodus ohne Benutzerentscheidung?

Diese Fragestellung ist n.u.M. eher akademisch, da zum einen die einschlägige Fachliteratur davon ausgeht, daß Rechtschreibfehlersysteme den menschlichen Korrektor zwar erheblich unterstützen, aber nicht ersetzen können. Ein vollautomatisch arbeitendes System im Batchmodus und ohne Einbeziehung von Benutzerentscheidungen führt nach den Ergebnissen zahlreicher Tests zu erheblichen Fehlentscheidungen des Systems. Da es kein Verfahren mit 100%iger Genauigkeit gibt und - schon aufgrund des Sprachwandels - nie geben wird, könnten bei Verzicht auf die Benutzerentscheidung mehr neue Fehler entstehen, als echte Fehler korrigiert würden. Man kann deshalb davon ausgehen, daß bei allen angebotenen oder in Zukunft angebotenen Verfahren immer um die Benutzer-Entscheidung gebeten wird.

Woraus besteht ein Rechtschreibfehlersystem?

Beim Kauf eines Rechtschreibfehlersystems (oder auch, wenn ein solches als Option mit einem Textsystem angeboten wird) ist es interessant, die wichtigsten Komponenten eines solchen Systems zu kennen; es läßt sich dann besser beurteilen, was man eigentlich einkauft und worauf man bei der Beurteilung eines Systems achten sollte.

Ein Rechtschreibfehlersystem besteht in erster Linie aus Wörterbüchern, je nach Qualität des Systems auch aus mehreren Wörterbüchern: einem sehr umfangreichen, generellen Hauptwörterbuch für beliebige Textbasen (mit oftmals bis zu 40.000 Einträgen), einem anwendungsspezifischen Wörterbuch, das sich der Benutzer für den speziellen Wortschatz seiner Anwendung selbst aufbauen kann (meist nur ein paar hundert Einträge), und zusätzlich kann noch ein allgemeines Schreibfehler-Wörterbuch existieren, das i.d.R. höchstens 200 - 300 besondere Schreibfehler enthält, bei denen eine spezielle Operation eingeleitet werden muß.

Ein typischer Algorithmus eines Rechtschreibfehlersystems vergleicht nun zunächst jedes Textwort mit dem kleinen anwendungsspezifischen Wörterbuch. Falls dort kein entsprechender Eintrag gefunden wird, wird das große generelle Hauptwörterbuch durchsucht. In Hinblick auf den großen Umfang dieser Listen und den daraus resultierenden Schnelligkeitsproblemen können hochfrequente Wörter bei diesem Vergleich im Arbeitsspeicher gehalten werden und dort zuerst überprüft werden.

Um den Umfang der Wörterbücher zu reduzieren, könnte auch eine Art Flexionsendungsreduktion und Derivationsreduktion auf eine Art "Wortstamm" (d.h. eine Rückführung gebeugter und abgeleiteter Wörter auf einen einzigen, hier auf einen künstlichen Wortstamm) enthalten sein. Im Wörterbuch müßte dann für alle Flexionsformen und Derivationen zu einem Wort nur ein Eintrag vorhanden sein. Beispiel: MAGNET für MAGNETEN, MAGNETISMUS, MAGNETISIEREN, MAGNETISIERT, MAGNETISIERE, MAGNETISCH, MAGNETISIEREND usw.

Falls also eine solche Flexionsform nicht gefunden wird, schneidet das System einen Endungsbuchstaben ab und vergleicht von Neuem usw. Ist der Vergleich innerhalb einer bestimmten "Endungs"-länge erfolgreich, wird angenommen, daß das Wort orthographisch richtig geschrieben ist. Andernfalls wird es als nicht identifizierbare Zeichenfolge markiert. Der Nachteil dieser Strategie ist natürlich eine erhöhte Rechenzeit. Die beste Strategie scheint es deshalb zu sein, möglichst viele hochfrequente Flexionsformen in das Wörterbuch aufzunehmen und für niederfrequente Wörter nur einen Stellvertreter in das Wörterbuch einzutragen und durch den Reduktionsalgorithmus behandeln zu lassen. Linguistisch ausgefeiltere Reduktionsalgorithmen dürften aus Rechenzeitgründen in nächster Zukunft wohl kaum realisierbar sein; man kann eher erwarten, daß in Anbetracht ständig neuer Speichermedien mit immer höherer Kapazität das Problem umfangreicher Wörterbücher zunehmend schwinden wird.

Damit in einer Anwendung nicht zuviele Textwörter als nicht identifizierbar bzw. orthographisch falsch markiert werden, muß ferner die Möglichkeit bestehen, daß der Benutzer selbst anwendungsspezifisch bestimmte Wörter von der Fehlerbehandlung ausschließen kann (z.B. Namen, Formeln, Acronyme, typische Abkürzungen z.B. für Gewichte, Größeneinheiten etc.) oder diese vom System als solche selbst erkannt werden.

Um den Benutzer nicht zu oft unnötig um seine Entscheidung zu bitten und damit die Akzeptanz des Systems zu gefährden, könnte das System eine weitere Strategie verfolgen: Nicht identifizierte Textwörter könnten einer Häufigkeitanalyse unterzogen werden, wobei man davon ausgeht, daß Textwörter, die mit einer entsprechend höheren Frequenz auftreten, keine Schreibfehler sein

können, auch wenn sie nicht identifiziert werden konnten; umgekehrt könnte man (zusätzlich) annehmen, daß Wörter mit entsprechend niedriger Frequenz mit großer Wahrscheinlichkeit Schreibfehler sind (z.B. wenn sie nur einmal auftreten).

Wie gesagt, sollte der Benutzer bei der Fehlerkorrektur möglichst wenig belastet werden. Ein gutes Rechtschreibfehlersystem sollte sich deshalb eine Benutzerentscheidung merken und damit selbstlernend oder selbstoptimierend sein. Beispielsweise sollte ein dem Benutzer als orthographisch falsch markiertes Textwort automatisch in das Hauptwörterbuch übernommen werden, falls es der Benutzer als richtig akzeptiert hat; umgekehrt sollte ein vom System angebotener Korrekturvorschlag, der vom Benutzer manuell nachkorrigiert werden mußte, in das Wörterbuch für allgemeine Schreibfehler eingetragen werden. Außerdem sollte das System während derselben Korrektursitzung eine Benutzerentscheidung automatisch auf alle identischen Fälle in der Textbasis anwenden, ohne noch den Benutzer zu fragen.

Falls man eine Rechtschreibfehlererkennung erwerben will, sollte nicht zuletzt unbedingt darauf geachtet werden, daß die Wörterbücher durch den Benutzer selbst gewartet werden können und der Benutzer hierbei nicht in eine umständliche und ggf. auch finanzielle Abhängigkeit eines Herstellers gerät, der für Wartung oder jedes Updating Gebühren berechnet. (Verfügbare Rechtschreibhilfesysteme vgl. in Anm. 1)

Grammatik-Prüfer

Man beklagt heute allgemein die mangelhaften Kenntnisse der deutschen Grammatik, die sich durch alle sozialen Schichten und Berufsqualifikationen feststellen lassen, nicht zuletzt auch in Seminar- und Klausurarbeiten an den Universitäten.

Einige Beispiele sollen die Vielfalt grammatikalischer Fehler zeigen; sie entstammen immerhin einem wissenschaftlichen Bericht, dessen dritte Korrekturfassung hier ausgewertet wurde:

Beispiele

"Dadurch ergibt sich natürlich eine deutliche Negativverschiebung **bei die** Bewertung ..."

"Die Titelverteilung, basierend **auf den** übereinstimmenden Urteil aller Tester, **zeigt** Abbildung 1 und Abbildung 2.

"Daher sollen die Skalenwerte nicht als absolute Größen interpretiert **werden sondern** nur dem Zahlenvergleich dienen."

"Falls **man**, die drei Extrembewerter aus der Wertung herausnimmt, erhält man das ..."

"Dann erhält man die **folgende** Ergebnisse:"

"Inklusive **die** Kontrolltitel hätte die Frage ..."

"..., wenn man den Testern entsprechend **ihrer** eigenen Angaben hinreichende Kenntnisse unterstellt."

"Aus den Statistiken dieses **Test** lassen ..."

Die Beispiele enthalten Rektionsfehler, Interpunktionsfehler, Numerus- und Kasusfehler. Damit sind natürlich noch längst nicht alle Fehlerquellen genannt; häufig finden sich auch Tempusfehler, falsche Wortstellungen und Satzbaufehler, falscher Gebrauch von Präpositionen, Genusfehler, falsche Relativsatzanschlüsse oder Artikelfehler u.v.m.

Die meisten o.g. Beispiele entstanden sicherlich als Flüchtigkeitsfehler. Gerade bei computergestützter Textverarbeitung lassen sich typischerweise häufig auch Fehler beobachten, die durch intensives Korrigieren (Löschungen, Einfügungen, Verschieben von Textblöcken etc.) neu entstehen. In dem Beispiel "Dann enthält man die folgende Ergebnisse" war der Artikel "die" vielleicht ursprünglich nicht enthalten; beim nachträglichen Einfügen könnte der Verfasser dann vergessen haben, die Flexionsendung von "folgende" anzupassen.

Besonders wer in der Ausbildung tätig ist oder häufig größere Dokumente erstellen muß, kann sicherlich bestätigen, daß hier keine ausgefallenen Beispiele ausgesucht wurden. Grammatikfehler werden nicht nur bei anderen beobachtet, sondern nahezu jeder übersieht mehr oder weniger häufig fehlerhafte Grammatikkonstruktionen oder ärgert sich nachträglich über peinliche Fehler in bereits versandten Papieren, die nicht korrigiert wurden.

Es ist deshalb naheliegend, daß im Zusammenhang mit der automatischen Textverarbeitung auch hierbei der Computer helfen soll.

Diese Aufgabe könnte von Grammatik-Prüfroutinen, sog. Syntax-Parsern, übernommen werden. Schon seit 20 bis 30 Jahren forscht man bekanntlich - auch in der BRD - vor allem im Rahmen der automatischen Sprachübersetzung an der Entwicklung leistungsstarker Syntax-Parser.

Es ist unmöglich, hier die Menge der verschiedenen Grammatikmodelle und Algorithmen vorzustellen, die bei der Entwicklung von Syntax-Parsern zur Anwendung kamen und immer wieder als neue Ansätze erfunden werden. Auf eine nähere Skizzierung kann in diesem Rahmen zudem verzichtet werden, da zumindest die Grundprinzipien diesem Leserkreis bestens bekannt sein dürften.

Schon ein relativ einfacher Parser könnte nun Sätze, die er nicht vollständig analysieren kann, markieren (z.B. auf Bildschirm invers darstellen). Der Benutzer muß dann selbst versuchen, eine richtige grammatikalische Form zu finden. Da die meisten Grammatikfehler Flüchtigkeitsfehler sind, dürfte der Benutzer bei dieser Aufgabe nur selten überfordert sein. Ein solcher einfacher Parser wäre also bereits eine große Hilfe.

Ähnlich wie bei der Rechtschreibung könnte der Parser aber auch gleich richtige Grammatiklösungen anbieten, die der Benutzer nur noch z.B. durch Tastendruck bestätigen bzw. verwerfen und selbst korrigieren muß.

Leider sind Syntax-Parser für die deutsche Grammatik noch weit davon entfernt, in der Textverarbeitung auf Microcomputern einsetzbar zu sein. Sie benötigen zum einen zuviel Rechenkapazität für die Analyseschritte und zuviel Speicherkapazität für die Wissensbasis. Zum anderen ist ihre Leistungsstärke noch nicht für eine praktische Verwendung geeignet; die heutigen Parser würden

noch zuviele Fehlentscheidungen liefern. Zumindest wird es jedoch in absehbarer Zeit (vielleicht schon in ein oder zwei Jahren) Grammatikhilfen für Microcomputer geben, die von den Herstellern **integriert mit den Rechtschreibhilfen** angeboten werden. Diese Integration ist sicherlich sinnvoll, weil der Benutzer bei der Verbesserung den Text nur einmal durchlaufen und nicht für jede Fehlerart ein eigenes Programm starten will (vgl. Anm.2).

Automatische Silbentrennung

Ähnlich wie bei der Rechtschreibfehlererkennung sind Verfahren zur automatischen Silbentrennung für Personalcomputer noch längst nicht selbstverständlich. Aber auch hier ist mit der Verbreitung leistungsstärkerer 16-Bit-Rechner bei den Produzenten von Textverarbeitungen zunehmendes Interesse festzustellen, und bei einigen Textverarbeitungen wird die Silbentrennung inzwischen bereits halb- oder vollautomatisch unterstützt (z.B. halbautomatisch bei WordStar von MicroPro, nach den allgemeinen Benutzererfahrungen mit einer Genauigkeit von ca. 60 - 70%, oder demnächst vollautomatisch bei WORD von Microsoft oder Wordstar 2000 von MicroPro).

Warum braucht man eigentlich eine automatische Silbentrennung?

In Anbetracht des Aufwandes, den ein Silbentrennprogramm an Speicherplatz und Rechenzeitverhalten stellen, fragt man sich vielleicht, ob ein solches Programm in einer Textverarbeitung überhaupt notwendig ist. Für die Integration einer (halb-)automatischen Silbentrennung gibt es allerdings einige gute Gründe:

Der wohl wichtigste Grund ist die Leistungssteigerung des Texterfassers. Es wurden Leistungssteigerungen von 20% - 40% beobachtet, die darauf zurückzuführen sind, daß die Schreibkraft endlos dahinschreiben kann, ohne sich um den Aufbau einer Zeile und die Worttrennung kümmern zu müssen.

Weitere Gründe resultieren aus Layout-bedingten Anforderungen. Beispielsweise ist bei sog. Blocksatz (d.h. rechtem Randausgleich) eine (halb-)automatische Silbentrennung nahezu eine unbedingte Voraussetzung; andernfalls entstehen im Satzbild zu große Wortzwischenräume, die vom Benutzer nicht mehr akzeptiert werden können.

Beispiel:

Dies ist ein
Beispiel für einen
Blocksatz ohne
automatische
Silbentrennung.

Es ist leicht einzusehen, daß die Notwendigkeit für ein Silbentrennprogramm zunimmt, je kleiner die Spaltenbreite eines Textes ist. Vor allem Textverarbeitungen, die auch Spaltensatz anbieten (z.B. WordStar 2000, TextStar von Omnitex in Rheinfelden u.v.a.), sollten deshalb unbedingt eine - zumindest halbautomatische - Silbentrennung beinhalten.

Wie leistungsfähig sind Silbentrennprogramme?

Über die manchmal kuriosen Ergebnisse von Silbentrennprogrammen wurde schon viel gelacht. Trotzdem läßt sich heute eine Trenngenauigkeit von über 99% erreichen. Wenn von Herstellern Genauigkeitsangaben z.B. von 99% genannt werden, bedeutet das, daß eine von hundert Trennungen falsch gemacht wird. **Man sollte unbedingt nachfragen, auf welcher Textbasis diese Ergebnisse erzielt wurden (z.B. Fachtexte, Zeitungstexte u.ä. oder Durchschnittsergebnis für verschiedene Textsorten) und wie groß die Textbasis war!** Ohne solche Informationen sind Leistungswerte völlig nichtssagend.

Eine Trenngenauigkeit von über 99% gilt gemäß der einschlägigen Literatur heute übrigens als Standardwert. Diese Anforderung sollte man an eine Silbentrennung stellen, wenn man eine Textverarbeitung mit Silbentrennung erwirbt. Trotzdem muß man darauf gefaßt sein, daß die derzeit auf Microcomputern installierten (halb-)automatischen Silbentrennungen diesen Standardwert nicht erreichen. Diese Verfahren basieren meist noch auf ganz simplen Algorithmen.

Die folgenden Informationen liefern einige Hinweise dafür, was ein gutes und was ein primitives Silbentrennungsverfahren ist bzw. wie komplex eine vernünftige Silbentrennung sein muß.

Wie funktioniert eine automatische Silbentrennung?

Programme für automatische Silbentrennung werden inzwischen auch für Personalcomputer von verschiedenen Herstellern angeboten (vgl. Anm. 3). Die ersten Versuche auf Großrechnern begannen allerdings schon Anfang der 60-iger Jahre, und im Laufe der Entwicklungsgeschichte verfolgte man im wesentlichen drei Lösungsansätze:

Zuerst versuchte man - als naheliegende Lösung - , den Wortschatz einer Sprache mit den jeweiligen Trennstellen eines Wortes abzuspeichern. Textwörter sollten dann einfach durch Vergleich mit dieser Liste getrennt werden. Dieses Verfahren erwies sich jedoch als völlig unwirtschaftlich und viel zu langsam. Ein Lexikonorientiertes Verfahren mit vielen tausend Einträgen gilt deshalb als veraltete Methode (, sofern nicht eine Notwendigkeit durch die Integration mit einer Rechtschreibhilfe gegeben ist).

Bei der zweiten Methode speicherte man für Buchstabenkombinationen, die an der "Nahtstelle" von Silben vorkommen, Wahrscheinlichkeitsregeln ab, über die entschieden wurde, wie eine entsprechende Kombination zu trennen ist. Der Lösungsansatz benötigte zwar relativ wenig Rechenaufwand, lieferte jedoch keine akzeptablen Genauigkeitswerte.

Die dritte Methode, die sich sowohl von der Leitungsqualität als auch im Hinblick auf die Rechneranforderungen inzwischen durchgesetzt hat, basiert auf einem logischen Regelwerk. Ein typischer Algorithmus könnte - vereinfacht dargestellt - z.B. folgendermaßen aussehen:

Wird bei der Texteingabe (oder bei nachträglicher Textformatierung) die rechte Spaltengrenze überschritten, wird das entsprechende Wort ab der Spaltengrenze Zeichen für Zeichen von rechts

nach links untersucht, bis sich der erste Vokal findet. Nach dem Vokal bzw. vor dem ersten Konsonanten liegt dann meistens die Trennstelle.

Beispiel: RE-DEN

Dieser Algorithmus ist jedoch noch wesentlich zu schwach und würde selbst für eine halbautomatische Trennung zu viele falsche Trennvorschläge liefern. Typische Fehlerquellen sind Wörter,

- die an der Trennstelle zwei oder mehrere Konsonanten vor dem identifizierten Vokal aufweisen,
- die aus mehreren Wörtern zusammengesetzt sind und bei denen an der Fuge zwischen den Kompositumbestandteilen das zweite Wortelement mit einem Vokal beginnt (z.B. HAUS-ECKE).

Einfache Verfahren, wie z.B. das Trennverfahren von WordStar, lösen lediglich noch das erste Problem dadurch, daß im Falle von mehreren Konsonanten nach dem ersten Konsonanten getrennt wird (kurioserweise aber nicht immer); das zweite Problem wird gar nicht behandelt. Damit erreicht WordStar nach den allgemeinen Erfahrungen ca. 60% - 70% korrekte Trennungen. Für halbautomatische Verfahren scheint dieser Wert bislang weitgehend akzeptiert worden zu sein, vor allem wohl, weil man an bessere Silbentrennungen noch nicht gewöhnt ist.

Um den für automatische Silbentrennungen erforderlichen Akzeptanzwert von über 99% Trenngenauigkeit zu erreichen, muß ein Algorithmus jedoch noch erheblich ausgebaut werden:

Es gibt zahlreiche, kritische Buchstabenkombinationen, die gesondert überprüft werden müssen und die Trennung nur nach statistischer Wahrscheinlichkeit entschieden werden kann. Ein Beispiel ist die Buchstabenfolge "gl": Vor "gl" wird getrennt bei ABERGLAUBE, nicht aber bei HANG-LAGE; steht vor "gl" ein Vokal, wird vor "gl" richtig getrennt bei KÄSE-GLOCKE, aber nicht bei TAUGLICH.

Solche - im Deutschen sehr häufigen - Problemfälle werden bei anspruchsvollen Silbentrennprogrammen durch Listen zu lösen versucht, die nicht trennbare Konsonantengruppen sowie im Hinblick auf zusammengesetzte Wörter Vor-, Nach- und Zwischensilben enthalten. Durch Erkennen z.B. der Vorsilbe "vor" kann dann VORARBEITER richtig getrennt werden; oder es wird eine als untrennbar gespeicherte Konsonantengruppe wie "br" wieder trennbar, wenn die Vorsilbe "ab" in der Vorsilbenliste enthalten ist, z.B. bei AB-RÄUMEN.

Für Ausnahmetrennungen müssen ferner Ausnahmelisten angelegt werden, z.B. für Wortzusammensetzungen mit Vokalanfang beim zweiten Kompositumelement (z.B. VOKAL-ANFANG). **Es ist - wie schon bei der Rechtschreibung - darauf zu achten, daß diese Ausnahmelisten selbst gewartet werden können und man nicht in die Abhängigkeit des Herstellers gerät!** Es sollten außerdem nicht ganze Wörter und seine sämtlichen Ableitungen in solche Listen eingetragen werden müssen, sondern nur eine (z.B. auf fünf Buchstaben) beschränkte Buchstabenanzahl, die für eine eindeutige Identifikation genügt.

Beispiel: ARBEI statt ARBEITER, ARBEITERIN, ARBEITEN, ARBEITERSCHAFT etc.

Formal und auch über Ausnahmelisten nicht lösbar, sind sog. Homographen (Wörter mit gleicher Schreibweise, aber unterschiedlicher Bedeutung), z.B. DAS SPIEL-ENDE, aber SPIE-LENDE KINDER. In solchen Fällen muß das Programm nach statistischen Wahrscheinlichkeitsregeln entscheiden.

Sehr komfortable Silbentrennprogramme können schließlich unsichere Trennungen umgehen und auf sichere Trennstellen ausweichen. Ferner können sie - besonders für Spaltensatz - die Möglichkeit der (grammatikalisch falschen) Zwangstrennung enthalten, wenn keine geeignete Trennstelle gefunden werden kann, trotzdem jedoch getrennt werden muß. Zuletzt könnte auch noch das Problem der "ästhetischen" Trennung angeführt werden (URIN-STINKT statt UR-INSTINKT); da solche Trennungen aber grammatikalisch korrekt sind, wird sie der Benutzer gerne akzeptieren, und ein zusätzlicher Aufwand wäre nicht gerechtfertigt.

Testbeispiele für die Qualität von Silbentrennprogrammen

Wenn man wissen will, ob eine Silbentrennung nur einen primitiven Algorithmus enthält oder tatsächlich leistungsfähig ist, kann man das Programm mit folgenden Beispielen schnell testen. (Die Beispiele sind größtenteils aus der einschlägigen Fachliteratur entnommen):

BEAT-MUSIK BE-ATMUNG	ANGEB-LICH VER-BLICHEN	ARBEIT-SAM ARBEITS-AMT
GUMMI-SCHLAUCH MISCH-LAUCHGEMÜSE	HÖF-LICH VER-PFLICHTET	HÖS-CHEN RÜ-SCHEN
BANK-NOTEN AN-KNOTEN	VERGNÜG-LICHEN VER-GLICHEN	GÄ-STE AUS-TRAGEN AN-STELLEN
NACH-TEIL NACHT-EILZUG	BESCHWÖRUNGS-EID SCHANTUNG-SEIDE	BUNDES-TAG
EI-NIGE EIN-IGELN	HILFS-ARBEITER SCHILF-SANDSTEIN	DEUTSCH-LAND

Wenn das Programm diese Beispiele richtig trennt, kann man weitgehend sicher sein, daß es sich um eine qualitativ hochwertige Silbentrennung handelt.

Silbentrennung während oder nach der Texterfassung?

Im Unterschied zur Rechtschreibfehlererkennung sollte n.u.M. die Silbentrennung und damit der Zeilenumbruch bereits bei der Texterfassung erfolgen. Voraussetzung dafür ist allerdings, daß das Silbentrennprogramm schnell genug ist und eine hohe Trennqualität hat. Andernfalls wird die Schreibkraft durch Trennkorrekturen aufgehalten und muß den Erfassungsfluß unterbrechen (oder nach der Texteingabe korrigieren) oder - im Falle unzureichender Schnelligkeit - ist das Satzbild wegen Umbruchsarbeiten des Programmes ggf. für die Schreibkraft unzumutbar stark in Bewegung.

Als Alternative zu echter simultanen Trennung am Zeilenende könnten deshalb auch größere Texteinheiten, z.B. ganze Absätze,

getrennt werden, während die Texterfassung simultan fortgesetzt werden kann.

Eine simultane Silbentrennung hat den Vorteil, daß die Umbruchs-gestaltung bereits bei der Texterfassung beobachtet und gesteuert werden kann.

Aufgrund der o.g. Probleme (Schnelligkeitsprobleme, zu geringe Trennqualität, unruhiges Satzbild) muß bei allen von uns getesteten Textverarbeitungen derzeit die (halb-)automatische Silbentrennung im Anschluß an die Texterfassung extra aufgerufen werden (bei WORD von Microsoft wird nach unseren Informationen in der kommenden Version eine simultane Silbentrennung eingebaut sein, ebenso bei Wordstar 2000 von MicroPro).

Interaktive Silbentrennung oder Batchmodus ohne Benutzerentscheidung?

Bei einer simultanen Silbentrennung kann ein interaktiver Modus nicht sinnvoll sein, da andernfalls die Schreibkraft die Texterfassung unterbrechen und die Trennstelle bestätigen oder korrigieren müßte. Damit würde die Steigerung der Schreibleistung, die durch automatische Silbentrennung und automatischen Umbruch ermöglicht werden soll, wieder zurückgenommen.

Dagegen ist die Einbeziehung der Benutzerentscheidung bei Silbentrennung nach der Texteingabe weitgehend problemlos und, falls keine ausreichende Trenngenauigkeit erreicht wird, unbedingte Voraussetzung (zur Verfügbarkeit von Silbentrennungen vgl. Anm. 4).

Indexe, Stichwortverzeichnisse, Sach- und Personenregister

Über fehlende oder schlechte Stichwortverzeichnisse/Indexe/Regi-ster ärgert man sich häufig. Viele haben sicherlich auch schon einmal versucht, selbst ein Stichwortverzeichnis o.ä. aufzu-bauen, und mußten dabei erleben, wie umständlich und langwierig solche Arbeiten sind. Die Mühseligkeit der Arbeit und die gleichzeitig hohen Anforderungen an die Sorgfalt sind der Grund, daß auf Nachweissysteme wie Stichwortverzeichnisse oder noch aufwendigere Register oft verzichtet wird oder diese nur mangelhaft und schlampig als lästige Pflicht an Dokumente angehängt werden. Mangelhafte oder fehlende Nachweissysteme führen jedoch dazu, daß Dokumente wie z.B. Handbücher, Bedienungsanleitungen, Dokumentationen etc. für Benutzer schlecht zugänglich und unhandlich sind. Vor allem folgende Mängel sind bekannt:

Übliche Mängel von Stichwortverzeichnissen

- Das Verzeichnis enthält zuwenig Suchbegriffe.
- Zu den Stichwörtern werden nur eine oder zuwenig Adressen oder auch falsche Adressen nachgewiesen, an denen sie im Dokument tatsächlich vorkommen.
- Zu den Stichwörtern wird (im Sinne eines Registers) zuwenig Information mitgeliefert, aus der die tatsächliche Bedeutung

eines Begriffs erkennbar ist, die dieser unter der jeweiligen Adresse im Dokument hat.

Traditionelle Erstellung von Stichwortverzeichnissen

Gute Stichwortverzeichnisse oder Register aufzubauen, ist heute noch sehr aufwendig und langwierig. Traditionell wird - z.B. in vielen Verlagen - immer noch mithilfe von Zettelkästen verfahren; ein Dokument wird nach Index-relevanten Begriffen durchgesehen, ein entsprechender Begriff auf einer Karteikarte zusammen mit seiner Adresse im Dokument eingetragen und anschließend wird das Dokument weiter durchsucht. Wird wieder ein Indexbegriff gefunden, muß erst im Zettelkasten nachgesehen werden, ob bereits eine Karteikarte angelegt wurde. Im positiven Fall wird die Adresse des aktuellen Begriffs in der Karteikarte eingetragen, im negativen Fall muß eine neue Karteikarte angelegt werden.

Ist der Index endlich im Karteikasten fertiggestellt, muß er erst noch texterfaßt werden und noch einmal alle Begriffe sowie die Adressen überprüft werden, ob sie einer "Gesamtschau" noch standhalten oder ein weiteres Mal modifiziert werden müssen.

Es läßt sich leicht erkennen, daß vor allem bei umfangreicheren Dokumenten schnell ein großer manueller/intellektueller Aufwand entsteht. Zu einem Dokument mit hundert oder mehreren hundert Seiten können in einem guten Index ohne weiteres mehr als 1000 Indexbegriffe mit oftmals mehreren Adressen enthalten sein.

Da inzwischen Dokumente zumindest im professionellen oder halbprofessionellen Bereich großteils auf (Micro-)Computern geschrieben werden, war es naheliegend, diese Aufgabe durch den Computer zu unterstützen.

Strategien computergestützter Verfahren

Niemand wird natürlich erwarten, daß eine vollautomatische Indexerstellung ohne jegliches menschliche Zutun möglich ist. Die endgültige Entscheidung, ob ein Begriff bzw. welche Adresse hierzu in einen Index aufgenommen werden soll, kann nur durch den Menschen erfolgen.

Die gegenwärtigen computergestützten Index-Erstellungsverfahren lassen sich nach zwei verschiedenen Strategien unterscheiden:

Strategie 1:

Der Benutzer muß bei der Texterfassung, in bestimmte "Trenner" eingeschlossen, **zusätzlich** zum eigentlichen Text den Indexbegriff miteingeben. Beispiel:

Das ist ein Beispiel, in dem Indexbegriffe /*Indexbegriff*/ enthalten sind, die zusammen mit ihrer Adresse /*Adresse*/ automatisch aus dem Text selektiert werden sollen.

Solche Begriffe werden beim Ausdrucken eines Dokuments unterdrückt und stören damit nicht das gedruckte Dokument. Das Verfahren selektiert schließlich anhand der Trenner die Indexbegriffe zusammen mit der Adresse aus dem Dokument und erstellt eine

sortierte Indexliste.

Das Verfahren beinhaltet zwei größere Nachteile:

- Es ist für den Benutzer immer noch sehr aufwendig, weil die Begriffe **zusätzlich** zum eigentlichen Text (und zu jeder relevanten Adresse erneut wieder) eingetippt werden müssen.

Angenommen, ein Index hätte tausend Indexbegriffe, jeder Indexbegriff durchschnittlich zwei Nachweisadressen, und man könnte davon ausgehen, daß die durchschnittliche Länge eines Indexbegriffs wenigstens zehn Zeichen beträgt, müßten zusätzlich 20.000 Zeichen sowie 8.000 Trennerzeichen miterfaßt werden.

- Der zweite Nachteil liegt darin, daß die Auswahl von Indexbegriffen und die Auswahl der relevanten Adressen bei der Texterfassung nur sehr unzuverlässig funktionieren kann: zum einen weiß man bei der Texterfassung meist noch nicht, was relativ zum aktuellen Standort im Dokument noch nachfolgt bzw. ob z.B. eine Adresse an späterer Stelle relevanter und die aktuelle Adresse überflüssig gewesen wäre. Derartige Mängel in Stichwortverzeichnissen kann man oft feststellen, wenn man den Index eines Buches wirklich überprüft. Zweitens werden viele relevanten Indexbegriffe und Adressen übersehen, weil man bei der Texterfassung bereits durch die Aufgabe der Texterstellung stark abgelenkt ist. Napoleonische Fähigkeiten finden sich bei uns aber sehr selten.

Ein bekanntes System, das auf dieser Strategie basiert, ist z.B. Starindex (von MicroPro) bzw. die in Wordstar 2000 integrierte Indexerstellung.

Strategie 2:

Eine neue Strategie wurde durch das Programmsystem ECOINDEX eingeführt (vgl. Anm. 5). Dieses Verfahren basiert auf Methoden, die in der Informationslinguistik als sog. KWIC, KWAC oder KWOC (KeyWord In Context, KeyWord and Context, KeyWord Out of Context) bekannt sind und bisher auf Großrechenanlagen implementiert waren.

Bei dieser Strategie wird davon ausgegangen, die Texterfassung und die Indexerstellung wieder in getrennte Arbeitsgänge aufzuteilen. Gleichzeitig soll im Unterschied zur ersten Strategie nicht nur die Qualität der Indexerstellung erheblich verbessert, sondern darüber hinaus auch die menschliche Arbeit wesentlich reduziert werden.

Ein Index-Erstellungsverfahren auf der Basis dieser zweiten Strategie "liest" nach der Dokumenterfassung das gesamte Dokument selbständig durch. Dabei wird quasi ein "Vorschlag" für einen Index erstellt, der bereits von den meisten "Trivialwörtern" (Begriffe ohne Bedeutung für ein Dokument) gereinigt ist. Dem Anwender wird dieser Vorschlag in einem interaktiven Editordialog angeboten; die angebotenen, potentiellen Indexbegriffe können entweder in vorliegender Form durch Tastendruck bestätigt werden oder über zahlreiche, speziell auf die Indexerstellung ausgerichtete Editorfunktionen (über 30 Funktionen) modifiziert werden oder natürlich auch aus dem Index ausgeschlossen werden. Die

angebotenen Vorschläge sind alphabetisch sortiert; die erste Fundstelle eines Begriffs wird dabei zuerst ausgegeben, weil davon ausgegangen wird, daß viele Index-relevanten Begriffe bei ihrem ersten Auftreten im Dokument zusammen mit ihrer Definition bzw. weiteren Ausführungen eingeführt werden und dies deshalb meist die wichtigste Fundstelle im Dokument ist. Die Index-Vorschläge können wahlweise mit einer beliebigen Kontextgröße (Anzahl der Wörter in der Textumgebung des Begriffs) ausgegeben werden.

Der Kontext spielt bei diesem Verfahren eine wichtige Rolle. Es gibt - wie im obigen Beispiel ersichtlich - sog. Homographen (Wörter mit gleicher Schreibweise, aber unterschiedlicher Bedeutung), die teilweise in den Index aufgenommen werden sollten, teilweise jedoch nicht. "Ehe" als Substantiv könnte z.B. Index-relevant sein, "Ehe" als Konjunktion am Satzanfang sollte dagegen sicher ausgeschlossen werden. Diese Vereindeutigung kann nur über die Kontextinformation entschieden werden. Hinsichtlich der Index-Entscheidung wird außerdem eine Zoom-Funktion angeboten, mit der der Benutzer direkt ins Dokument eintreten und prüfen kann, ob ein Begriff an dieser aktuellen Dokumentstelle tatsächlich Index-relevant ist.

Die Adresse des Wortes kann wahlweise aus einem Dokumenten(-einheiten-)namen (z.B. Band 1, Band 2, oder Teil I u.ä.), der Seitenzahl, der Zeilenzahl oder auch noch der Wortnummer in der Zeile ausgegeben werden, da eine Seitenangabe alleine oft eine relativ ungenaue Adresse darstellen kann.

Die Adresse könnte auch spaltenweise untereinander ausgegeben werden.

Es ist schließlich auch möglich, den Kontext zu einem Indexbegriff mitausgeben zu lassen. Der im Index-Vorschlag angebotene Kontext kann zu diesem Zweck ebenfalls editiert werden. Die Kontextumgebung kann zu jedem Indexbegriff individuell eingestellt werden, Unwichtiges kann aus dem Kontext gelöscht und durch drei Pünktchen ersetzt werden.

Mithilfe des Kontexts kann der Leser eines Dokuments wesentlich gezielter suchen, da der Indexbegriff über den Kontext vereindeutigt wird. Durch diese zum Indexbegriff mitgelieferten Zusatzinformationen sollen Register-ähnliche Verzeichnisse ermöglicht werden, die über Textbegriffe wesentlich mehr aussagen als normale Indexverzeichnisse.

Im selben Arbeitsgang ist es bei ECOINDEX auch möglich, parallel ein Personenregister aufzubauen.

Nach dieser intellektuellen Nachbearbeitung des Index-Vorschlags durch den Benutzer kann dieser zuletzt noch das gewünschte Layout des Indexverzeichnisses festlegen, den Rest erledigt das System. Der fertige Index wird in einer Datei abgelegt und könnte ggf. noch einmal in einen Texteditor geladen und weiter editiert werden.

Textverarbeitungswerkzeuge für Dokumentare, Sprachforscher, Literaturwissenschaftler und ähnlich Gesinnte

Wörterbücher, Frequenzwörterbücher, rückläufige Wörterbücher, Zeilen- und Satzkonkordanzen und Thesauruserstellung

Als Sprachforscher, Textlinguist o.ä. erinnert man sich bestimmt noch an die Zeiten, als man wenig genüßlich die Josephine Mutzenbacher auf Lochkarten erfaßte, um anschließend über ein Konkordanz-Programmsystem herauszufinden, welche Ausdrucksvarianten zu "Pudern" seinerzeit im Wiener Straßenleben existierten. Die Umständlichkeit und Langwierigkeit der Texterfassung und die Mühen, bis man endlich die gewünschten Output-Listen erhielt, verdarben damals oftmals die Freude an der Wissenschaft.

Ganz anders ist die Situation heute. Immer größere Textmengen werden inzwischen mithilfe von Textverarbeitungssystemen produziert und bleiben langfristig für eine EDV-Anwendung auswertbar. Alle Textsorten sind vertreten, Briefe, Protokolle, Berichte, ganze Bücher. Nahezu sämtliche sozialen Schichten und Berufsqualifikationen schreiben inzwischen auf Microcomputern, nicht nur Autoren, sondern auch Schüler, Studenten, Sekretärinnen, Ämter, gelegentlich auch der Chef persönlich. Eine Fülle von Textmaterial entsteht hier ganz selbstverständlich, das sich nach den unterschiedlichsten (textlinguistischen, soziolinguistischen, corpuslinguistischen usw.) Gesichtspunkten auswerten ließe. Wenn nur die entsprechende Software dafür existierte!

Gibt es spezielle Textverarbeitungswerkzeuge für den Sprachforscher, Literaturwissenschaftler, Dokumentar und ähnlich Gesinnte?

Werkzeuge aus der 'Vergangenheit'

Programmsysteme zur Erstellung von Wortformenwörterbüchern, Frequenzwörterbüchern und Konkordanzen existieren bereits seit langem. Sie wurden hauptsächlich an Universitäten entwickelt und wurden auf deren Großrechnern implementiert.

Endbenutzer dieser Systeme sind bis heute in der Regel Sprach- und Literaturwissenschaftler, aber auch Mitarbeiter von Bibliotheken, Verlagen, Dokumentationsstellen u.ä., d.h. also meistens EDV-Laien. Besonders für solche Benutzergruppen weisen diese Systeme schon aus der Sicht der Anwendbarkeit und Bedienung erhebliche Mängel auf:

- Die Programmsysteme sind ausschließlich auf Großrechner implementiert, zu denen ein Endbenutzer kaum oder nur erschwert Zugang hat und mit dessen Bedienung er in der Regel nicht vertraut ist.
- Die Programmsysteme selbst sind vom Endbenutzer nicht bedienbar oder setzen eine intensive Schulung voraus. Fast immer müssen sie von Experten bedient werden, von denen der Benutzer abhängig ist.
- Die Programmsysteme sind nicht auf Textsystemen oder Microcomputern einsetzbar, die heute für die Texterstellung komfortable und bedienungseinfache Schreib- und Editierfunktionen (Textbearbeitungssoftware) anbieten und an vielen Arbeitsplätzen be-

reits verfügbar sind. Die Texte müssen deshalb erst auf Großrechenanlagen wiederum von Experten überspielt werden. Dies wird zusätzlich erschwert, weil die Programmsysteme bestimmte Input-Formate voraussetzen und deshalb Anpassungsprogramme notwendig werden. Ggf. müssen die Texte sogar neu erfaßt werden.

- Die Programmsysteme benötigen derart umfangreiche Betriebsmittel, daß sie im Betrieb eines Rechenzentrums häufig nur zu Sonderzeiten geladen werden können.

Die Programmsysteme ließen sich folglich nicht dort einsetzen, wo sie tatsächlich gebraucht werden, nämlich am Arbeitsplatz des Endbenutzers.

Was ist heute möglich?

Inzwischen wurde z.B. mit CORDA ein auch für den Microcomputer geeignetes Programmsystem entwickelt (vgl. Anm. 6). Wichtige Kriterien bei der Entwicklung waren es, daß ein zumindest gleich mächtiges Funktionsangebot zur Verfügung gestellt werden konnte wie bei entsprechenden Programmsystemen auf Großrechenanlagen sowie daß das System leicht und sicher vom Endbenutzer selbst bedient werden kann, ohne dabei von einem umständlichen Beratungsdienst abhängig zu sein. Die Bedienung von CORDA geschieht vollständig Dialog-orientiert; das System kann auch von einem EDV-Laien sicher bedient werden, nicht zuletzt deshalb, weil beim Dialog zwischen Benutzer und System ständig Plausibilitätskontrollen über die eingegebene Parametrierung stattfinden. Diese sollen verhindern, daß falsch parametrierte Systemläufe zustandekommen; die Erstellung umfangreicher Wörterbücher oder Konkordanzen benötigt bekanntlich sehr lange Lauf- und Verweilzeiten im Rechner, und wie oft hat sich ein Anwender solcher Systeme schon geärgert, wenn er nach vielen Stunden, ggf. sogar Tagen, seinen Output aus dem Rechenzentrum erhielt und dieser nicht dem beabsichtigten Ergebnis entsprach, weil bei der Parametrierung ein kleiner Fehler unbemerkt blieb oder ein Parameter falsch verstanden wurde.

CORDA erstellt nicht nur simple Wortformenlisten, sondern auch sog. rückläufige Wörterbücher, die nach Endungen oder auch eingeschränkt auf ganz bestimmte Endungen sortiert sind. Es können lemmatisierte Wörterbücher hergestellt werden, wobei zu Wortformen über einen eingebauten Spezialeditor die entsprechenden Grundformen erzeugt werden. Für Wortschatz-Untersuchungen oder Textlinguistik lassen sich auch Frequenz-Wörterbücher herstellen, die das Häufigkeits-Vorkommen absolut und/oder relativ in einem Dokument oder einer ganzen Dokumentensammlung angeben. Es lassen sich schließlich auch nahezu beliebige Arten von Konkordanzen herstellen, (zentrierte) Zeilen- und Satzkonkordanzen mit den verschiedenen Darstellungsmöglichkeiten KWIC (keyword in context), KWAC (keyword and context) und KWOC (keyword out of context).

Das System enthält noch eine ganze Reihe weiterer Möglichkeiten, die sich in diesem Rahmen nicht alle darstellen lassen; z.B. können über Negativlisten oder Positivlisten bestimmte Wörter von der Verarbeitung ausgeschlossen bzw. nur solche zugelassen werden. Die Verarbeitungsmöglichkeit ganzer Dokumentenbasen erlaubt

es darüberhinaus, Deskriptorlisten für Literaturdatenbanken zu generieren oder die Voraussetzung für einen Thesaurusaufbau zu liefern. Nicht unterstützt wird allerdings bislang die Relationierung von Thesaurusbegriffen (Unterbegriff/Oberbegriff-Relation, Synonyme etc.).

Transfer von Microcomputer-Daten auf Fotosatzanlagen

Obwohl Textverarbeitungen auf Microcomputern schon sehr leistungsfähig sind, kann damit der Fotosatz nicht ersetzt werden. Die wichtigsten Gründe dafür sind:

- Die meisten Textverarbeitungen bieten nicht bestimmte wichtige Leistungen von Fotosatzanlagen (z.B. zu wenig Schrifttypen und zu wenig Schriftgrößen u.ä.).
- Die Druckerausgaben eignen sich bei höheren Ansprüchen (z.B. Büchern, Werbeunterlagen) nicht als Vorlagen für fotomechanische Reproduktion (z.B. zu kleine Vielfalt an Schrifttypen, zu schlechter Schwärzungsgrad, keine echte Proportionalschrift; diese Restriktionen reduzieren sich allerdings mit großem Tempo durch neue, immer erschwinglichere Technologien, z.B. Laserdrucker, die von manchen Textverarbeitungen bereits unterstützt werden).

Andererseits ist jedoch die Texterfassung und vor allem die Überarbeitung und Korrektur eines Textes bei Microcomputern erheblich komfortabler und schneller, als auf einer Fotosatzanlage. Nach den bisherigen Erfahrungen hat sich deshalb in vielen Fällen folgende typische Situation herausgebildet:

- Der Text wird auf Microcomputer erfaßt und bis zur Endfassung redigiert. Es werden auch bereits die typographischen Codes miterfaßt (oder nachträglich auf Microcomputer eingefügt), da das Einfügen typographischer Codes auf Fotosatzanlagen relativ langsam vor sich geht oder sogar je nach Anlage Umbruchberechnungen auslöst, die nicht unterdrückt werden können und damit die Eingabe bei umfangreicheren Texten stark verlangsamt.
- Der Text muß anschließend irgendwie auf die Fotosatz-Anlage transferiert werden. Meistens geschieht das heute durch Austausch von Disketten: Der Fotosatz erhält vom Kunden eine Diskette, die von der Fotosatz-Anlage gelesen werden kann.

Der Weg hat jedoch mehrere Nachteile:

- Die Fotosatz-Anlage kann keine auf modernen Microcomputern eingesetzten Disketten lesen, wenn sie mit hardsektorierten Disketten arbeitet (z.B. Berthold tps6000).
- An die Fotosatz-Anlage muß ein bestimmtes Disketten-Format geliefert werden. Das führt in Anbetracht unterschiedlicher Disketten-Größen (8'', 5 1/4'' und 3 1/2''), ein- und doppelseitigen Disketten sowie verschiedenener Schreibdichten zu Problemen.

Sofern diese Probleme überhaupt gelöst werden, entsteht oft folgende Situation:

- Der Kunde ist auf eine bestimmte Fotosatz-Firma weitgehend festgelegt, die sein Format verarbeiten kann.
- Umgekehrt kann ein Fotosatz-Unternehmen zahlreiche Aufträge nicht annehmen, die nicht im richtigen Disketten-Format geliefert werden können.

Aufgrund dieser Situation bemüht man sich derzeit, Kopplungen zwischen Microcomputern und Fotosatz zu ermöglichen. Allgemein verwendbare Lösungen sind aber noch sehr selten oder - auch für mittlere Fotosatzfirmen - oftmals zu teuer; beispielsweise kosten Disketten-Formatkonverter, die eine große Anzahl von Diskettenformaten unterstützen, ca. 30.000.- DM (vgl. Anm. 7).

Eine billigere und trotzdem flexiblere Lösung ist die Koppelung einer Fotosatzanlage mit einem Microcomputer. Der Microcomputer kann dabei selbst noch als Erfassungsplatz eingesetzt werden, ferner wird preisgünstige Standardsoftware, auch spezifische Branchenlösungen für Auftragsverwaltung, Finanzbuchhaltung, Kalkulation etc. angeboten, und vor allem können über Konvertierprogramme für wenige hundert DM Kaufpreis fast alle heute für Personalcomputer üblichen Diskettenformate verarbeitet werden (vgl. Anm. 8).

Bei der Übertragung vom Microcomputer auf die Fotosatzanlage können spezifische Steuerzeichen einer Textverarbeitung herausgefiltert werden, solche Codes können aber auch von dem Transfer-Programm geeignet interpretiert und in die typographischen Codes automatisch übersetzt werden (vgl. Anm. 9). Damit wird ein Großteil der Satzkosten eingespart und die Produktionsgeschwindigkeit wird erheblich gesteigert.

(Dieser Beitrag ist ein Auszug aus dem Buch von **Ammon, R.v.; Kopelent, U.; Müller-Zantop, S.: Textverarbeitung auf IBM-Personalcomputern**. Erscheint im Carl Hanser Verlag/München.

Eine Demo-Textverarbeitung sowie Demo-Disketten zu Rechtschreibhilfe, Silbentrennung, Index-Erstellung, (Frequenz-)Wörterbücher/Konkordanzen sind erhältlich bei ECO Institut für Electronische Communication und Organisation GmbH, Eichelberg 8, 8411 Pielmühle, Tel.: 0941/82509).

Anmerkungen:

- /1/ Sehr leistungsstarke Rechtschreibhilfen, die speziell auch auf die deutsche Sprache ausgerichtet sind und die bereits in der heute im professionellen Bereich geforderten Programmiersprache C verfügbar sind, werden von der Softex GmbH in Saarbrücken und von der Unternehmensberatung Dieckmann in Hannover-Waldemark angeboten.
- /2/ Teilweise kann die Rechtschreibhilfe der Softex GmbH bereits Grammatikfehler erkennen, z.B. Komma-Fehler, falsche Groß/Kleinschreibung auch nach Interpunktionszeichen und demnächst ggf. Rektions- und Numerus-Fehler.
- /3/ Z.B. wiederum von der Unternehmensberatung Dieckmann oder der Softex GmbH in Saarbrücken.
- /4/ Microcomputer-geeignete und sehr leistungsstarke Silbentren-

nungen, die ebenfalls wieder bereits in der Programmiersprache C zur Verfügung stehen, werden von der Unternehmensberatung Dieckmann und der Softex-GmbH - auch integriert mit der Rechtschreibhilfe - angeboten.

- /5/ ECOINDEX ist ein Produkt des ECO Instituts, 8411 Pielmühle.
- /6/ CORDA ist ein Produkt des ECO Instituts.
- /7/ Z.B. der GICO-Konverter der GEVA-GmbH in Aachen.
- /8/ Z.B. mit dem im Großhandel erhältlichen XENOCOPY oder dem DISKMANAGER von Omnitex in Rheinfelden.
- /9/ Ein solches Transfer-Programm ist das Programm TRANS des ECO Instituts.

SC - EIN INTELLIGENTES HILFSSYSTEM FÜR SINIX

Ch. Kemke

FR 10.2 Informatik IV
Universität des Saarlandes
Im Stadtwald 15
D-6600 Saarbrücken 11

Zusammenfassung

Der SINIX¹ Consultant SC ist ein natürlichsprachliches Hilfesystem für SINIX-Benutzer, das seit Mai 1985 am Lehrstuhl für Künstliche Intelligenz und Datenbanksysteme, Universität Saarbrücken, entwickelt wird². SC soll sowohl natürlichsprachliche Fragen des Benutzers über Konzepte und Kommandos des SINIX-Systems beantworten können, als auch aktiv, ohne explizite Aufforderung durch den Benutzer, Hilfestellungen und Ratschläge für die Bewältigung einer Aufgabe geben. Anhand eines ersten Systementwurfs wird die prinzipielle Arbeitsweise des SINIX Consultants, insbesondere die Aspekte Planung/Problemlösung und Repräsentation des SINIX-Wissens, beschrieben.

1 Einleitung

Die in den letzten Jahren stark zunehmende Verbreitung von Rechenanlagen hat zur Folge, daß ein ständig wachsender Personenkreis mit den darauf installierten Betriebs- und Anwendungssystemen umgehen muß. Die Komplexität dieser Systeme bedingt, daß der Benutzer nur unvollständige Kenntnisse darüber hat, welche Möglichkeiten und Funktionen sie bieten und wie er diese nutzen kann.

Manchmal fehlt ihm zur Lösung seiner Probleme nur eine spezifische Auskunft über ein Detail, in anderen Fällen benötigt er genaue Anweisungen darüber, wie er eine komplexe Aufgabe erledigen kann. Unkenntnis oder falsche konzeptuelle Vorstellungen von der Funktionsweise des Systems können zudem zu einer fehlerhaften oder umständlichen Arbeitsweise des Benutzers führen.

¹ SINIX ist ein UNIX-Derivat der Siemens AG.

² Die Entwicklung von SC wird partiell von der Siemens AG im Rahmen des Projektes 'Innovative Informationsinfrastrukturen' finanziert

Um bessere Kenntnis über das System zu erlangen, benötigt der Benutzer Informationen über dessen Konzepte, die Existenz von Kommandos, ihre Anwendung und Semantik. Die Benutzer von Rechenanlagen sind meistens darauf angewiesen, sich mehr oder weniger autodidaktisch in die Handhabung des Systems einzuführen. Die zur Verfügung stehenden Handbücher sind i.a. zu unstrukturiert und umfangreich, um den Anfänger oder Fortgeschrittenen adäquat unterstützen zu können. Die beste Informations- und Hilfsquelle ist oft der versierte Kollege, der Hinweise und Ratschläge geben und Fragen, die ein spezielles Problem betreffen, zielgerichtet beantworten kann.

Da kundige Fachleute oft nicht greifbar sind, ist es naheliegend, gerade im Hinblick auf die wachsende Verbreitung von Arbeitsplatzrechnern und Personal Computern, für diese Aufgabe spezielle Hilfesysteme einzusetzen. Der Anspruch an derartige Hilfesysteme, quasi einen menschlichen Berater zu ersetzen, kann durch konventionelle Schlüsselwort- oder Menü-basierte Systeme bei weitem nicht erfüllt werden. Ein intelligentes Hilfesystem muß in der Lage sein, dem Benutzer seine speziellen Fragen zu beantworten, ihn bei aktuell auftretenden Schwierigkeiten zu unterstützen, ihm bei der Lösung schwieriger Aufgaben zu assistieren und -noch weitergehend- als Tutor seine Arbeitsweise zu korrigieren und seinen Wissensstand zu verbessern. Diese Anforderungen bedingen nicht nur, daß das Hilfesystem Probleme in dem jeweiligen Anwendungsbereich lösen und darüber Auskunft geben kann. Es muß Fähigkeiten haben, die Zielsetzungen des Benutzers und seine konzeptuellen Vorstellungen vom Anwendungssystem zu erkennen, um sinnvolle, adäquate Hilfestellungen zu geben. Dabei soll das System nicht nur in der Lage sein, einen zusammenhängenden Dialog zu führen. In einigen Situationen ist es notwendig, daß das System die Initiative ergreifen und dem Benutzer unaufgefordert Ratschläge erteilen kann.

Es sind bis zum jetzigen Zeitpunkt im Rahmen von Forschungsprojekten einige Hilfesysteme konzipiert worden, die jeweils einen Teil der oben genannten Probleme angehen. Der UNIX Consultant UC [21], entwickelt von Wilensky et al. an der University of Berkeley, und der UNIX Computer Consultant UCC [8] von Douglass (Los Alamos National Laboratory) und Hegner (University of Virginia) sind Hilfesysteme für das Betriebssystem UNIX. Zielsetzung bei UC und UCC ist das Beantworten natürlichsprachlicher Anfragen über UNIX-Kommandos und -Konzepte. Das System WIZARD von Finin (CMU) und Shrager (University of Pennsylvania) unterstützt Benutzer des Betriebssystems VMS, indem es sie hinweist auf effektivere Kommandos, wenn sie zu lange oder ineffiziente Kommandofolgen bzw. Kommandos benutzen. WIZARD ist damit im Gegensatz zu UC und UCC ein aktives Hilfesystem. Als *aktiv* werden solche Hilfesysteme bezeichnet, welche die Fähigkeit besitzen, in Abhängigkeit vom Verhalten des Benutzers selbständig, ohne Aufforderung durch den Benutzer Hilfestellungen zu geben. *Passive* Hilfesysteme sind im Gegensatz dazu Hilfesysteme, bei denen der Benutzer explizit um Hilfe bitten muß; hierunter fallen z.B. die oben genannten Schlüsselwort-basierten Systeme und auch die meisten natürlichsprachlichen Hilfesysteme. Die Unterscheidung zwischen aktivem und passivem Hilfesystem wird gerade bei den Systemen PASSIVIST und AKTIVIST von Fischer et al. [12] getroffen. Es handelt sich hierbei um an der Universität Stuttgart entwickelte Hilfesysteme für den Editor BISO. PASSIVIST beantwortet natürlichsprachliche Anfragen zu BISO, wobei ggf. eine schrittweise Erklärung von Arbeitsabläufen stattfindet. AKTIVIST analysiert die vom Benutzer eingegebenen Befehlssequenzen und erzeugt eine Meldung, falls ein anwendbares komplexes Kommando oder eine spezielle Funktionstaste nicht verwendet wurde. ADVISOR [13], von

Genesereth am MIT entwickelt, unterstützt Benutzer bei der Einarbeitung in das Formelmanipulationssystem MACSYMA. Der Benutzer kann ADVISOR um Rat fragen, wenn er bzgl. des Effektes einer Kommandofolge in seiner Erwartung enttäuscht wurde. ADVISOR versucht dann, mittels eines Debuggers den Fehler im Plan des Benutzers zu finden. MENO [22] von Woolf und McDonald (University of Massachusetts) ist ein Tutorssystem für PASCAL, das ebenfalls falsche konzeptuelle Vorstellungen des Benutzers vom Aufgabenbereich entdecken soll. Grundlage für MENO's Arbeitsweise ist eine Wissensbasis, die sowohl korrektes Wissen über Programmierung in Form von Plänen enthält, als auch typisches fehlerhaftes Wissen als Abweichung von diesen Plänen. Dieses Wissen wird zur Erzeugung von Tests verwendet, in denen überprüft werden kann, ob der Benutzer ein falsches Verständnis grundlegender PASCAL-Konzepte hat. Ist dies der Fall, gibt MENO eine Erklärung des mißverstandenen Konzeptes aus.

Hauptaspekt der genannten Hilfe- bzw. Tutoringsysteme ist das Beantworten natürlichsprachlicher Anfragen über Funktionen und Konzepte des jeweiligen Systems und/oder das Erteilen von Ratschlägen bei einer fehlerhaften oder ineffizienten Arbeitsweise. Dies ist auch die globale Zielrichtung bei der Entwicklung des SINIX Consultants SC. SC soll ein intelligentes Hilfesystem für SINIX-Benutzer sein, das sowohl natürlichsprachliche Anfragen über die Definition von SINIX-Konzepten und -Kommandos und deren Verwendung beantwortet, als auch eine Erkennung des globalen Benutzerziels betreibt und ggf. einschreitet, **um** dem Benutzer unaufgefordert Hilfestellungen bei der Durchführung seiner Aufgabe zu geben.

Entsprechend der weitreichenden Aufgabenstellung sind bei der Entwicklung eines solchen Beratungssystems verschiedene Teil- und Spezialgebiete der Künstlichen Intelligenz involviert. Dies sind im einzelnen die Analyse und Generierung natürlichsprachlicher Ausdrücke, Dialogführung, Planerkennung, Plangenerierung, Tutoring, Benutzermodellierung und -zentral für das gesamte System- Wissensrepräsentation. Zudem sind tiefgehende Kenntnisse über das Anwendungsgebiet erforderlich, insbesondere, um ein geeignetes Modell des Anwendungssystems zu erstellen und um Pläne zur Durchführung typischer Aufgaben herauszufinden. Dieses Spektrum an beteiligten Themenbereichen kennzeichnet die Problematik und den Forschungscharakter der Entwicklung eines intelligenten Hilfesystems wie SC. Anforderungen an SC, Ideen zur Konzeption und damit verbundene Problematiken werden in vorliegendem Papier anhand eines ersten Entwurfs skizziert.

2 Überblick über SC

Der SINIX Consultant SC soll -wie bereits in der Einleitung angedeutet- ein intelligentes Hilfesystem für SINIX-Benutzer sein, das sowohl in einem passiven Modus als natürlichsprachliches Hilfesystem arbeitet, als auch aktiv beratend und unterstützend tätig wird.

Der Benutzer kann in natürlicher Sprache formulierte, schriftliche Fragen über die Funktion und Konzeption von SINIX stellen. Die natürlichsprachlichen Eingaben des Benutzers werden in SC einem Analyseprozeß unterworfen, der als Ergebnis eine **semantische** Repräsentation des pragmatischen Gehalts der Äußerung liefert. Der anschließende Planungs- bzw. Problemlösungsprozeß extrahiert die zur Beantwortung notwendigen Inhalte aus der SINIX-Wissensbasis. Die von SC erzeugte Antwort wird aus einer internen, formalen Repräsentation generiert, evtl. unter Zuhilfenahme fester Antwortmuster und vorgefertigter Texte (z.B. Be-

schreibung der Struktur von Objekten oder der Wirkung von Kommandos). In seiner Funktion als Hilfesystem ist es nicht ausreichend, daß SC nur einzelne Fragen analysieren und beantworten kann. Es muß vielmehr in der Lage sein, eine kooperative Kommunikation mit dem Benutzer zu führen. Dazu müssen das vorherige Dialoggeschehen und individuelle Eigenschaften des jeweiligen Benutzers (seine Ziele, sein spezieller Wissensstand usw.) berücksichtigt werden. Informationen hierüber finden sich im Kontextmodell und im Benutzermodell. Das Kontextmodell enthält Darstellungen referierter Objekte, Aktionen und Ereignisse mit zusätzlicher semantischer und syntaktischer Information. Das Benutzermodell spiegelt den Wissensstand und die Intentionen des Benutzers wider.

Als aktives Hilfesystem soll SC Pläne und Zielsetzungen des Benutzers erkennen und ggf. Ratschläge für die Bewältigung einer Aufgabe geben, d.h. Hinweise auf Kommandos oder Pläne, die dem Benutzer vermutlich nicht bekannt sind und die Durchführung seiner Aufgabe erleichtern oder überhaupt erst ermöglichen. Hierzu betreibt SC eine Plan- bzw. Zielerkennung auf der Basis der vom Benutzer eingegebenen Kommandos bzw. seiner Fragen an SC. Hat SC den Plan bzw. das Ziel des Benutzers erkannt, kann die Planungskomponente tätig werden, um einen (u.U. besseren) Plan zu generieren. Dieser Plan wird von der Antwortgenerierung in eine adäquate Form umgesetzt, die dem Benutzer als Hilfestellung präsentiert werden kann.

Das von SC zur Problemlösung und Planung benötigte Wissen über das SINIX-System ist explizit formuliert in separaten Wissensquellen gespeichert. Die Wissensrepräsentation beruht auf einem Modell der SINIX-Benutzeroberfläche, in dem Objekte und Kommandos als Spezialisierungen genereller Konzepte dargestellt sind. Das SINIX-Modell dient als Basis einer Semantik-Definition der SINIX-Kommandos, die den Effekt bzw. das Ziel von Kommandos beschreibt. Die Darstellung des Wissens über das SINIX-System wird zur Beantwortung von Fragen, zur Planung und zur Planerkennung benötigt.

Die vorläufige Konzeption der Systemarchitektur ist in Abb.1 dargestellt. Die Funktionsweise der einzelnen Komponenten und die benötigten Daten- bzw. Wissensbasen werden in den folgenden Abschnitten beschrieben.

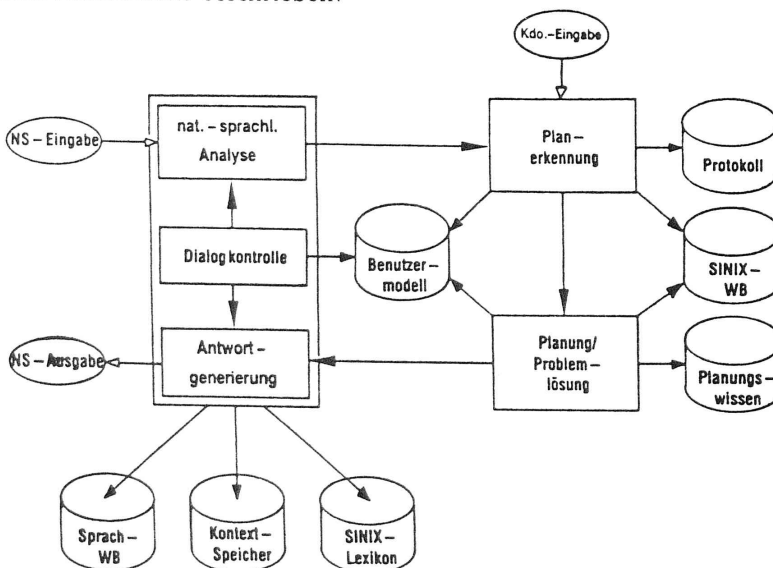


Abb.1 : SC - Systemarchitektur

2.1 SC - Komponenten

2.1.1 Natürlichsprachliche Analyse

Die natürlichsprachliche Schnittstelle bringt viele generelle Probleme der Analyse und Generierung natürlicher Sprache mit sich (z.B. Bildung und Auflösung von Referenzen und Ellipsen), einige Phänomene treten in dieser Anwendung jedoch in besonderem Maß auf, z.B.

- die Verwendung von Synonymen, Paraphrasen und Umschreibungen

Ein typisches Problem ist, daß der Benutzer einen Kommandonamen nicht kennt, aber er kennt eine entsprechende Bezeichnung einer anderen Kommandosprache oder er benutzt eine umgangssprachliche Formulierung (z.B. 'aufhören' oder 'logout' statt 'end').
- die Einschränkung auf bestimmte Fragetypen

Es treten vermutlich im wesentlichen nur wenige verschiedene Frageformen auf, die Informationen über die Definition von Konzepten bzw. Kommandos und die spezielle Verwendung von Kommandos, insbesondere mit zusätzlichen Bedingungen, fordern. Frageformen sind z.B. "Wie kann ich/man ..." (z.B. "Wie kann ich das Dateiverzeichnis wechseln?") und "Wie ..." (z.B. "Wie wechsle ich das Dateiverzeichnis?") mit Ergänzung zu einer Frage nach Struktur oder Definition eines Kommandos oder Konzeptes (in diesem Fall die Umschreibung des cd-Kommandos durch den Terminus 'Dateiverzeichnis wechseln').
- die Formulierung von Fragen unter Verwendung intentionaler Ausdrücke

Oft treten Fragen als indirekte Sprechakte auf, die durch intentionale Ausdrücke charakterisiert sind, z.B. "Ich möchte ...", "Ich will ...", "Ich beabsichtige ..." (z.B. "Ich möchte eine Datei löschen."), oder explizit eine Erklärung fordern, z.B. "Ich möchte wissen ...", "Kannst du mir sagen ..." (z.B. "Ich möchte wissen, wie man eine Datei löscht.") und ähnliches. Diese Ausdrücke, an die sich eine Formulierung entsprechend der oben erwähnten Fragetypen anschließt, können als Idiome verarbeitet werden.
- die Verwendung von Schlüsselwörtern

Die Einschränkung der Diskurswelt auf das SINIX-System bringt einige Erleichterungen bzgl. der natürlichsprachlichen Analyse mit sich. Da es sich um ein technisches Gebiet mit einer entsprechenden Fachterminologie handelt, ist die Bedeutung spezifischer SINIX-Begriffe i.a. klar definiert. Mehrdeutigkeiten und die Schwierigkeiten der Semantik- Definition, wie sie z.B. in Systemen auftreten, die mit Alltagswissen oder Wissen über geistes- bzw. gesellschaftswissenschaftliche Bereiche umgehen, treten bei SC nur in geringem Maß auf.

Die lexikalische, syntaktische und semantische Analyse der Anfrage erfolgt auf der Basis sprachbezogener Wissensquellen unter besonderer Berücksichtigung der Klassifizierung in Fragetypen. Synonyme und Umschreibungen können anhand des SINIX-Lexikons als spezifische, SINIX-betreffende Termini erkannt und durch Standardformen ersetzt werden. Ellipsen und Referenzen werden durch Zugriff auf den Kontextspeicher soweit möglich aufgelöst. Die erkannte Frageform und die angesprochenen SINIX-Begriffe sind Grundlage einer formalen Repräsentation der Anfrage, d.h. einer Formalisierung der Benutzerintention hin-

sichtlich des propositionalen Teils des Sprechaktes. Im Anschluß an die Analyse der Anfrage wird der Kontextspeicher anhand der erkannten SINIX-Begriffe aktualisiert.

Beispiel 1

Ben: "Wie kann ich jemandem eine Nachricht schicken?"

Frageklassifizierung :

'Wie kann ich ...' - Frage nach Kommando-Definition/Plan

Durch Zugriff auf das SINIX-Lexikon können für die Repräsentation relevante semantische Konstituenten durch Standardbegriffe ersetzt werden :

jemand \Rightarrow Benutzer

Nachricht \Rightarrow Nachricht

schicken \Rightarrow schicken

Nachricht schicken \Rightarrow Post senden / mail <Ben.kennung> (xsend/write)

Der Kontext-Speicher wird um die erkannten Begriffe und zusätzliche semantische und syntaktische Informationen erweitert :

Post senden - Aktion (Superkonzept)

Nachricht - Objekt der Aktion (Akkusativ-Objekt)

schicken - Aktionsverb

Benutzer - Beteiligter (Dativ-Objekt)

Ben: "Und eine geheime?"

Zur Ergänzung eines unvollständigen Ausdrucks wird auf den Kontext-Speicher zugegriffen. Elidierte syntaktische Konstituenten können unter Berücksichtigung syntaktischer und semantischer Restriktionen aus dem Kontextspeicher ergänzt werden.

Bindung des Adjektivs an das Akkusativ-Objekt :

geheime -> Nachricht

Die vollständige Ergänzung liefert :

Wie kann ich jemandem geheime Post senden?

Die besondere Bedeutung der Frageformen und der speziellen SINIX- Terminologie bzgl. der semantischen Interpretation der Anfrage legt die Verwendung eines semantisch-orientierten, pattern-basierten Parsers nahe; eine vollständige syntaktische Analyse wird in vielen Fällen nicht notwendig sein. Es muß allerdings auch berücksichtigt werden, daß der Benutzer zur Klärung seines Problems nicht nur einfache Fragen stellt. Möglicherweise entwickelt sich ein Dialog, in dem der Benutzer Rückfragen stellt, auf Fragen von SC antwortet, oder eine Folge von Aussagen über seine vorausgegangenen Arbeitsschritte macht, um seine Problematik darzustellen.

2.1.2 Planerkennung

Die Ziel- bzw. Planerkennung in SC ist abzugrenzen von dem, was in natürlichsprachlichen Systemen herkömmlicherweise unter diesen Begriffen verstanden wird. Es handelt sich bei SC um eine Planerkennung hinsichtlich der Benutzeraktionen mit dem SINIX-System, d.h. welches Ziel der Benutzer durch Anwendung von Kommandos bei seiner Arbeit auf

dem SINIX-System erreichen will, und nicht um eine Zielerkennung hinsichtlich der Dialogführung. Fragen und Äußerungen des Benutzers müssen jedoch in der Planerkennung als wesentliche Hinweise auf die Benutzerintention aufgenommen werden. Bei anderen Systemen (z.B. UC) wird unter Zielerkennung oder Zielanalyse die Extraktion der Benutzerintention hinsichtlich des propositionalen Teils des Sprechaktes, z.B. bei indirekten Sprechakten, verstanden. Diese Art von Zielerkennung wird in SC zum Teil von der Sprachanalysekomponente übernommen (Verarbeitung indirekter Sprechakte), in subtileren Fällen jedoch auch von der Planerkennungskomponente bzw. einem Submodul 'Zielerkennung', das zu diesem Zweck auf das Benutzermodell zugreift. Das Benutzermodell kann diesbezüglich aufschlußreich sein, da es generelle Intentionen eines durchschnittlichen Benutzers sowie Zielsetzungen des individuellen Benutzers enthält.

Die Planerkennung in SC basiert in erster Linie auf der Folge abgesetzter Kommandos und außerdem auf den gestellten Fragen des Benutzers, die zusammen als Historie des Benutzerhaltens in der Protokoll-Datei festgehalten sind.

Beispiel 2

Ben: {erstellt Datei}

Ben: crypt <Dateiname> {Datei verschlüsseln}

Ben: cat <Dateiname> | mail <Ben.kennung> {Datei als Post senden}

SC nimmt an : Intention des Benutzers ist 'Senden geheimer Post', erkannt durch die Folge 'crypt, cat, mail'. SC erteilt daraufhin einen Ratschlag :

SC : "Wenn Sie jemandem geheime Post senden wollen, ist es einfacher, das xsend-Kommando zu benutzen."

Hat SC den Plan eines Benutzers und somit auch sein globales Ziel erkannt, kann es z.B. durch Vergleich mit in der Wissensbasis gespeicherten Plänen feststellen, ob es eine bessere Möglichkeit gibt, das Ziel zu erreichen, als den vom Benutzer angewandten Plan. Es kann z.B. eine umständliche, zu lange Kommandofolge durch Verwendung eines speziellen Befehls oder des pipeline-Konzeptes kürzer und effizienter formuliert werden. Existiert ein besserer Plan, generiert SC autonom einen Hinweis für den Benutzer, der ihm z.B. in einem separaten Fenster - evtl. erst nach expliziter Aufforderung - gezeigt wird.

2.1.3 Planung/Problemlösung

Fragen an SC können bzgl. ihrer Zielrichtung und ihres Lösungsaufwandes in drei Kategorien eingeteilt werden :

- Fragen, die Definitionen betreffen

Die Problemlösung ist hier durch einfachen Zugriff auf konzeptuelles Wissen möglich, z.B. Zugriff auf die Wissensbasis anhand eines Schlüsselwortes, welches das angesprochene Objekt oder Kommando eindeutig bezeichnet.

Beispiel 3

Ben: "Was ist eine Nachricht?"

SC : "Eine Nachricht besteht aus einer Kopfzeile, die Absender, Datum und Uhrzeit enthält, und einem Text."



Ben: "Wie schicke ich jemandem Nachrichten?"

SC: "Sie können Nachrichten schicken mittels des mail-Befehls in der Form 'mail <Ben.kennung>'. Andere Befehle sind xsend und write."

SC generiert nur den ausführlicheren Hinweis für den mail-Befehl, da er der Standardbefehl zum Verschicken von Post ist; xsend und write werden als speziellere Befehle angesehen.

- Fragen, die die Anwendung von Kommandos betreffen

Hierbei ist ein assoziativer Zugriff auf eine Wissensbasis mit (mehreren) Attributwerten notwendig, die z.B. den Haupteffekt des Kommandos oder Objekte, die an der Aktion beteiligt sind, spezifizieren.

Beispiel 4

Ben: "Wie kann ich ältere Post lesen?"

Post lesen \Rightarrow mail

ältere Post \Rightarrow zeitl. vorausgehende Post \Rightarrow Inhalt (Briefkasten (Zeiger-auf-Nachricht - i)) , $i = 1, \dots, \text{Zeiger-auf-Nachricht}$

Es wird nach Kommandos gesucht, deren Effekt die Ausgabe älterer Post ist, wobei eine formale Repräsentation dieses Effektes als Suchindex dient.

Mögliche Befehle: mail -r; mail -p; mail mit Eingabe-Taste oder '+', 'q'.

SC: "Sie verwenden hierzu den mail-Befehl mit oder ohne Parameter. Geben Sie 'mail' an, so wird die zuletzt erhaltene Nachricht ausgegeben; wenn '?' erscheint, können Sie durch Drücken der Eingabe-Taste oder Eingabe von '+' die jeweils nächstältere Nachricht lesen oder durch Eingabe von 'q' die Ausgabe der Post beenden. Bei Eingabe von 'mail -p' wird der gesamte Briefkasteninhalt ohne Abfrage ausgegeben. Bei Eingabe von 'mail -r' werden die Nachrichten in der Reihenfolge ausgegeben, in der sie gesendet wurden."

- Fragen, die komplexe Aktionen/Pläne betreffen

Diese Fragen können nur durch Zugriff auf Pläne mit Hilfe von Planungs- oder Problemlösungsprozessen bewältigt werden.

Beispiel 5

Ben: "Wie kann ich verhindern, daß jemand meine Post liest?"

Post lesen \Rightarrow mail, xget

Das Verhindern einer Aktion bzw. eines Plans kann semantisch dargestellt werden dadurch, daß die Vorbedingungen zur Ausführung des Plans nicht erfüllt sind. Vorbedingung für die Anwendung des xget-Kommandos ist, daß der Benutzer das Schlüsselwort kennt. Eine andere Möglichkeit ist, Plänen und evtl. auch Konzepten einen Zweck zuzuordnen. Z.B. wäre der Zweck des Konzeptes 'Geheime Post', daß andere Benutzer die Post des betreffenden Benutzers nicht lesen können. Das Konzept 'Geheime Post' beinhaltet einen Plan mit den Kommandos enroll, xsend und xget. Erste durchzuführende Aktion ist also das Festlegen eines Schlüsselwortes durch das enroll-Kommando.

SC: "Legen Sie ein Schlüsselwort für geheime Post an."

Oder, falls SC davon ausgeht, daß der Benutzer das Konzept 'Geheime Post' nicht kennt :

SC : "Wissen Sie, was 'Geheime Post' ist?"

Der Benutzer reagiert mit einer entsprechenden Frage :

Ben: "Was ist geheime Post?"

SC antwortet mit einer Erklärung des Konzeptes 'Geheime Post', das mehrere Kommandos beinhaltet :

SC : "Post an Sie kann verschlüsselt werden. Dazu müssen Sie durch Angabe von 'enroll' die Festlegung eines Schlüsselwortes veranlassen. Schickt jemand Ihnen Post mit 'xsend', so können Sie diese nur durch 'xget <Schlüsselwort>' lesen."

Es besteht die Möglichkeit, Pläne explizit als Pseudo-Kommandos oder Konzepte abzuspeichern. Nachteil hierbei ist die aufwendige Wissensakquisitionsphase, in der die Pläne aufgestellt werden müssen, und eine mangelnde Flexibilität des Systems: In Situationen, in denen es mit nicht-gespeicherten Plänen konfrontiert wird, kann es nur mit einem Ausfall reagieren; eine 'kreative' Problemlösung kann nicht erfolgen. In SC ist jedoch eine flexible, 'intelligente' Planungskomponente vorgesehen, die autonom neue Pläne generiert.

Wesentlich für die Planungskomponente ist die semantische Repräsentation der Kommandos. Ähnlich wie in UCC werden in SC Kommandos u.a. durch Angabe von Vor- und Nachbedingungen und eine formale Beschreibung ihrer Semantik charakterisiert. Es ist allerdings zu beachten, daß diese Semantik-Beschreibung auf einem konzeptuellen Niveau erfolgt (nämlich der Vorstellung des Modellierers von der Funktionsweise des Systems) und somit abstrakt und unvollständig ist. Eine vollständige Semantik-Beschreibung hätte eine weitgehende Modellierung oder sogar Rekonstruktion des SINIX-Kommandointerpreters zur Folge.

In SC wird die Semantik-Beschreibung der Kommandos unterteilt in Haupt- und Nebeneffekte. Der Haupteffekt bezeichnet die in erster Linie im allgemeinen verfolgte Absicht des Benutzers, wenn er das Kommando eingibt (z.B. ist der Haupteffekt des rm-Kommandos, daß die benannte Datei nicht mehr existiert). Nebeneffekte sind Auswirkungen des Kommandos, die i.a. nicht Hauptanliegen des Benutzers sind (z.B. hat das Dateiverzeichnis nach Löschen einer Datei einen Eintrag weniger), aber unter gewissen Gesichtspunkten wesentliche Zustandsänderungen zur Folge haben können (z.B. kann ein leeres Dateiverzeichnis gelöscht werden).

Die Plangenerierung in SC basiert auf der Idee, Kommandos zu Plänen zu komponieren, wobei sich die Semantik der entstehenden Kommandofolge aus einer 'Verrechnung' der Haupt- und Nebeneffekte der beteiligten Kommandos ergibt. Die Verkettung von Kommandos durch pipelining wird bei dieser Betrachtungsweise eine spezielle Art der Komposition sein.

2.1.4 Antwortgenerierung

Die Antwortgenerierung gestaltet sich in einem Hilfesystem wie SC komplexer als in vielen anderen natürlichsprachlichen Systemen, da Erklärungen entsprechend dem Wissensstand des Benutzers, der aktuellen Dialogsituation und allgemeinen pädagogischen Maßstäben abgefaßt werden sollen. Es muß z.B. berücksichtigt werden, daß der Benutzer Konzepte nicht kennt, die bei der Beantwortung einer Frage verwendet werden müssen. In einigen Fällen ist eine

Überbeantwortung von Fragen angebracht; in anderen Fällen sollte eine Antwort partitioniert werden (z.B. Erklärung eines komplexen Plans), so daß der Benutzer zwischenzeitlich entsprechende Aktionen durchführen kann oder daß er, um weitere Auskünfte zu erhalten, erst explizit nachfragen muß (z.B. bei Hinweis auf weitere Kommandos oder detailliertere Erklärung). Dadurch soll vermieden werden, daß der Benutzer nicht benötigte Informationen erhält oder durch das Informationsangebot überlastet wird.

Zum Teil können Antworten aus vorgefertigten Texten bestehen; insbesondere bei Definitionen oder bei der Erklärung der Wirkung von Kommandos kann in die jeweilige Struktur, die das Objekt oder Kommando charakterisiert, ein Definitionstext als Attribut aufgenommen werden. Eine derartige Antwortgenerierung ist allerdings für ein komplexes Dialogverhalten und in dem Moment, wo das System selbständig Pläne erzeugt, die es dem Benutzer anbietet, nicht mehr ausreichend. SC muß dann in der Lage sein, aus einer formalen semantischen Repräsentationskonstruktion ggf. unter Verwendung von Antwortfragmenten einen natürlichsprachlichen Ausdruck zu generieren.

2.1.5 Dialogsteuerung

Es ist nicht ausreichend, daß SC einzelne Fragen analysieren und beantworten kann. In seiner Funktion als Hilfesystem muß es kooperative Kommunikationsfähigkeiten besitzen. Es muß in der Lage sein, einen zusammenhängenden Dialog über ein bestimmtes Themengebiet zu führen, eine längere Erklärung über die Funktionsweise eines Kommandos o.ä. zu geben, dem Benutzer beim Herausfinden einer Fehlerursache zu assistieren, Rückfragen des Benutzers zu handhaben, Erklärungen zu partitionieren und in Abhängigkeit vom Benutzermodell unterschiedlich zu gestalten und autonom Hinweise zu generieren, wenn es fehlendes oder falsches Wissen des Benutzers entdeckt.

Diese Anforderungen bedingen eine übergeordnete Kontrolleinheit, die in Abhängigkeit von der Dialogsituation die Analyse- bzw. Generierungskomponente anstößt und die für ihre Funktion notwendige Information über die Dialogsituation bereitstellt.

2.2 SC - Wissens- / Datenbasen

2.2.1 Sprachliche Wissensquellen

Die Sprach-Wissensbasis enthält das für natürlichsprachliche Systeme klassische morphologische, lexikalische, syntaktische und semantische Wissen, das von der Analyse- bzw. der Generierungskomponente benötigt wird, um aus einem natürlichsprachlichen Ausdruck eine formale semantische Repräsentation zu erzeugen und vice versa. Da die Sprach-Wissensbasis vom Diskursbereich unabhängig sein soll, um ggf. Module anderer Sprachverarbeitungssysteme übernehmen zu können, und da auf die speziellen SINIX-Begriffe noch von anderen Komponenten zugegriffen wird, wird die SINIX-Terminologie in einem getrennten Lexikon verwaltet.

2.2.2 SINIX-Lexikon

Das SINIX-Lexikon besteht aus einer Tabelle, in der SINIX-Begriffe und zugehörige Synonyme, z.B. entsprechende Bezeichnungen anderer Kommandosprachen, umgangssprachliche Umschreibungen usw., verzeichnet sind. Äquivalenten Ausdrücken wird im Lexikon eine

Standardform zugeordnet, die bei der weiteren Verarbeitung verwendet wird. Diese Standardform dient als eindeutige Kennung des betreffenden SINIX-Konzeptes in der Wissensbasis, so daß sie einen Schlüssel für den Zugriff auf die Wissensbasis darstellt. Das Lexikon kann nicht nur von der Analysekomponente genutzt werden, sondern auch bei der Sprachgenerierung, um einige Variationsmöglichkeiten für die Antwortformulierung bereitzustellen. Außerdem wird das Lexikon verwendet, um durch einen SC vorgeschalteten Filter erkennen zu können, ob es sich bei einer Benutzereingabe um ein SINIX-Kommando oder eine Anfrage an SC handelt (s. Abschnitt 3).

2.2.3 Kontextspeicher

Der Kontextspeicher enthält die während des Dialogs und in SINIX-Meldungen angesprochenen Objekte und Aktionen, d.h. insbesondere SINIX-Objekte, -Kommandos und -Konzepte, ggf. auch Satzfragmente oder verwendete Frageformen, zusammen mit syntaktischer und semantischer Information, durch die die Menge möglicher Referenzen eingeschränkt wird (s.a. 2.2.1).

Der Kontextspeicher wird aufgebaut aus Inhalten der Protokoll-Datei und Ergebnissen der Analyse- und der Generierungskomponente. Die Analyse- und die Antwortgenerierungskomponente greifen auf den Kontextspeicher zu, um Referenzen oder Ellipsen zu erkennen bzw. zu bilden.

2.2.4 Benutzermodell

Das Benutzermodell in SC spiegelt die Intentionen und den Wissensstand des aktuellen Benutzers wider. Ausgehend von einem à-priori-Benutzermodell, das Annahmen über den Wissensstand und allgemeine, grundlegende Intentionen eines durchschnittlichen Benutzers enthält, wird im Lauf des Dialogs ein individuelles Benutzermodell entwickelt. Grundlage für die Adaption des à-priori-Modells ist eine Klassifizierung möglicher Benutzer in Stereotypen gemäß unterschiedlicher Qualifikationen (z.B. absoluter Anfänger, grundlegende Kenntnisse vorhanden, kennt anderes Betriebssystem) und eine Strukturierung des Wissensbereichs, die i.w. dem Aufbau des Wissens in der SINIX-Wissensbasis entspricht. Anhand der vom Benutzer verwendeten und erfragten Begriffe, des von SC vermittelten Wissens und der von SINIX erzeugten Fehlermeldungen kann SC Rückschlüsse ziehen darüber, welche Konzepte dem Benutzer bekannt bzw. unbekannt sind. Anhand dieser Kenntnis kann der Benutzer einem entsprechenden Stereotypen zugeordnet werden. Für einen Stereotypen gelten bestimmte Vor/-annahmen über den Wissensstand der durch ihn repräsentierten Benutzerklasse, die von SC nun auch für den aktuellen Benutzer angenommen werden. Außerdem können durch Informationen über die Strukturierung des Wissensbereichs Rückschlüsse auf das Wissen bzw. Unwissen des Benutzers gezogen werden; z.B. kann SC annehmen, daß der Benutzer vom gesamten Message-System keine Kenntnisse hat, wenn ihm das grundlegende Kommando 'mail' unbekannt ist. Weiterhin kann SC schließen, daß der Benutzer insbesondere nichts vom Konzept 'Geheime Post' und den dazugehörigen Kommandos 'xsend', 'xget' und 'enroll' weiß.

Das Benutzermodell wirkt sich auch auf die Antwortgenerierung aus, da die Differenziertheit einer Antwort dem jeweiligen Wissensstand des Benutzers angepaßt sein soll.

2.2.5 Protokolldatei

Die Protokolldatei enthält eine aufbereitete Aufzeichnung sämtlicher Interaktionen zwischen Benutzer, SC und SINIX, also des gesamten Sitzungsverlaufs. Die Protokolldatei kann unterteilt werden in vier Bereiche :

- die vom Benutzer eingegebenen Kommandos
- die vom Benutzer gestellten Fragen (formale Repräsentation)
- die von SC generierten Antworten (formale Repräsentation)
- aufbereitete Meldungen von SINIX (insbes. Fehlermeldungen)

Die in der Protokolldatei gespeicherten Informationen werden verwendet

- zur Aktualisierung des Kontextspeichers

Der Benutzer kann während des Dialogs Begriffe oder Sachverhalte referieren, die in vorhergehenden Äußerungen erwähnt wurden. Im Unterschied zu herkömmlichen Dialogsystemen muß berücksichtigt werden, daß der Benutzer sich auch auf SINIX-Meldungen, also Texte, die nicht Bestandteile des Dialogs sind, beziehen kann.
- zur Anpassung des Benutzermodells

Die vom Benutzer verwendeten Kommandos, die von SC vermittelten Konzepte und die von SINIX gemeldeten Fehler lassen Rückschlüsse auf den Kenntnisstand des Benutzers zu, die in das aktuelle, individuelle Benutzermodell eingehen.
- zur Planerkennung

Um eine globale Benutzerintention, eine zielgerichtete Handlungsfolge erkennen zu können, beobachtet die Planerkennungskomponente die Sequenz der vom Benutzer eingegebenen Kommandos. Fragen des Benutzers können evtl. auch aufschlußreich sein, um einen von ihm verfolgten Plan zu erkennen.

2.2.6 SINIX-Wissensbasis

Zentral für SC ist die SINIX-Wissensbasis. Das Wissen über SINIX wird in Form eines Modells der Benutzeroberfläche dargestellt, wobei die Repräsentation von (virtuellen) Objekten und Kommandos grundlegend ist. Außer den 'realen' Systemobjekten (Dateien, Prozesse usw.) und spezifischen Kommandos werden auch 'virtuelle' Konzepte, die nicht eine direkte Entsprechung in SINIX-Konzepten finden, und Meta-Konzepte formuliert, z.B. generelle Konzepte für 'Kommando' und 'Aktion' oder ein Konzept für 'Geheime Post', das die zugehörigen Kommandos und Objekte subsumiert. Tätigkeiten, die mehrere einzelne Handlungsschritte beinhalten, können durch 'Pseudo-Kommandos' repräsentiert werden, denen Pläne, d.h. Sequenzen primitiver Kommandos, zugeordnet werden.

Die SINIX-Wissensbasis ermöglicht SC, Fragen über Struktur und Handhabung von SINIX zu beantworten. Diese können Aufbau und Definition von (virtuellen) Objekten und Kommandos betreffen, sowie die Anwendung von Kommandos (z.B. welche Parameter eingegeben werden müssen) und Zusammenhänge zwischen Einheiten der Wissensbasis (z.B. wird der Inhalt einer Briefkasten durch Senden einer Nachricht verändert, eine Nachricht beinhaltet einen Text, eine Postbox ist eine spezielle Datei usw.).

Die Zuordnung von Plänen, also Folgen primitiver Kommandos, zu Pseudo-Kommandos resp. Benutzerintentionen sind Grundlage der Planerkennung.

Unter Verwendung des SINIX-Modells soll die Planungskomponente in der Lage sein, selbständig einen Plan zu generieren, durch den der Benutzer das in seiner Anfrage formulierte Ziel erreichen kann. Notwendig hierzu ist eine formale Darstellung der Semantik einzelner Kommandos, aus der sich bei Komposition von Kommandos zu Handlungsfolgen resp. Plänen deren Semantik bestimmen läßt.

Vorbedingungen für die Anwendung von Kommandos oder Plänen sind wesentlich, um den Benutzer bei der Suche nach Fehlerursachen zu unterstützen.

Eine Unterteilung der Domäne in Themenbereiche, die sich bei der Wissensrepräsentation ergibt, kann bei der natürlichsprachliche Analyse zur Bestimmung des Fokus genutzt werden, wodurch z.B. bei Ambiguitäten zwischen den verschiedenen Interpretationen entschieden werden kann.

Wesentliche Zielsetzung bei der Wissensrepräsentation ist eine möglichst weitgehende Strukturierung der Domäne durch die Formulierung von Abstraktionen und Prinzipien, also durch Herausbildung von konzeptuellem Wissen. Dadurch wird zum einen eine geringe Redundanz der Darstellung erreicht; zum anderen ist die Verwendung derartigen Wissens grundsätzlich notwendig, um intelligentes Verhalten zu zeigen.

Grundlegend für das bisherige Vorgehen bei der SINIX-Wissensrepräsentation ist eine Unterteilung darzustellender Konzepte in eine Welt der Objekte und eine Welt der Aktionen. Diese Trennung erfolgt unter dem Gesichtspunkt, daß die Darstellung von Objekten in erster Linie auf einer Strukturbeschreibung basiert, während Aktionen durch Zustandsänderungen bei Objekten charakterisiert werden können. Bei der Formulierung von Repräsentationskonstruktionen für Kommandos ist zu berücksichtigen, daß diese sowohl einen statischen Aspekt (ihr syntaktischer Aufbau) als auch einen dynamischen Aspekt (ihre Semantik, das Bewirken einer Zustandsänderung) besitzen.

Die Wissensrepräsentation auf der Basis eines Modells der SINIX-Benutzeroberfläche ist aus verschiedenen Gründen nicht unproblematisch. Die Modellierung ist ein subjektiver Vorgang, bei dem der Designer sein konzeptuelles Modell von der Struktur und Funktionsweise des Systems festhält. Die konzeptuellen Vorstellungen von einem System sind für individuelle Benutzer unterschiedlich. In der Entwurfsphase muß ein Modell für die Repräsentation in SC ausgewählt werden. Nun kann der Benutzer des SC-Hilfesystems über ein anderes konzeptuelles Modell als SC verfügen, woraus sich u.U. 'Verständnisschwierigkeiten' beim Einsatz von SC ergeben.

Außerdem stellt sich die Frage nach dem Detailliertheitsgrad der Repräsentation, der davon abhängt, ob ein eher Maschinen-orientiertes, genaues Modell gewählt wird oder ein eher Benutzer-orientiertes, abstraktes Modell. Wird z.B. eine Generierung von Antworten auf der Basis einer formalen Semantikepräsentation von Kommandos - statt Ausgabe eines vorgefertigten Textes - angestrebt, müssen auch Auswirkungen eines Kommandos, die für den Benutzer nicht direkt sichtbar sind sondern auf Implementierungsebene liegen, in der Semantikdefinition berücksichtigt werden (z.B. Wechsel des Systemmodus bei Aufruf eines Editors). Dies kann zu einem aufwendigen, detaillierten Modell führen, das für die Formulierung und Handhabung der Repräsentationskonstruktionen zu komplex ist. Es ist also eine Einschränkung der Beschreibung auf wesentliche Merkmale notwendig, wodurch das Modell natürlich inexakt wird und zu Fehlinformationen führen kann.

Die bisherigen Ansätze zur Wissensrepräsentation in SC lagen vorwiegend auf dem Themenbereich 'Kommunikation mit Benutzern'. Es wurde insbesondere die Darstellung von Kommandos behandelt, da sie sowohl bezüglich der Beantwortung von Fragen von zentraler Bedeutung sind, als auch hinsichtlich der Wissensrepräsentation interessante Probleme liefern.

Es wurden zunächst frameartige Datenstrukturen definiert, in denen generelle und spezifische Kommandokonzepte abgebildet werden können. Parallel hierzu wird zur Zeit untersucht, inwieweit die aus KL-ONE entstandene Wissensrepräsentationssprache KRYPTON [2] [3] [4], die Vererbungsnetze und Prädikatenlogik als Darstellungsformen vereint, zur Repräsentation des SINIX-Wissens geeignet ist.

Zur Veranschaulichung ist in Abb.2 eine für das mail-Kommando gefüllte Struktur dargestellt, die die wesentlichen Attribute des Kommando-Konzeptes enthält. ¹

```

Name : mail
Superkonzept : Post-senden
  Parameter : Benken ; Nachricht
  Haupteffekt : push (Briefkasten (Benken, Nachricht))
  Nebeneffekt : increment (Zeiger-auf-Nachricht (Briefkasten (Benken)))
Subkommandos : Texteingabe
Kommandoende : '.' ; 'CTRL d'
Plankennzeichen : plan xxxx
  Vorbedingung : <status = 0>
  Nachbedingung : <status = 0>
  Funktion : Mittels des mail-Befehls können Sie Post senden. Geben Sie das Kommando in der Form 'mail <Benutzerkennung>' an. Anschließend geben Sie den Text ein. Die Nachricht wird durch Eingabe von '.' oder 'CTRL d' abgeschickt.
    
```

Abb.2 : Repräsentation eines Kommandos

2.2.7 Planungswissen

Hierunter fällt das Wissen darüber, wie Pläne erzeugt werden können. In der Literatur sind verschiedene Planungsmechanismen bekannt, die zum Teil in SC Anwendung finden können, z.B. Strategien zur Konfliktauflösung, kooperative Planung mehrerer Agenten (SC und Benutzer), hierarchisches und nicht-lineares Planen.

Im Hinblick auf die Plangenerierung in der SINIX-Domäne müssen zudem spezielle Operatoren definiert werden, die bei der Komposition von Kommandos zu Plänen deren Semantik bestimmen.

Es ist zur Zeit noch offen, ob fertige Pläne in die SINIX-Wissensbasis aufgenommen oder getrennt hiervon abgespeichert werden sollen.

Benken ist Abkürzung für Benutzerkennung

3 Implementierungsaspekte

Der SINIX Consultant wird auf einem SIEMENS-Arbeitsplatzsystem 5815 (identisch mit XEROX 1108-105) entwickelt, das speziell für Interlisp-D ausgelegt ist und darauf aufbauend das objektorientierte Programmiersystem LOOPS zur Verfügung stellt. Zur Unterstützung des Entwurfs und zur Präsentation der Systemarchitektur wurde das interaktive Programm PRÄSENT erstellt, mittels dessen Struktur und Funktion des Gesamtsystems sowie einzelner Komponenten anhand von Graphiken dargestellt und durch Beispiele erläutert werden können. Es wird angestrebt, das lauffähige, getestete Programm anschließend auf SIEMENS-Arbeitsplatzrechner PC MX-2 zu portieren. Zu diesem Zweck wird zur Zeit im SC-Subprojekt PORTFIX FranzLisp unter UNIX auf die MX-Rechner übertragen. Die Programmierung auf dem APS 5815 erfolgt zunächst in LOOPS; es wird erwogen für die Wissensrepräsentation KRYPTON oder eine äquivalente Repräsentationssprache zu installieren. Die Benutzeroberfläche des SINIX-Systems wird auf dem APS 5815 durch einen Kommando-Emulator nachgebildet, der SINIX-Kommandos erkennt und mit entsprechenden SINIX-Meldungen reagiert. Eine andere Möglichkeit besteht darin, das System SC auf dem APS 5815 zu belassen und die Kommunikation zu den PC MX-2 über ein lokales Netz (Ethernet mit TCP/IP) zu bewerkstelligen. Die SC-Version auf den MX-2 muß um eine Kommunikation zwischen SC und dem Betriebssystem erweitert werden. Die Anbindung von SC an SINIX erfolgt durch einen Filter, den die LOGIN-Shell nach Eingabe der Benutzererkennung aufruft, falls der Benutzer mit SC arbeiten möchte (s. Abb.3). Der Filter trennt Anfragen an SC und SINIX-Kommandos. Bei Kommando-Eingabe wird eine zweite Shell gestartet, die das Kommando ausführt; gleichzeitig wird das Kommando für die Planerkennung an SC übergeben. Natürlichsprachliche Anfragen werden nur an SC geleitet.

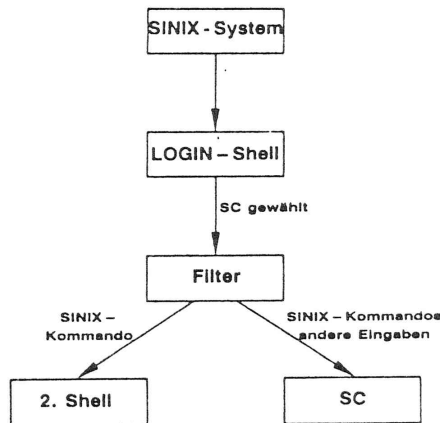


Abb.3 : Anbindung SC - SINIX

4 Schlußwort

Inzwischen wurde die Realisierung eines lauffähigen Prototypen begonnen. Momentane Arbeiten zur Prototyp-Erstellung sind die genaue Festlegung und Implementierung der Wissensrepräsentation. Diese soll vorerst auf den Themenbereich 'Kommunikation zwischen Benutzern' beschränkt werden, allerdings auch grundlegende Konzepte beinhalten.

Weiterhin wird ein Parser in Anlehnung an DYPAR entworfen und implementiert. Für die Prototyp-Version ist eine einfache Schnittstelle vorgesehen, die zur Verarbeitung direkter Fragen ausreicht. Weitergehende Kommunikationsfähigkeiten können durch Ausbau der Dialogsteuerung sowie Entwicklung und Integration des Benutzermodells erzielt werden. Die Planungs- bzw. Problemlösungskomponente muß die zur Beantwortung einfacher Fragen notwendigen Zugriffsoperationen auf die SINIX-Wissensbasis enthalten; komplexe Planungs- und Planerkennungsprozesse können zu einem späteren Zeitpunkt der lauffähigen Grundversion hinzugefügt werden. Die Antwortgenerierung arbeitet zunächst nur mit Antwortmustern und fertigen Definitionstexten, die den Konzepten zugeordnet sind.

Auf theoretischer Seite sind empirische Untersuchungen vorgesehen, die Aufschluß über Art und Aufbau der potentiellen Fragen an einen SINIX Consultant und die Wissensstrukturierung bei SINIX-Benutzern ergeben sollen.

Literatur

- [1] Arens, Y. : Using Language and Context in the Analysis of Text
Proc. IJCAI-81, Vancouver, British Columbia, 1981, S.52-57
- [2] Brachman, R.J. / H.J. Levesque : Competence in Knowledge Representation
Proc. AAAI-82, Pittsburgh, PA, 1982, S.189-192
- [3] Brachman, R.J. / J.G. Schmolze : An Overview of the KL-ONE Knowledge Representation System
Fairchild Technical Report No 655 , Sept. 1984
- [4] Brachman, R.J. / R.E. Fikes / H.J. Levesque : KRYPTON - Integrating Terminology and Assertion
Proc. AAAI-83, Washington, DC, 1983, S.31-35
- [5] Chin, D. : An Analysis of Scripts Generated in Writing between Users and Computer Consultants
AFIPS Proc. National Computer Conference 53, Las Vegas, NV, 1984, S.637-642
- [6] Chin, D. : Knowledge Structures in UC, the UNIX Consultant
Proc. 21th Meeting of the ACL, Cambridge, MA, 1983, S.159-163
- [7] Deering, M. / J. Faletti / R. Wilensky : PEARL - A Package for Efficient Access to Representations in LISP
Proc. IJCAI-81, Vancouver, British Columbia, 1981, S.930-932
- [8] Douglas, R.J. / S. Hegner : An Expert Consultant for the UNIX-Operating System: Bridging the Gap between the User and Command Language Semantics
Proc. CSCSI / SCEIO, Saskatoon, Saskatchewan, 1982, S.92-96

- [9] Faletti, J. : PANDORA - A Program for Doing Commonsense Planning in Complex Situations
Proc. AAAI-82, Pittsburgh, PA, 1982, S.185-188
- [10] Finin, T.W. : Providing Help and Advice in Task Oriented Systems
Proc. IJCAI-83, Karlsruhe, Germany, 1983, S.176-178
- [11] Finin, T.W. / J. Shrager : An Expert System that Volunteers Advice
Proc. AAAI-82, Pittsburgh, PA, 1982, S.339-340
- [12] Fischer, G. / A. Lemke / T. Schwab : Knowledge-Based Help Systems
Contribution to CHI '85, San Francisco, 1985
- [13] Genesereth, M.R. : The Role of Plans in Automated Consultation
Proc. IJCAI-79, Stanford, 1979, S.311-319
- [14] Hobbs, J.R. / D.A. Evans : Conversation as Planned Behavior
Cognitive Science 4, 1980, S.349-377
- [15] Rich, E. : Building and Exploiting User Models
Proc. IJCAI-79, Stanford, 1979, S.720-722
- [16] Rychener, M.D. : A Semantic Network of Production Rules in a System for Describing Computer Structures
Proc. IJCAI-79, Stanford, 1979, S.738-743
- [17] Steinacker, I. : Syntaktisches und semantisches Parsing und seine Integration im System VIE-LANG
Bericht der Österreichischen Studiengesellschaft für Kybernetik, 1984
- [18] Thomas, J.C. : Artificial Intelligence and Human Factors
IBM Research Report RC 10823, Computer Science, 1984
- [19] Wahlster, W. : Natürlichsprachliche Systeme
in W. Bibel / J.Siekmann (Hrsg.) : Künstliche Intelligenz Springer, Berlin, 1982, S.203-283
- [20] Wilensky, R. / Y. Arens : PHRAN - A Knowledge-Based Natural Language Understanding
Proc. 18th Meeting of the ACL, Pennsylvania, 1980, S.117-121
- [21] Wilensky, R. / Y. Arens / D. Chin : Talking to UNIX in English : An Overview of UC CACM, Vol 27, No 6, 1984, S.574-593
- [22] Woolf, B. / D.D. McDonald : Human-Computer Discourse in the Design of a PASCAL Tutor
COINS Technical Report 83-27, 1983

Ein Microcomputer-Programmsystem für *Autoren, Handbuch-Schreiber* u.ä.

ECOINDEX

für die Herstellung von
Stichwortverzeichnissen, Indizes, Registern

ECOINDEX verarbeitet alle und beliebig viele *ASCII-Textdateien*, also auch eine
ganze Datenbasis aus vielen Dokumenten;
ECOINDEX verarbeitet auch *Dateien von Textverarbeitungssystemen* wie *WordStar*
(von *MicroPro*) und *MS-WORD* (von *Microsoft*);
ECOINDEX kann auch *down loaded* Dokumente verarbeiten, die aus einem
Informationssystem *recherchiert* wurden.

ECOINDEX ist in der Programmiersprache „C“ geschrieben und läuft derzeit
unter den Betriebssystemen *MS-DOS/PC-DOS* und *CP/M-86, CCP/M* und allen
compatiblen Betriebssystemen.

ECOINDEX kostet nur 750,- DM (ohne MWSt.)

Ein Microcomputer-Programmsystem für die *Corpuslinguistik,*
Textlinguistik und alle *Wortschatz- und Textuntersuchungen*

CORDA

für die Herstellung von
Wortformenwörterbüchern, lemmatisierten Wörterbüchern, Frequenz-
wörterbüchern, rückläufigen Wörterbüchern,
(zentrierten) Zeilen- und Satzkonkordanzen,
KWIC, KWAC, KWOC-Ausgabeformate,
Wortstatistik (absolut, relativ)

CORDA besitzt ansonsten alle oben beschriebenen Leistungsmerkmale
von *ECOINDEX*.

CORDA kostet nur 1.000,- DM (ohne MWSt.)

Die Bedienung und Parametrierung beider Systeme ist durch selbsterklärende
Bedienführung und eingebaute Plausibilitätskontrollen *idiotensicher*.

Weitere Informationen bei

ECO Institut für *Electronische Communication* und *Organisation GmbH*
Eichelberg 8, 8411 Pielmühle, Telefon: (09 41) 8 25 09

**REMARKS
CONCERNING THE USE OF EXPERT SYSTEM APPROACH
IN INFORMATION RETRIEVAL**

G. Salton

Department of Computer Science
Cornell University
Ithaca NY 14843

1. Expert System Approach

In recent years it has become commonplace to suggest that an expert system approach be used as a solution to problems in many areas of computer applications. Expert systems are characterized by a dynamic adaptation to the problem and its environment, a multi-faceted system design in which the user may be asked to confirm or approve particular modes of action, and an incorporation of so-called expert capabilities including reasoning components that take available data and prevailing conditions to construct problem solutions.

In information retrieval environments, the expert approach is normally based on the use of stored knowledge bases characterizing the subject matter under consideration and the world-at-large, a natural language capability allowing the user to communicate with the system in natural language, a user characterization capable of specifying user requirements, and automated search and retrieval capabilities that can adapt the search effort to the user needs.

The knowledge base may be thesaurus-like in the sense that the main entries consist of concepts with connections to broader, more generic concepts, or to narrower concepts, or to other related terms. [1,2] Alternatively general semantic nets may be used where a great variety of different relationships can be specified between included terms or concepts. Rule sys-

tems have also been used for the representation of knowledge in particular fields [3], as have term-document networks where the terms are related to other terms and to documents, and documents are related to other documents and to terms. [4-5]

In an automated search system based on user-system interaction, the natural language system may be charged with the analysis of the natural language queries and with the management of user interviews and dialogues. In one recent proposal, noun groups are extracted from user query statements by morphological and syntactic methods, and the correctness or applicability of each extracted noun group is checked against thesaurus entries by a semantic validation process. [1] An approximate matching process is then used to compare the output of the natural language analysis system with the vocabulary provided for document indexing.

Normally, an attempt is made to characterize the user needs in terms of the amount of material to be retrieved (high recall versus high precision), the types of documents to be obtained (general or specific, practical or theoretical, etc.), the languages and publication dates of the documents, and so on. This can be done by conducting interviews with the users in the form of questions and answers prior to each search effort. [2,5-7] Alternatively, or in addition, long-term characterizations of particular users may be kept specifying the degree of user expertise in the subject area, and the degree of familiarity with the retrieval system. [5]

A search strategy is eventually chosen based on the user characterization, the output of the linguistic analysis performed on the user queries, and the use of the knowledge base to supplement, or refine, the query terminology proposed by the users. Among the possible search strategies that may be considered are text scanning methods, conventional inverted file searches, quorum searches based on the number of matching term combinations, as well as combinations of several different search strategies. [8-10]

In summary, the expert system approach is designed to provide more sophisticated tools for the retrieval of relevant items than are available in conventional retrieval situations, while at the same time offering transparency and user-friendliness by keeping the user population as a central part of the system.

2. Observations

In principle, the use of expert system approaches in information retrieval must be welcomed, because substantially better retrieval operations would be obtained if all the systems components were properly integrated and working. In practice one wonders, however, whether the existing efforts will really succeed for two main reasons:

- a) many of the postulated components are presently unavailable and their construction or development appears to be very difficult;
- b) none of the postulated components has been evaluated, and their effectiveness must be taken on faith at the present time.

Concerning the first point, one must recall that the linguistic tools needed to perform the free text query analysis are not currently at hand. Thus, it is very difficult to extract good noun groups from a query text, and it is even harder to classify diverse noun groups into sets of semantically equivalent phrases as is required in an automatic phrase matching system. The usability of completely unrestricted natural language inputs must therefore be questioned. Secondly, the required knowledge bases which describe the document and query domains are not normally available, and no consensus exists about how a knowledge base should be built and structured. In particular, the well-known ambiguities arise about the nature of the concepts incorporated into knowledge bases -- whether general or specific, assertive or descriptive, intensional or extensional and so on. Third, the know-how that must be used

in implementing the automated retrieval experts is not readily available beyond general conjectures about broad and narrow searches being useful respectively, to nonspecialists and specialist users.

A fundamental question also arises about the desirability of burdening the user with many difficult chores in the course of the retrieval process, such as assigning generality levels to terms [2], deciding on degrees of applicability of rules [3], and suggesting spelling variations or term synonyms. [1]

As an example of the uncertainty in effectiveness of some of the proposed expert system operations, consider the use of the knowledge bases for the expansion of query (or document) terminology. A desirable query reformulation step is often stated in quite simple terms: [1]

- a) if the query is judged too narrow, broaden it by replacement of the narrower terms by generically broader terms extracted from the knowledge base;
- b) if the query is judged too broad, construct a narrower query by using more specific terms from the knowledge base;
- c) if the query does not fit the document corpus, reformulate it using terms from the knowledge base that are semantically close.

There are of course initial difficulties about an accurate characterization of queries as narrow, or broad, or outside the corpus. The main problems arise, however, with what appears to be a relatively uncontrolled navigation in the knowledge base while still maintaining close control over the query terminology. Many years ago, procedures such as those now contemplated were in fact used in an experimental retrieval system. In particular, a hierarchical term arrangement was constructed for a sample collection of documents in electrical engineering (the IRE collection) and terms which occurred jointly

in queries and documents were "expanded" by adding new concepts that were

- a) hierarchically superior (expansion by parent)
- b) hierarchically inferior (expansion by children)
- c) hierarchically on the same level (expansion by sibling)
- d) related by cross-referencing.

The main conclusion at the time was that the contemplated vocabulary expansion operations were spectacularly unsuccessful. Some quotations from the evaluation report may be useful: [11]

- It is seen that in each case the standard thesaurus process alone (using no term expansion) is superior; the equality hypothesis (in the comparison of the methods) can be rejected unequivocally for the expansion by children, siblings and cross-references. ...
- A question exists only for the expansion by parent. This expansion does in fact improve retrieval performance for certain recall levels.
- In general, the term expansions tend to produce large-scale disturbances in the information identifiers attached to the documents and queries ... most often the change in direction specified by the hierarchy option is too violent, and the average performance of most hierarchy procedures does not appear to be sufficiently promising to warrant their immediate integration in an analysis system for automatic document retrieval "

There is no suggestion that the results of these early experiments can be extrapolated directly to cover the current efforts. However, the operations implemented at the time are very similar to what is now being suggested as an approach to modern retrieval system design. One has reason to be careful in

accepting many of the currently unevaluated design proposals for expert system approaches as effective solutions to the retrieval problem.

References

- [1] E. Defude, Different Levels of Expertise for an Expert System in Information Retrieval ACM-SIGIR Conference on Research and Development in Information Retrieval, Montreal. June 1985.
- [2] G. Guida and C. Tasso. An Expert Intermediary System for Interactive Document Retrieval, Automatica, 19:6. 1983, 759-766.
- [3] B.P. McCune, R.M. Tong, J.S. Dean and D.G. Shapiro. Rubric: A system for Rule-Based Information Retrieval, Proc. IEEE Seventh International COMP-SAC Conference. Chicago, 1983.
- [4] W.B. Croft and T. Parenty. A Comparison of a Network Structure and a Database System Used for Document Retrieval. Technical Report, Computer Science Department. University of Massachusetts, 1984.
- [5] W.B. Croft and R.H. Thompson. An Expert Assistant for a Document Retrieval System. Proc. RIAO Conference. Grenoble. France. March 1985.
- [6] A.S. Pollitt, A Front-End System: An Expert System as an On-Line Search Intermediary Aslib Proceedings, 36:5. May 1984. 229-234.
- [7] R.N. Oddy, Information Retrieval through Man-Machine Dialogue. Journal of Documentation, 33:1-14. 1977.
- [8] G. Salton and M.J. McGill. Introduction to Modern Information Retrieval. McGraw Hill Book Company. New York, 1985.
- [9] C.J. van Rijsbergen. Information Retrieval. 2nd Edition. Butterworths. London, 1979.
- [10] C.W. Cleverdon. Optimizing Convenient Access to Bibliographic Databases. Information Services and Use. Vol. 4, 1984. 37-47.
- [11] G. Salton and M.E. Lesk. Computer Evaluation of Indexing and Text Processing. Journal of the ACM, 15:1. January 1968. 8-36.

A LEXICAL FUNCTIONAL GRAMMAR SYSTEM IN PROLOG

A. Eisele

Department of Linguistics
University of Stuttgart
D-7000 Stuttgart

Abstract

This paper describes a system in PROLOG for the automatic transformation of a grammar, written in LFG formalism, into a DCG-based parser of the language. It demonstrates the main principles of the transformation, the representation of f-structures and constraints, the treatment of long-distance dependencies, and left recursion.

Finally some weak points of the system and possibilities for overcoming them are discussed.

Introduction

In order to improve our knowledge about natural language, it is desirable to have a high-level description language which can be used to test grammars on a computer, but which is independent of the details of the implementation. For linguists without knowledge of programming languages, a system for writing and testing grammars on a computer should be offered.

At the University of Stuttgart such a system has been implemented in PROLOG, which uses the formalism of lexical-functional grammar [Bresnan, Kaplan 82] as its description language.

The system makes it possible for the user to enter grammar rules and lexical entries directly in the form described by Bresnan/Kaplan. The input is translated into PROLOG rules, which form a top down parser in definite clause grammar style.

The whole system is written in PROLOG II [Caneghem 82]*, which allows for delaying subgoals by using the predicates FREEZE and DIF. These features are used to optimise the evaluation of the f-

structures. Equations and constraints associated to a grammar rule are evaluated as soon as the rule is used; and thus wrong parses can be rejected as soon as constraints are violated. One of the main problems using DCG grammars - the prohibition of using left-recursive grammar rules - is solved by automatically converting these rules to right-recursive ones in such a way that the f-structures remain the same.

Principles of the Parser

Bresnan gives an algorithm in three steps to get the f-structure associated to a sentence. Firstly the constituent structure of the sentence is discovered, using only the context-free parts of the grammar. In a second step, the equations and constraints associated to the grammar rules used are collected and the meta-variables are replaced by real variables. In a third step, the system of equations is solved and the f-structure of the input sentence is found.

Considering that the functional description can be used to reduce the syntactical ambiguities of the input sentence and that the process of solving the equations is independent of their order, parsing the sentence and solving the equations can be done in parallel.

In doing so, there must be the possibility to denote partial f-structures, i.e. f-structures which result from solving some of the equations and which are still to be extended when considering further information.

We represent partial f-structures as an 'open ended' list of pairs:

$$[A_1 = V_1 , A_2 = V_2 , \dots , A_n = V_n \mid _]$$

where the A_i stand for (atomic) attributes and the V_i for the values associated to these attributes. These values are either atomic, terms denoting semantic forms, themselves f-structures or the term 'set(S)' where S stands for an (open ended) list of f-structures denoting a set.

* Although PROLOG II differs from the standard PROLOG syntax and does not allow for writing grammar rules in DCG formalism, I use the syntactic features of CPROLOG for better readability.

An equation that expresses the identity of two f-structures can thus be evaluated by inserting into both structures the features missing with respect to the other, and then unifying the variables that stand for the rest of the lists.

A procedure which performs this action can easily be written in PROLOG :

```
del(F,[F|X],X) :- !.
del(F,[E|X],[E|Y]) :- del(F,X,Y).

merge(X,X) :- !.
merge([A=V1|R1],F2) :- del(A=V2,F2,R2),
                        merge(V1,V2),
                        merge(R1,R2).
```

Despite its name, the predicate `del` can be used to insert an element at the end of an open ended list if it is not already a member of this list. In both cases, the third argument is instantiated to a list which is identical to the second argument minus the first.

The predicate `merge` takes two f-structures as its arguments and expands them such that they become identical if they do not contain contradicting values for any attribute at any level; if they do so, the predicate fails and nothing is changed.

(The treatment of sets is omitted here for the sake of simplicity).

```
Example: The goal      merge([subj=[spec=def,
                             num=sg,
                             pred=girl |RSubj1] |RS1],
                             [pred=hand(subj,obj2,obj),
                             tense=present,
                             subj=[num=sg |RSubj2] |RS2])).
```

yields the assignments

```
RS1 = [pred=hand(subj,obj2,obj),
        tense=present|RS2]
RSubj2 = [spec=def,pred=girl|RSubj1]
```

Using only the predicate `merge` for expressing equality, all LFG rules with equations can be transformed into DCG clauses.

For example

$$S \rightarrow NP (\hat{=} \text{SUBJ})=v \quad VP \hat{=}v$$

is translated into the DCG rule

$$s(S) \rightarrow np(NP), \{ \text{merge}([\text{subj}=NP|_], S) \}, \\ vp(S).$$

Treatment of Constraining Equations

After merging two f-structures, they remain identical during the whole further computation, i.e. each extension to one of them also affects the other. This feature can be used, in connection with the coroutines of PROLOG II, to treat constraining equations in a simple and efficient way: goals which test constraints can be 'frozen' to variables that are part of an f-structure. If such a variable is instantiated later (maybe while parsing a different constituent of the input sentence) the goal is 'woken up' and if the test now fails, backtracking is invoked.

In this way all kinds of constraining conditions in LFG formalism and even the tests for completeness and coherence of the f-structures can be evaluated locally - there is no need to 'gather' them in extra variables and test them afterwards.

Specifically, we proceed as follows:

- Existential constraints are treated by inserting the attribute into the f-structure, but the associated value remains variable (if it isn't already known). To the PROLOG variable that represents this value the condition $\neq \text{nil}$ is bound using 'freeze'. After parsing, all the remaining variables in the structures are set to the value 'nil', so a missing value can be detected. (In general, the violation of an existential constraint can not be detected sooner, because the parser cannot know if it will still get a value for the attribute).
- Negative existential constraints are treated by assigning the value 'nil' to the attribute. The value 'nil' is interpreted as non-existence of the feature, so the change of the f-structure is

not an essential one. The contradicting case in which both a positive and a negative existential constraint are required, is detected immediately.

- Constraining equations are evaluated using the predicate `merge_c`, which compares the parts of its arguments that are already known and freezes the further tests to the remaining variables. As soon as a contradiction appears, backtracking is invoked.

- Negative constraining equations are handled by the predicate 'different' which uses the PROLOG II primitive 'dif'. Dif fails as soon as its arguments are unified, but waits if they still contain parts represented by different variables.

- Completeness of f-structures is tested by existential constraints on the sub-structures required by the semantic form. We think that the mere existence of a required sub-structure is not enough. For example, verb entries often introduce a partial f-structure for the subject by specifying its number. This should not lead to the acceptance of a sentence without subject. For that reason we use existential constraints on the 'pred' of a structure to test if it is there.

- Coherence of an f-structure means a negative existential constraint concerning all governable functions that are not required by its semantic form. We use the predicate 'ngf', which checks all attributes of the remaining structure as soon as they are built in and fails if they are governable functions.

Since the invocation of 'frozen' tests can interact with the procedure that merges f-structures, the real definition of 'merge' had to be slightly altered.

Definitions of some of the predicates used are given in the appendix.

Example: the lexical entry
promised:

```
V, (^ TENSE) = PAST
    (^ PRED) = 'PROMISE <(^ SUBJ)(^ OBJ)(^ VCOMP)>'
    (^ VCOMP TO) = c +
    (^ VCOMP SUBJ) = (^ SUBJ)
```


becomes

```
v(V) --> [promised],
      (merge(V, [tense=past,
                pred=promise(subj,obj,vcomp),
                subj=[pred=PSUBJ|RSUBJ],
                obj=[pred=POBJ|_],
                vcomp=[to=PLUS,
                      pred=PVCOMP,
                      subj=[pred=PSUBJ|RSUBJ]|_]
                |R2]),
      freeze(PLUS,PLUS=+),
      freeze(PSUBJ,PSUBJ\=nil),
      freeze(POBJ,POBJ\=nil),
      freeze(PVCOMP,PVCOMP\=nil),
      ngf(R2) ).
```

where the first freeze subgoal ensures that, whenever PLUS gets instantiated, it must be set to +, the other freeze subgoals ensure that the preds of the structures demanded by the semantic form must not be set to nil (completeness) and the ngf constrains the further expansions of the f-structure to attributes that are not governable functions (coherence).

Treatment of Long-Distance Dependencies

To handle long-distance dependencies correctly, the conditions for proper instantiation given by Bresnan/Kaplan have to be satisfied. They concern:

- The relation between the domain roots and the long distance controllers.
- The one-to-one assignment between domain root and controllee
- The observance of the crossing limit

Also, bounding nodes, i.e. nodes that are excluded from the control domains of higher nodes, have to be handled correctly.

One of the tasks, the identification of the domain roots, only depends on the grammar and is performed during the transformation of the grammar rules into PROLOG clauses.

The main job, the assignment between domain root and controllee is performed as follows:

Each goal has two extra parameters for input and output of a controller list. These lists act as a global stack on which the controllers are pushed. Each element of the stack refers to a node which dominates the current goal, and which is a domain root.

A domain root adds an element to the stack before the parser enters its control domain and removes one after the domain is left. The element that is pushed consists of the class name (eg. [+wh]) of the controller and its actual variable.

If a controllee appears, the stack is searched for the first element with the same class name (for crossing limit n the first n+1 matching elements can be chosen). Now the controllee can use the actual variable of the controller.

To check the correct assignment between domain root and controllee the stack element is replaced by a 'receipt'. The domain root tests this when it removes the stack element.

As an example, the transformation of an LFG rule with controller and of an entry for a controllee are given:

```
NP -> eps      ^ = ( <= np).
```

```
np(CLO,CL1,Fnp) -> [],
                    {subst(np/Fnp,np,CLO,CL1)}.
```

```
S' ->      NP (^ Q) = (NP => wh)
          (^ FOCUS) = v
          v = (S => np)
bounding S  ^ = v.
```

```
s_bar(CLO,CL1,Fs_bar) -> np([wh/Q|CLO],[wh|CL1],Fnp),
                           {merge([q=Q,focus=Fnp|_],Fs_bar)},
                           s([np/Fnp],[np],Fs_bar).
```

Treatment of Left Recursion

Definite Clause Grammars do not allow left-recursive grammar rules to be used because they lead to loops in the top-down parser.

This is a serious shortcoming for a natural language system since many linguistic phenomena can be most naturally described with left-recursive rules (coordination, possessive NPs etc.).

In the theory of formal languages there exist several algorithms to convert a grammar containing left recursion into an equivalent grammar that does not.

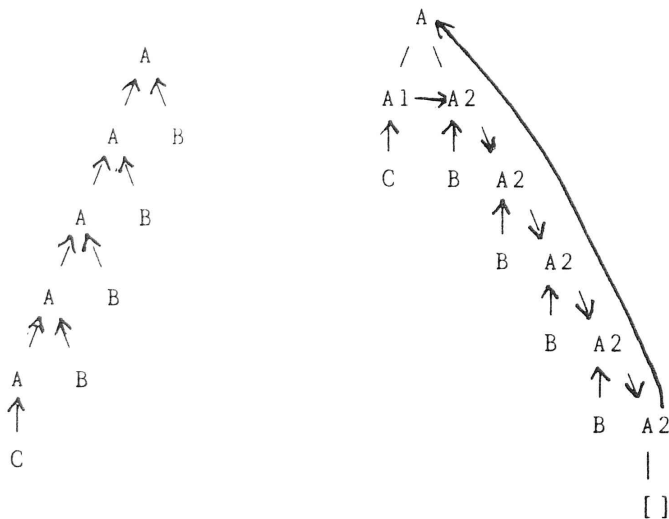
But equivalency means here that the same language is generated with different assignment of c-structures to the input sentences. In LFG the c-structures are essential for the correct evaluation of f-structures, so a transformation must provide a way to get the right interpretation of the functional description.

At the moment, our system performs an automatic conversion of locally left-recursive rules into rules which are right-recursive and which produce the same f-structures.

The method is best shown with an example:

The rules $A \rightarrow A B.$ $A \rightarrow C.$
 can be transformed to the equivalent grammar
 $A \rightarrow A_1 A_2.$ $A_2 \rightarrow B A_2.$ $A_1 \rightarrow C.$
 $A_2 \rightarrow [].$

This transformation can be performed for every local left recursive symbol in a one-to-one manner, where only the two extra rules on the left are needed. But now the c-structure of an input string C B B B B is completely different:



In the old grammar one could say that each A node takes the f-structure(s) of its subordinate A and B or C node(s) and uses them to build its own f-structure.

In the new grammar, the A1 node behaves just as the A node dominating the C node in the old grammar, but the A2 nodes can be considered as taking the f-structure of the left-sister A1 node or the **dominating** A2 node and that of the subordinate B node to build its own f-structure.

If the f-structures are evaluated this way and passed downward, then the f-structure belonging to the A node will be assigned to the A2 node which expands to the empty string.

We have to provide an extra parameter for passing the f-structure of A through all the A2 goals.

The following example shows the above grammar augmented with variables for f-structures and with goals which evaluate these f-structures:

```
a(FH) -> a(FA), {goal_a(FH,FA)},
          b(FB), {goal_b(FH,FB)}.
```

```
a(FH) -> c(FC), {goal_c(FH,FC)}.
```

This would be converted to the right-recursive DCG program that yields the intended f-structures:

```
a(FH) -> a1(F1), a2(F1, FH).
```

```
a2(FA, F) -> [], {merge(FA, F)}.
```

```
a2(FA, F) ->          {goal_a(FH, FA)},
                   b(FB), {goal_b(FH, FB)},
                   a2(FH, F).
```

```
a1(FH) -> c(FC), {goal_c(FH, FC)}.
```

Experience with the System

The LFG system, which has been written by Jochen Doerre and myself, using ideas from Werner Frey and Uwe Reyle, runs on a VAX 780 and has been tested with different grammars of English, German and French, the latter consisting of about 30 grammar rules and over 200 lexical entries.

Our expectations concerning the parse times were not quite fulfilled.

We see mainly three reasons for long parse times:

- The language PROLOG II, which was originally designed to run on an Apple and was afterwards adapted to a VAX, does not seem appropriate for high-efficiency language analysis.

- The grammars contain many optional constituents and constituents that can be repeated arbitrarily often, which are handled automatically by introducing new nonterminal symbols and empty productions for them. This leads to a high amount of backtracking in the top-down parser, which does not regard the input string before it has expanded its expectation down to the lexical level.

WOZU UND ZU WELCHEM ENDE MACHEN WIR SPRACHORIENTIERTE KI

H.-D. Lutz

Erziehungswissenschaftliche Hochschule
Rheinland-Pfalz
Rheinau 3-4
D-5400 Koblenz

Zehn Jahre ist es her, daß J.Siekman in AISB European Newsletter einen kleinen Aufsatz veröffentlichte mit dem Titel: "German Intelligence Becomes Artificial" und dem Unterton, man hätte in der BRD ja lange nicht die Zeichen der Zeit erkannt, sei nun aber aufgebrochen, den scientific gap zu schließen. Ebenfalls 1975 fand sich hierzulande die Gruppe zusammen, die heute unter der offiziellen Bezeichnung "Fachausschuß 1.2. (KI und Mustererkennung) in der GI" firmiert, aber besser bekannt ist als KI-Gruppe. Man sieht, wir haben organisatorisch aufgeholt.

Was nun die institutionelle und inhaltliche Etablierung an den Hochschulen und den kontinuierlichen und breiten Transfer von Forschungsproblemen und -resultaten in die Lehre angeht, so habe ich allerdings meine Zweifel, daß wir bereits genauso weit sind.

"Ansätze der sprachorientierten KI-Forschung" (der Kürze halber im folgenden: SOKI) soll das nächste Themenheft des GLDV-Forums heißen. Ich finde diese Formulierung angemessen. Denn Ansätze können "vielversprechend" sein, können sich aber auch als "steckengeblieben" erweisen, können in die "richtige" Richtung gehen oder in eine "falsche" Richtung weiterentwickelt werden zu "guten", aber auch zu "schlechten" Zwecken.

Wenn man die allgemeine KI-Euphorie mit den KI-Realitäten kontrastiert, so habe ich meine Zweifel, ob die hochgesteckten Erwartungen, die in die KI-Forschung gesetzt werden, in Erfüllung gehen (können).

Keinen Zweifel kann es aber für mich geben, daß KI-Ansätze bzw. -entwicklungen nicht in die Richtung gehen dürfen, die gerade von militärischer Seite intendiert wird. Daß davon nicht nur immer das jeweils andere KI-Teilgebiet, sondern inzwischen jedes KI-Teilgebiet, also auch die SOKI betroffen ist, soll das folgende Zitat belegen:

"Um die wissenschaftliche und technologische Basis einer breiten Anwendung der KI im militärischen Bereich zu schaffen, wurde von DARPA unter dem Titel 'Strategic Computing Initiative' (SCI) eine Initiative zur Forschung und Entwicklung im Bereich der KI mit 10jähriger Laufzeit eingeleitet. Das 80-seitige Dokument, in dem die Ziele dieser Initiative beschrieben werden, wurde am 28. Oktober 1983 veröffentlicht. Die im Rahmen des Programms zu entwickelnden 'intelligenten' Computersysteme wollen über "beispiellose Fähigkeiten verfügen und versprechen einen enormen Zuwachs für unsere nationale Sicherheit und ökonomische Macht."

Die Ziele der SCI werden von Robert S.Cooper, dem Direktor von DARPA, und seinem Stellvertreter Robert E.Kahn wie folgt beschrieben: Die SC-Initiative "wird eine neue Generation superintelligenter Computer für militärischen Gebrauch hervorbringen. Diese Maschinen werden menschenähnliche Fähigkeiten haben, die ihnen gestatten, zu sehen, zu denken, zu planen und die Aktionen von militärischen Systemen auf dem Gefechtsfeld zu überwachen."

Für die ersten fünf Jahre wurden Gesamtkosten von rund 600 Millionen US-Dollar vorgesehen; bis zum Ende des Jahrzehnts sollen für SCI insgesamt eine Milliarde US-Dollar aufgewendet werden.

In dem Dokument wird für jede der drei Teilstreitkräfte ein Projekt zur militärischen Anwendung der KI ausgeschrieben. Diese drei Forschungs- und Entwicklungsprojekte wurden ausgewählt, weil von ihrer Nutzung ein großer Zuwachs der 'Verteidigungskraft' angesichts der Szenarien künftiger Schlachten erhofft und darüber hinaus "in ihnen die Entwicklung der Technologie sozusagen gebündelt und angeregt wird und sie schließlich ein Fertiglabor darstellen, mit dem nachgewie-

sen werden kann, ob die Technologie in der Tat funktioniert". Indem diese drei konkreten Zielvorgaben umgesetzt werden, soll die technologische Basis einer völlig neuen Generation 'intelligenter' Computersysteme geschaffen werden. Als Anwendung für das US-Heer ist die Entwicklung eines fahrerlosen Landfahrzeugs zur Aufklärung ausgeschrieben, das mittels visueller Sensoren und seines Bordrechners in der Lage ist, sich selbst zu einer vorgegebenen Position zu steuern, feindliche Kräfte zu erkennen und Informationen zu übermitteln. Ein mit einem Echtzeit-Expertensystem zusammenarbeitendes Bildverarbeitungssystem soll Straßen und Begrenzungen erkennen, Objekte ausmachen und identifizieren und eine Karte des Geländes anlegen, während sich das Fahrzeug mit einer Geschwindigkeit von bis zu 60 km/h bewegt. Um diese Aufgaben in der geforderten Zeit erfüllen zu können, muß der Bordcomputer 10 bis 100 Milliarden (von Neumann-)Instruktionen pro Sekunde verarbeiten. (Die leistungsfähigsten heutigen Rechner verfügen über eine Rechengeschwindigkeit von 30 bis 40 Millionen Instruktionen pro Sekunde). Darüber hinaus soll der Bordrechner weniger als 250 kg wiegen und weniger als ein Kilowatt elektrischer Leitung verbrauchen. Das Navigations-Expertensystem soll mit einer Wissensbasis von 6500 Regeln ausgestattet sein und 7000 Regeln pro Sekunde verarbeiten können. (Gegenwärtige Expertensysteme verfügen über bis zu 2000 Regeln bei einer Verarbeitungsgeschwindigkeit von 50 bis 100 Regeln pro Sekunde). Der Auftrag zur Entwicklung eines Prototyps für ein solches autonomes Landfahrzeug ging im Herbst 1984 an die Firma Martin Marietta. Die zweite, für die US-Luftwaffe vorgesehene Anwendung ist ein automatischer 'Pilotenassistent'. Damit soll das Problem gelöst werden, daß der Pilot physisch damit überfordert ist, die Fülle der Informationen, die sich aus der komplizierten Technik des Kampfflugzeugs und aus der komplexen Gefechtsituation ergeben, schnell zu verarbeiten. Der automatische 'Assistent' soll den Piloten bei der Handhabung und Überwachung der zahlreichen Fluß- und Waffenkontrollsysteme unterstützen, so daß sich der Pilot auf die taktischen und strategischen Ziele konzentrieren kann. Die natürlichsprachliche Schnittstelle zwischen dem Pilot und dem automatischen 'Assistent' beruht auf Techniken der Spracherkennung und -erzeugung sowie der graphischen Repräsentation komplizierter Informationen. Die Wissensbasen des zu entwickelnden Expertensystems werden "signifikant größer sein als irgendeine der bisher projektierten. Allein die Überwachung der grundlegenden Flugsteuerungssysteme (Antrieb, Elektrik und Hydraulik) erfordert beispielsweise mehrere tausend Regeln."

"Hier finden sich erneut die Ziele von Miniaturisierung und Robustheit der Hardware, aber auch Ziele, mit komplexen, integrierten, wissensgestützten Systemen zu verarbeiten, die hundertmal schneller sein müssen als heutige Systeme, und Ziele für unmißverständliche Sprachkommunikation bei besonders großem Lärm."

Auch die Aufgabenstellung des für die US-Marine geplanten seegestützten Kampfführungssystems für Flugzeugträger - dem dritten Projekt im Rahmen der SCI - besteht darin, die unermesslichen Datenmengen, durch die künftige Schlachten gekennzeichnet sein werden, zu bewältigen und die Kommandeure dabei zu unterstützen, zu einer Einschätzung der Gefechtslage zu kommen. Gestützt auf eine sehr große Informationsbasis über militärische Strategie und auf aktuelle Gefechtsdaten, soll das System mögliche Maßnahmen vorschlagen und deren mußmaßliche Ergebnisse vorhersagen; dann für die von dem Kommandeur gewählte Option detaillierte Aktionspläne entwickeln, diese Pläne an die Ausführenden weitergeben und die Befehlshaber Fortschritte der Ausführung unterrichten. Diese Informationen sollen in natürlicher Sprache unterstützt durch Graphikern ausgegeben werden.

Die Bewältigung des Funktionsablaufs erfordert nach den Angaben von DARPA ein 'verteiltes Expertensystem' mit etwa 20000 Regeln und einer Verarbeitungsgeschwindigkeit von 10 Milliarden Instruktionen pro Sekunde. "Allein das natürlich-sprachliche System erfordert eine Verarbeitungsgeschwindigkeit von ungefähr einer Milliarde Instruktionen pro Sekunde."

Die im Rahmen dieser konkreten Projekte erzielten Fortschritte im Bereich der KI sollen über diese spezifischen Anwendungen hinaus die Möglichkeit zum Einsatz 'intelligenter' Computer in militärischen Systemen auf breiter Front eröffnen. So sollen etwa Bildverarbeitungs- und Expertensysteme, wie sie im Rahmen des Landfahrzeug-Projekts entwickelt werden sollen, auch in anderen auto-

nomen Systemen, wie Marschflugkörpern oder Untersee-Robotern, Verwendung finden. Die KI-Anwendung für die Marine soll beispielsweise im Bereich des Heeres als taktische Kampfführungssysteme auf Korps-, Divisions- und Bataillonsebene eingesetzt werden.

Das wichtigste Ziel von DARPA dürfte jedoch, so die amerikanischen Computerwissenschaftler Willie Schatz und John W. Verity, die Entwicklung für das 'Star Wars'-Konzept benötigten Computertechnologie sein...."

(Wilhelm, Rudolf: Strategic Computing Initiative: Ein kurzer Überblick.
In: Bickenbach, J.; Keil-Slawik, R; Löwe, M.; Wilhelm R. (Hg.), Militarisierte Informatik. Marburg 1985 (=Schriftenreihe Wissenschaft und Frieden. Nr.4: hrsg. v. Bund demokratischer Wissenschaftler. Forum Informatiker für Frieden und gesellschaftliche Verantwortung. Regionalgruppe Berlin, und Forum Naturwissenschaftler für Frieden und Abrüstung), 75-79, hier: 75-77)

Ist das folgende (britische under-) statement ein Gegenargument?

"The millennial language processing system will be expectation-driven in a very rich sense, and will need to "know what's going on" ahead of time in a pretty profound way. How we move forward from our current almost total incapacity in this area I don't know."

(Thompson, H., Natural language processing: a critical analysis of the structure of the field, with some implications for parsing.

In: Sparck Jones, K.; Wilks, Y. (Eds.), Automatic Natural Language Parsing. New York etc. 1983, 22-31, hier: 28)

Und läßt sich die Quintessenz des folgenden Zitats nun wieder als Gegenargument zum Gegenargument ins Feld führen?

"Wahlster betonte nochmals seine Zielgröße "Linguistik (als Knowledge Engineering) für die AI." Er betonte den großen Aufwand, der in der AI-Forschung zu betreiben ist (Implementierung). Es fiel das Stichwort dieser Diskussion: Fleiß. 80 % sei Fleißarbeit. Sehr wenige deutsche Grammatiken lägen vor." (Aus: Panel "Linguistik und Künstliche Intelligenz" (Zusammenfassung: Th.T.Ballmer). In: Laubsch, J. (Hrsg.), GWAI-84. 8th German Workshop on Artificial Intelligence, Wingst/Stade, October 8-12, 1984. Berlin etc. 1985, 265-282, hier: 281)

Meine Befürchtung ist allerdings, daß es schon nicht mehr um eine inhaltliche Diskussion darum geht, ob etwas "machbar" ist. Schon garnicht darum, wie es um die Qualität des "Produkts" bestellt ist (Wahlspruch: "it works!").

Es geht vielmehr um nichts weniger als um eine nicht-triviale Antwort auf die Frage: Wozu und zu welchem Ende machen wir SOKI?

Natürlich müssen wir die Arbeit der letzten zehn Jahre kritisch sichten mit dem Ziel, all das zusammenzustellen, was wir können, aber auch, was wir nicht können.

Natürlich müssen wir z.B. im Rahmen von 'Synthese natürlichsprachlicher Äußerungen' genau untersuchen, was alles nötig ist, um einen morphologisch und syntaktisch korrekten output zu erhalten, der darüberhinaus auch Kriterien der kommunikativen Adäquatheit genügt (und dazu braucht's eben eine ganze Menge an epistemologischem Wissen, an Inferenzregeln, an Kenntnissen über den bisherigen Dialogverlauf, über den aktuellen Aufgabenkontext, dynamischer Partnermodellierung, usf.).

Natürlich werden wir zusammenstellen müssen, was es mit "neuen" Anläufen auf sich hat: GPSG, LFG, DCG u.ä. einerseits und nicht-klassische Logiken andererseits, u.v.a.

Vielleicht bekommen wir auf diese Weise zumindest eine Antwort auf die Frage, warum in dem Artikel "AI - Where Are We" von BOBROW und HAYES (eds.) in ARTIFICIAL INTELLIGENCE 25 (1985), 375-415 von SOKI praktisch nicht die Rede war.

Aber all dies wird nur eine Teilantwort auf die zentrale Frage sein.

Um die Frage vollständig beantworten zu können, werden wir den "holistischen Ansatz", den die KI ja für sich reklamiert, viel weiter auslegen müssen.

SECHS THESEN ZUR SPRACHORIENTIERTEN KI-FORSCHUNG

A. Fauser

TA Triumph Adler
D-8510 Fürth

These 1:

KI ist für die sprachorientierte Forschung (und, bitte schön, auch für die Entwicklung von Systemen) ein wichtiges Hilfsmittel. Ihren vollen Wert gewinnt sie erst in der konzertierten Aktion: mit Linguistik und Informationswissenschaft. Die Schönheit der Repräsentationssprache kann nicht die Unzulänglichkeit der Wissensbasis verschleiern. Und fachliches Wissen, das nicht mit guten Hilfsmitteln implementiert werden kann, ist für die LDV genau das Papier wert, auf dem es eben nur gedruckt werden kann.

These 2:

Natürlichsprachliche Dialogsysteme dürfen entweder nur sehr wenige oder müssen sehr viele sprachliche Phänomene beherrschen, wenn sie akzeptabel sein wollen. Im ersten Fall lernt der Benutzer, was das System kann - und hat damit eine (hoffentlich) verhältnismäßig komfortable Kommandosprache. Im anderen Fall lernt er, was das System nicht kann - der Lernaufwand ist auch hier quantitativ gering. Doch ein Problem bleibt: wie lehre ich dem häufig bemühten "Gelegenheitsbenutzer", daß mein System leider keine "cleft sentences" beherrscht und auch beim Bach-Peters-Paradox versagt. Sind die Alternativen Frust durch "Trial and Error" oder ein Linguistik-Studium?

These 3:

Der Vorteil natürlichsprachlicher Dialogsysteme kommt besonders dann zum Tragen, wenn die Eingaben lang und/oder komplex sind. Sonst sind Menütechnik, Funktionstasten, direkte Manipulation oder Kombinationen davon häufig ebenbürtig. Statt lange zu tippen, wird man lieber sprechen wollen - so ist man's im Umgang mit Sprache auch gewohnt. Ergo: die Kopplung oder Integration mit akustischer Spracheingabe wird, die Erkennung fließend gesprochener Sprache vorausgesetzt, erst den vollen Erfolg bringen.

These 4:

Systeme zur Indexierung oder Deskriptorengewinnung werden - bei allen Fortschritten - in nächster Zeit für den praktischen Einsatz am besten wie MÜ-Systeme genutzt: sie liefern einen guten Vorschlag, der aber meist noch überarbeitet werden muß.

These 5:

Die Deskriptorengewinnung ist eine Vorstufe für den Retrievalprozeß. Sie muß deshalb, wenn sie benutzerfreundlich sein will, die Vorlieben der (potentiell) später Suchenden wenn

möglich einbeziehen. Das heißt zum Beispiel konkret: wenn ich bestimmte Lieblingsstichworte habe, sollten sie möglichst in der Klassifikation berücksichtigt werden, denn mit diesen werde ich das Dokument vermutlich zuerst suchen.

These 6:

Der Erfolg des Einsatzes von KI-Techniken bei der Dokumentenklassifizierung hängt nicht von der Zahl und Menge der Dokumente ab, sondern davon, ob sie zu einem genügend kleinen und abgrenzbaren Aufgabengebiet gehören. Analog zur Entwicklung von Expertensystemen gilt es, solche zu finden und auf ihnen Erfahrungen mit wissensbasiertem Filing und Retrieval zu sammeln. Dieser Weg ist eine Gratwanderung: auf der einen Seite der Problemsumpf des "Textverstehens", auf der anderen terminologische Spiegelfechtereie ("Mein Thesaurus ist meine Wissensbasis"). Aber mancher Gipfel ist eben am besten über einen Grat zu besteigen.

THESEN ZUR SPRACHORIENTIERTEN KI-FORSCHUNG BZW. ZUR KI-ORIENTIERTEN SPRACHFORSCHUNG

P. Schefe

Universität Hamburg
D-2000 Hamburg

F ü r :

- 1) Für eine informatisch orientierte KI ist die Befassung mit der automatischen Behandlung natürlicher Sprache sehr nützlich, wenn nicht unabdingbar, da sie ein wesentliches Korrektiv für epistemologisch inadäquate Modellierung von Wissen darstellt.
- 2) Für die Linguistik ist sie Hilfsmittel und wesentliche Ergänzung, da Formalisierungen leichter überprüfbar bzw. in ihren (formalen) Konsequenzen überschaubar werden. Der algorithmische Aspekt der Sprache kommt überhaupt erst in den Blick. Neben Kompetenzmodellen lassen sich auch Performanzmodelle (kognitive Simulation) explorieren.

G e g e n :

- 1) Die Befassung einer informatisch orientierten KI mit natürlicher Sprache führt zu der Annahme, man könne durch sogenannte natürlich-sprachliche Schnittstellen größere Benutzerfreundlichkeit erreichen. Das Gegenteil ist der Fall.
- 2) Die Einführung von KI-Methoden in die Linguistik grenzt wesentliche Aspekte der Sprache aus. Die Sprache als historisches Phänomen, die Entstehung und der Wandel von Bedeutungen im sozialen Kontext z.B. gerät aus dem Blick. Es droht die Spaltung in eine an Verwertungsinteressen orientierte, förderungswürdige Sprachforschung und eine förderungsfreie Diskussionswissenschaft.

THESEN ZUR SPRACHORIENTIERTEN KI

U. Hahn

Universität Konstanz
 Informationswissenschaft
 Projekt TOPIC II
 Postfach 5560
 D-7750 Konstanz 1

Einschätzungen zur sprachorientierten KI werden vorzugsweise aus der Perspektive von Erfahrungen formuliert, die aus Arbeiten am Forschungsgegenstand "gesprochene Sprache" oder "interaktive Dialogformen" herrühren. Die folgenden Thesen rücken "schriftsprachliche Texte" als eine dritte Anwendungsklasse sprachorientierter KI in den Vordergrund, ohne daß damit eine ausschließliche Konzentration auf textorientierte KI verbunden ist. Zusätzlich wirkt auf diese Aussagen der Hintergrund einer informationswissenschaftlichen Arbeitsumgebung ein, in der sprachanalytische Konzepte primär unter einer funktionalen Perspektive Eingang finden, inwieweit sie nämlich die Lösung natürlichsprachlicher Probleme bei der Entwicklung von Informationssystemen unterstützen (informationelle Sprachanalyse).

Die in dieser Wissenschaftsumgebung formulierten Forschungsfragen im Bereich der automatischen Textanalyse waren lange an einer Suboptimierung bestehender Informationssystemstandards - referenzorientierten Dienste, wie Indexing und Klassifikation - ausgerichtet. Letztlich wurden die mit den verfügbaren lexikalisch-statistischen und strukturell grammatikalischen Methoden dabei erarbeiteten Verfahren weder aus dem Rahmen kontrollierter Experimentalumgebungen herausgelöst noch - und das scheint weitaus gravierender - fruchtbare und signifikante Anschlußprobleme zur begründeten Fortsetzung dieser Forschungslinie aus ihnen abgeleitet. Auch die lange Stagnation im Bereich des automatischen Übersetzens - als Beispiel einer anspruchsvolleren Form des Textverstehens in rein linguistischer Zuständigkeit - steht exemplarisch für ein tiefgreifendes sprachanalytisches Methodendezifit, dessen Kern offen-

sichtlich darin besteht, lexikon- bzw. syntax-lastige Grammatikspezifikationen als dominierenden Träger von Sprachbeschreibungsmodellen zu begreifen.

T_1 Das Paradigma der sprachorientierten KI ist für die (informationelle) Sprachanalyse zweifach attraktiv:

- Es fördert die Definition von Forschungsthemen zu anspruchsvolleren Formen des Textverstehens (und damit insbesondere eine substantielle Erweiterung des programmatischen Rahmens im (textorientierten) Information Retrieval)
- Es scheint mit seiner Methodologie und seinem methodischen Instrumentarium auch die für die Lösung entsprechender definierter Forschungsprobleme und Durchführbarkeit von solchen Projekten notwendigen Konzepte und Techniken bereitzustellen bzw. deren Entwurf und Realisierung zu unterstützen (im Unterschied zu statistisch bzw. rein grammatikalisch ausgerichteten linguistischen Verfahren)

Dabei ist der Definitionsprozeß längst von der Durchführungsphase abgelöst worden. Sie zeigt für die textorientierte KI im Moment eine große Vielfalt konkurrierender Forschungsthemen. So zählen analytisch betrachtet zu den momentan etablierten (anwendungsbezogenen) Forschungsproblemen im Bereich von Textverstehenssystemen:

- die Synthese von Wissensbasen auf der Basis des Verstehens von Eingabetexten (Faktenextraktion)
- die Generierung von Texten aus Repräsentationsstrukturen in Wissensbasen
- das Übersetzen von Texten
- intelligente Formen des textorientierten Information Retrieval (Fakten, Textextrakte, inhaltliche Synopsen) auf der Basis interaktiver natürlichsprachlicher oder grafischer Zugriffsmodi

Neue, wesentlich von der Verfügbarkeit elektronischer Texterstellung- und Textverwaltungssysteme geforderte Themen betreffen:

- intelligente Formen des Mailing/-Messaging (etwa die durch Benutzerprofile gesteuerte Textverteilung im Rahmen wissensbasierter Büro-Informationssysteme)
- intelligente Editoren, die neben trivialen Funktionen (etwa Schreibfehlerkorrektur) zusätzlich über entwickelte Möglichkeiten zur Syntax-, Semantik- und Stilberatung verfügen

Die in T₁ ausgedrückte hohe Attraktivität des paradigmatischen Rahmens der KI bietet zweifellos eine fruchtbare methodische Hülle für die Textverstehensforschung. Das ändert aber nichts an der nach wie vor chronischen konstruktiven und empirischen Unterspezifikation ihres zentralen Forschungsgegenstands, da die in anderen Sektoren der Sprachverstehensforschung (gesprochene Sprache, Dialogsituationen) erarbeiteten Ergebnisse (Grammatiken/Parser, Systemarchitekturen usw.) keineswegs unbezogen auf den Bereich des Textverstehens übertragen werden können:

T₂ Den elaborierten Konstrukten zur Modellierung von Weltwissen - einer der wesentlichen Punkte für die Attraktivität des KI-Paradigmas - und den linguistischen Modellen aus anderen Anwendungsklassen der sprachorientierten KI stehen in der textorientierten KI keine vergleichbaren methodischen und empirischen Beiträge gegenüber, die dem eigentlichen Gegenstand "schriftsprachlicher Text" angepaßt sind.

Die textorientierte KI hat die intensiv erforschten Vertextungsmechanismen von Dialogen (unterschiedliche Formen von Anaphora, Ellipsis usw.) übernommen, ist aber über diese rudimentäre Stufe einfacher Textkohäsion kaum hinausgekommen. Zwar wird mit einer Reihe texttypischer Konstrukte experimentiert, ohne daß sich jedoch eine theoretische Vereinheitlichung abzeichnet:

- die Übernahme des in der Kognitionswissenschaft entwickelten Konzepts inhaltlich unterschiedlicher Detaillierungsebenen von Textrepräsentationen (Mikro-/Makropropositionen), mit dem kurz- und langfristige Erinnerungs- (Recall) und Zusammenfassungsleistungen rekonstruiert werden können
- das aus der strukturellen (nicht-generativen!) Linguistik entlehnte Makrostrukturenmodell von Texten (story grammars), in dem globale, textsortenspezifische Organisations-schemata von Texten postuliert werden (eine Schema-Orientierung von Textverstehensprozessen legen auch entsprechend geprägte Repräsentationskonstrukte für das Weltwissen (Scripts, Frames) nahe)
- das als Folge der Prozeßorientierung der sprachorientierten KI eingeführte Abarbeitungsprinzip des variabel tiefen (interessen-, aufmerksamkeits- usw.) abhängigen Textverstehens, mit dem die Verarbeitungskomplexität von (realistischen) Text(meng)en heuristisch gezielt reduziert werden kann
- die über alternative Inferenzregel- oder Belief-Systeme steuerbare unterschiedliche Perspektivität des Textverstehens (entsprechend der These, daß es nicht ein kanonisches Verständnis eines Text gibt)
- die Sensitivität des Textverstehens gegenüber dem Neuigkeitswert von in Texten enthaltenen Informationen und ihrer Verträglichkeit mit bestehenden Strukturen in Textwissensbasen (active memory)
- textverstehenstypische Organisationsformen der Speicherung von Textrepräsentationsstrukturen (TAUs, MOPs) zur Unterstützung inhaltsorientierter Zugriffsverfahren (conceptual indexing)

T_3 Die Forschungsdefizite der textorientierten KI konzentrieren sich auf die Bereiche:

- (a) qualitative Textdimensionierung
- (b) Textpragmatik: Textsortenvielfalt
- (c) quantitative Textdimensionierung

(a): Textorientierte KI bevorzugt für Textverstehensexperimente die Verwendung von Minimaltexten (Folgen von wenigen Sätzen), deren textstrukturelle Komplexität bewußt klein gehalten wird bzw. auf einen interessierenden Simulationsfaktor (Verknüppungsprinzipien wie Kausalität, Tempusvarianz, Erzähl-/Verstehensperspektive usf.) analytisch reduziert wird. Mit der Abkehr von Spielwelten auf der Ebene der Wissensrepräsentation muß für die textorientierte KI auch die Hinwendung zu "realistischen" Textausprägungen verbunden sein - und dies sind mit der zunehmenden Verfügbarkeit von elektronischen Textproduktions-, -distributions- und -verwaltungssystemen routinemäßig erstellte Lang- oder Volltexte (Briefe, Protokolle, Zeitschriftenartikel u.ä.). Für diesen Texttyp fehlt es aber in der textorientierten KI - im auffälligen Gegensatz zur Elaboriertheit der Weltwissensrepräsentationskonstrukte - an geeignet ausdifferenzierten Text-(wissens)modellen, deren theoretische Fixierung in Form von Textgrammatiken betrieben werden müßte. Zu deren zentralen Konzepten wird die Beschreibung der lokalen Verkettung von einzelnen Aussagen (Textkohäsion) und der globalen Wohlgeformtheit von Aussagekomplexen (Textkohärenz) gehören. Die Nichtbeachtung dieser Phänomene führt systematisch im Fall

- der Textkohäsion zur Invaliddität
 - der Textkohärenz zur Unterstrukturierung
- d.h. empirisch unkorrekten und strukturell inadäquaten, defizitären Textwissensbasen.

(b): Das Vorherrschen von human interest/factors-Texten - vorwiegend triviale Narrativtexte (kurzen literarischen Gebrauchstexten) - hat die pragmatischen Modellierungsschwerpunkte der textorientierten KI stark in den Bereich personenbezogener, emotionaler Konstrukte, sog. 'human character traits + dispositions' verzerrt, so daß ein erhebliches textsortentypisches Modellierungsdefizit für informativ Texte, Überzeugungstexte usw. (analog etwa den Besonderheiten von problemlösungs-, überzeugungs- oder beratungsorientierten Dialogen) besteht. Insbesondere resultiert aus diesen individualpsychologisch ausgerichteten Simulationsexperimenten zu emotionalen Faktoren des Textverstehens eine breite konzeptionelle Lücke zu Anforderungen textorientierter Informationssysteme, die mit höher aggregierten Konstrukten von in Benutzerklassen "generalisierten" Benutzertypen und daraus ableitbaren Benutzermodellen (Laie, casual user, Experte usf.) operieren und auf deren Forschungsagenda dementsprechende gruppen- und klassenorientierte Benutzerkonstrukte zu informativen, persuasiven usw. Faktoren des Textverstehens und textorientierten Information Retrieval stehen.

(c): Als informationswissenschaftliches Desiderat für die textorientierte KI drängt sich die Verarbeitung realistisch dimensionierter Textmengen oberhalb der quantitativen Ebene von Spielkollektionen (Analyse weniger, handverlesener Texte) auf. Daraus resultieren (besonders für die Textsorte "informativer

Text") einschlägige Forschungsfragen wie:

- semantische Konsistenz von Textwissensbasen, also die Wahrung der inhaltlichen Widerspruchsfreiheit der Textrepräsentationsstrukturen im Verlauf des Textanalyseprozesses
- Zeit-Konsistenz von Textwissensbasen, also die Problematik der Konsistenzhaltung von Wissensbasen durch die zeitliche Abfolge aufeinander bezogener Texte (dies umfaßt die Zeitdimension in Texten, zur Produktionszeit - dem Zeitpunkt der Verfassung eines Texts - und zur Systemanalysezeit! (ähnlich dem Zeitversionen-Problem im Datenbankkontext))
- modellmäßige Beherrschbarkeit von 'Very Large Knowledge Bases': Wissensverwaltungsaspekte von großen Mengen von Textwissensstrukturen, etwa Kapselungsprinzipien (Clustering) von Wissensbanksegmenten, d.h. inhaltlich ausgerichtete Fokussierungsmechanismen zur Auswahl von und zum Zugriff auf relevante Wissensbanksegmente
- Wissensakquisitionsroutinen für den manuellen bzw. halb-automatischen Aufbau und die Pflege von Weltwissensbeständen "mittlerer" Reichweite - in Analogie zu Theorien "mittlerer Reichweite" (HOMANS) -, wie sie für die informationelle Analyse breiter domänenspezifischer Textkollektionen benötigt werden, durch intellektuelles Knowledge Engineering
- Wissensakquisitionsroutinen für den automatischen Aufbau und die Pflege von Weltwissensbeständen "mittlerer" Reichweite durch maschinelles Lernen, das durch Verfahren automatischer Textanalyse operationalisiert ist

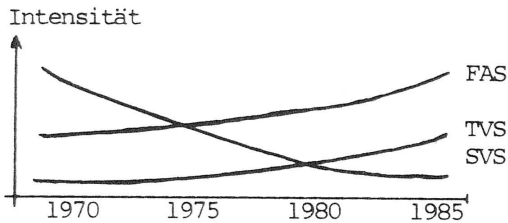
Infrastrukturell wird die Lösung dieser Forschungsprobleme und die

Einbindung von entsprechenden textorientierten Information-Retrieval- und Wissensakquisitions-Front-ends in Experten- oder Frage-Antwort-Systeme einen wesentlichen methodischen Beitrag zur Beherrschbarkeit und Operabilität immer komplexer werdender intelligenter Informationssysteme leisten, da deren Quantitätsprobleme, besonders in den Bereichen Wissensakquisition, -repräsentation und -verwaltung, durch die Migration in Produktionsumgebungen zunehmend wahrnehmbarer in Qualitätsprobleme umschlagen.

Die Forschungsthemen für anspruchsvolle Formen des Textverstehens werden angesichts der oben skizzierten Wissenslücken selbst mittelfristig nicht abgearbeitet sein - sie zählen zur langfristig anzulegenden Grundlagenforschung. Ein wesentlicher Faktor bei der Lösung der dabei auftretenden Probleme wird die Förderungsintensität im Vergleich zu konkurrierenden Forschungsbereichen der sprachorientierten KI sein.

T_4 Im Gegensatz zur internationalen Forschungsszene der sprachorientierten KI ist die Forschungs- und Förderungsintensität im nationalen Bereich in den Sektoren Systeme zum Verstehen gesprochener Sprache (SVS), Frage-Antwort-Systeme (FAS) und Textverstehenssysteme (TVS) annähernd gleichgewichtig verteilt. Daraus resultiert ein überraschender (auch für die nicht KI-orientierte LDV und Informationslinguistik geltender) Positionsvorteil der nationalen Textforschung, der bei steigendem Bedarf an intelligenten Textverarbeitungssystemen durch zunehmende Elektronisierung der Textproduktion, - distribution und -speicherung auch für die industrielle Entwicklung von Produktionssystemen vorteilhaft sein könnte.

Im internationalen Forschungsbetrieb der sprachorientierten KI ist die Intensität der Förderung (Finanzmittel) und Forschung (Mannjahre, Publikationen) zu Frage-Antwort-Systemen stets um ein Vielfaches höher gewesen als zu Textverstehenssystemen, während die Förderung von Systemen zum Verstehen gesprochener Sprache annähernd achsenverkehrt zu der der Frage-Antwort-Systeme verläuft:



Demgegenüber steht eine nationale Forschungs- und Förderungstradition, in der (natürlich unter Beachtung der unverhältnismäßig kleineren "Stückzahlen") SVS- (HEX), FAS- (PLIDIS, BACON, HAM-RPM/ANS) und TVS-Projekte (KIT, SEMSYN, TOPIC, CON3TRA) annähernd gleichgewichtig gefördert wurden. Aus dieser günstigen Konstellation sollte zusammen mit dem historisch gewachsenen Theorievorsprung der europäischen und deutschen Textlinguistik durch Beibehaltung dieser im Effekt balancierten Förderungspolitik und Forschungspraxis festgehalten werden, um bei steigender Anwendungsrelevanz des Textsektors wissenschaftlich besetzte Positionen einfacher ausbauen zu können. Dies wäre eine grundsätzlich andere Ausgangsposition als in den vergleichbaren SVS- und FAS-Bereichen, in denen ein gewaltiger internationaler Forschungsvorsprung mühsam egalisiert werden mußte.

Ein genereller Effekt einer weiter auf Diversifikation ausgerichteten Praxis der Förderung der Grundlagenforschung im Bereich der sprachorientierten KI besteht darin, daß die sich abzeichnenden Großprojekte insbesondere unter europäischen organisierten Trägerschaften (ESPRIT, EUROTRA u.ä.) mit ihrer starken Betonung von Entwicklungs- statt Grundlagenforschungsaufgaben die notwendige Theorieunterfütterung erhielten. Dabei soll-

te auch der Eindruck nicht weiter verstärkt werden, daß es sich bei der Entwicklung ambitionierter, den Prinzipien der sprachorientierten KI verpflichteter natürlichsprachlicher Systeme um die Bestimmung einer ingenieurwissenschaftlichen Restkonstante handelt. Denn anders als in technisch-naturwissenschaftlichen (Vision, Robotics und großen Teilen von Expertensystementwicklungsumgebungen) und formalwissenschaftlichen Anwendungskontexten (Theorem-Proving) der KI, die durch ein vergleichsweise entwickeltes und robustes Kernwissen gekennzeichnet sind, unterliegen Sprachverstehenssysteme noch empirischen und konstruktiven Engpässen, so daß die z.T. fraprierenden wissenschaftlichen Durchbrüche in anderen Anwendungsgebieten nicht zwangsläufig zur Prognose für die kurzfristige Durchsetzung von natürlichsprachlichen Systemen am Hochtechnologiemarkt taugen.

T_5 Das momentan verfügbare Wissen über natürliche Sprachen ist hinreichend, um in ihrer Leistungscharakteristik attraktive, an industriellen Markterfordernissen ausgerichtete natürlichsprachliche Produktionssysteme für die Erkennung gesprochener Sprache, natürlichsprachliche Schnittstellen, automatische Übersetzung usf. mit klar umrissenen Grenzen in bezug auf ihre Funktionalität zu lancieren. Solche sprachindustriellen Verwertungsprojekte setzen dabei eindeutig auf aus KI-Perspektive eher konventionellen (grammatik-basierten) Methodenstandards der LDV auf (INTELLECT, TAUM-METEO etc.).

Dieser evolutionäre Weg wird ergänzt durch mittelfristig angelegte Entwicklungsarbeiten, in deren Zentrum durch die dezidierte Anwendung des entwickelteren Methodenrepertoires der KI vor allem ein immenser Zuwachs an pragmatisch defi-

nierter Funktionalität natürlichsprachlicher Systeme (Handlungsberatung, Erkennung von und Reaktion auf Benutzerintentionen usw.) steht. Entsprechende Arbeiten befinden sich aber erst in einem vorindustriellen Reifestadium prototypischer Entwicklungsprojekte.

Das rasant zunehmende Interesse der DV-Industrie an Sprachverarbeitungs-konzepten und -systemen wird damit in Übereinstimmung mit klassischen ingenieurwissenschaftlichen und ökonomischen Prinzipien kanalisiert: Den Schwerpunkt einigemaßen stabilen und ausdifferenzierten Wissens über natürliche Sprachen bilden Teil-Grammatiken einiger ausgewählter Sprachen, zu denen nur mit Einschränkungen auch das Deutsche gezählt werden kann. Konsequenterweise beziehen sich am Technologiemarkt orientierte, auf Robustheit abgestellte Sprachanalyse-systeme ganz erheblich auf diese schon weitgehend abgesicherten Vorarbeiten. Die für komfortable Sprach-

verstehenssysteme so essentielle Entwicklung pragmatischer und auf Inferenzmechanismen abgestützter semantischer Systemkomponenten kann diesen Grad der Absicherung und Elaboriertheit nicht nachweisen - deswegen ihre jetzige Zuordnung zu Laborentwicklungstätigkeiten, die aber bereits ein beträchtliches Maß an Erwartungen in die Zugkraft der selbst noch experimentellen KI-Methodik beweist.

Zusätzliche Verfeinerungen bestehender und das Angebot bislang unbekannter System-Features haben aber nur dann eine Marktchance, wenn sie jenseits akademischer Sinnfälligkeit auch gewünschte (Bedarf) und den Benutzern/Kunden einsichtige und von ihnen gesuchte (Akzeptanz) Verbesserungen mit sich bringen. In diesem, der enthusiastischen Aufbruchstimmung der KI-Industrie nachgeschalteten Filter liegt letztlich die realistische Chance einer Einwirkung von Konzepten der sprachorientierten KI auf industriell organisierte Märkte und die eines ingenieurhaft geprägten Umgangs mit sprachlichem Wissen.

DER SCHATZMEISTER INFORMIERT, BITTET, MAHNT ...

1. Diejenigen, die bereits dem Einzugsverfahren angeschlossen sind, mögen sich bitte nicht wundern, wenn der Jahresbeitrag 1985 erst kürzlich von ihrem Konto abgebucht wurde. Das hat seine Richtigkeit.
2. Wer seine Anschrift und/oder seine Bankverbindung wechselt, ist herzlich aufgefordert, mir dies mitzuteilen. So ist z.B. nicht möglich, die Jahresbeiträge der Mitglieder BRECHT, GEHRKE, GUCKLER, HITZENBERGER, KUHLIN, WINKLER per Lastschrift einzuziehen, weil jeweils die mir allein bekannte Bankverbindung nicht mehr stimmt.
3. Wer sich dem Einzugsverfahren anschließen möchte, kann dies gerne tun, indem er mir eine Einzugsermächtigung zuschickt, auf der neben Name, Anschrift, Bankverbindung, Betreff (=Jahresmitgliedsbeitrag GLDV) nur noch der Satz notiert sein muß: Hiermit ermächtige ich ... Im übrigen verfügt Ihre Bank über die entsprechenden Vordrucke.
4. Wer nun partout darauf besteht, nicht abbuchen zu lassen, sondern selbst überweisen zu wollen, möge dies - so es nicht schon geschehen ist - möglichst umgehend tun - am besten auch schon gleich für 1986. Die Gesellschaft braucht jede Mark, nicht zuletzt weil die Herstellung, d.h. der Druck des LDV-Forums eine Menge Geld kostet.

Burkhard Schaefer/Fachbereich 3/Universität Essen Gesamthochschule/Postfach 103764/4300 Essen 1

EILIGE THESEN ZUR ZUKUNFT NATÜRLICHSPRACHLICHER SYSTEME

K. Morik

Technische Universität Berlin
Fachbereich Informatik
Fachgebiet Computergestützte Informationssysteme
Projektgruppe KIT
Sekretariat FR 5-8
Franklinstr 28/29
D-1000 Berlin 10

Das Gebiet der natuerlichsprachlichen Systeme (NLS) hat sich seit einiger Zeit schon stark aufgefaechert. Daher muesste eine gruendliche Vorausschau auf Entwicklungen von NLS auf weitaus mehr Aspekte eingehen, als ich es hier tue. Ich mochte im folgenden nur kurz einige Thesen skizzieren, die als Vorschlaege fuer ausfuehrlichere Diskussionen zu verstehen sind.

1. These: Stellung von NLS im Zyklus von Systemen

Ein einfaches Zyklusmodell von Systemen unterscheidet zwischen

- > der Vorphase, in der ein bestimmtes Paradigma gewaehlt wird und ein Kernprozess im Ansatz implementiert wird,
- > der experimentellen Phase, in der ein Gesamtsystem implementiert wird, das die im gewaehlten Paradigma wichtigsten Phaenomene modelliert (tiefe Verarbeitung),
- > der anwendungsorientierten Phase, in der das Experimentalsystem optimiert, Loesungen verallgemeinert und die Performanz verbreitert wird (breite Verarbeitung),
- > der Entwicklungsphase, in der das anwendungsorientierte System fuer den Einsatz aufbereitet wird, und schließ- der Einsatzphase, in der das System verkauft und eingesetzt wird, wobei weitere Verbesserungen (Versionen) erstellt werden.

Die Entwicklungs- und Einsatzphase ist meist parallel zu einer erneuten Experimentalphase, in der auf die Erfahrungen mit dem ersten Experimentalsystem und anwendungsorientierten System aufgebaut wird. Bei einem Paradigmenwechsel wird auch erneut in die Vorphase eingetreten.

Im Bereich von NLS sind zur Zeit 4 Systeme in der Einsatzphase, einige Projekte entwickeln in einem zweiten

Durchlauf Experimentalsysteme, gleichzeitig sind aus der Linguistik neue theoretische Entwicklungen aufgenommen worden: situational semantics, DRT, GPSG, LFG. Es ist also zu jeder Phase mindestens ein System(-entwurf) angebar. Die ist ein Grund fuer die Vielfalt im Gebiet von NLS.

2. These: Funktionen von NLS

Wurden frueher oft NLS mit Anfragensystemen fuer Datenbanken gleichgesetzt, so fehlt daeuer nun jede Berechtigung. NLS werden nicht nur als Zugangssysteme entwickelt, sondern als Hilfssysteme, Erklaerungssysteme, zur Einfuehrung in ein bestimmtes System, zur Unterstuetzung bei interaktiver Problemloesung - kurz: ueberall dort, wo vorgefertigter Text an den Benutzer ausgegeben wurde, ist ein Ansatzpunkt fuer eine wissensbasierte Behandlung natuerlichsprachlicher Texte gegeben. Diese Tendenz - weg von vorgefertigten Meldungen hin zu einem natuerlichsprachlichen Dialog - ordnet NLS einen Platz als Komponente in einem anderen, meist einem der klassischen EDV-Systeme (z.B. Editor, Mail, Betriebssystem) zu. Bekannte Beispiele sind WIZARD und UC.

Zusaetzlich ist eine Funktion wieder aktuell geworden, die einige Zeit nicht offen angesprochen werden konnte: die maschinelle Uebersetzung. Ohne hochfliegende Ansprueche, einfach aufgrund des grossen praktischen Bedarfs ist in einigen Firmen an maschineller Unterstuetzung fuer Uebersetzungen gearbeitet worden. Die Formen der Unterstuetzung, die jetzt angeboten werden koennen, sind weit mehr als nur Suche im Lexikon. NLS sind also nicht nur Zugangssysteme, sondern Unterstuetzungssysteme in einem weiten Sinne.

3. These: Zugang zu neueren Systemen

Mit den Expertensystemen ist ein neuer Systemtypus entstanden, der fuer einen natuerlichsprachlichen Zugang andere Anforderungen stellt und auch andere Moeglichkeiten bietet als die Datenbanksysteme. Besonders interessant sind hier die Ansaetze, Zugang zu beiden Systemtypen zu bieten wie z.B. XCALIBUR. Diese Entwicklung verlauft parallel zu den Versuchen Datenbanken und Expertensysteme zu integrieren. NLS in der Funktion als Erklaerungssysteme haben den Anstoss zu einer angemesseneren Repraesentation in Expertensystemen gegeben: verschiedene Arten von Wissen muessen getrennt repraesentiert werden, kausale Beziehungen modelliert werden, verschiedene Detailliertheitsgrade zur Vefuegung gestellt werden, damit adaequate Erklarungen produziert werden koennen. Ein Projekt, das diesen Ansatz verfolgt ist z.B. das EES am USC/ISI (Swartout).

4. These: Generierung natuerlichsprachlicher Aeusserungen

Stand lange Zeit die Analyse natuerlichsprachlicher aeusserungen im Vordergrund, so ist jetzt die Generierung von Saetzen und Texten zu einem lebhaften Forschungsgebiet geworden. Dies scheint mir sowohl mit den vielfaeltigen Funktionen von NLS zusammenzuhaengen (s. 2. These), als auch mit dem fortgeschrittenen Stand von NLS (s. 1. These). Erklarungs- und Hilfssysteme sind ganz wesentlich textproduzierend. Maschinelle uebersetzung beinhaltet das Verstehen imselben Masse wie die Generierung. Insofern geben diese Funktionen von NLS Anstoss zu einer eingehenden Beschaeftigung mit Verfahren zur Sprachproduktion. Gleichzeitig ist die Zeit reif fuer eine nicht mehr auf die Verstehensseite beschaenkte Sprachverarbeitung. Waehrend die NLS auf dem Markt noch auf die Analyseseite konzentriert sind, beruecksichtigen die Experimentalsysteme und anwendungsorientierten Systeme schon

gleichgewichtig das Sprachverstehen und -generieren.

5. These: zur kommunikativen Adaequatheit

Die Abbildung von Aussagen auf einen Weltausschnitt ist das klassische Gebiet der Semantik. Dass semantische Fragestellungen zu loesen sind, um ein NLS zu konstruieren, ist allgemein akzeptiert. Dass pragmatische Probleme ein wichtiges Forschungsgebiet ausmachen, wird nur langsam anerkannt. Die Beruecksichtigung der Kommunikationssituation, in der sich Benutzer und System befinden, ist aber notwendig, sollen nicht nur richtige, sondern auch kommunikativ adaequate Aeusserungen vom System ausgegeben werden. Unter den Stichworten Partnermodellierung, Sprechakterkennung und -erzeugung, Ueberbeantwortung und Kooperativitaet sind schon seit laengerer Zeit pragmatische Aspekte behandelt worden. Durch den Einsatz von NLS erhalten diese Arbeiten nunmehr eine grosse praktische Relevanz: die Akzeptanz von NLS wird sich nicht zuletzt nach der kommunikativen Adaequatheit der Systemaeusserungen richten. Die gegenueber den 70er Jahren weitaus besseren Hardware-Bedingungen ermoeglichen zusaetzlich zu der Modellierung des Weltausschnitts die Modellierung der Kommunikationssituation. Die vielfaeltigen Funktionen von NLS (wenn auch nicht unbedingt jeweils desselben NLS) machen die Kommunikationssituation bewusst. Wird sogar angestrebt, dasselbe NLS in verschiedenen Rollen mit Benutzern interagieren zu lassen, so ist die explizite Modellierung dieser Rollen unabdingbar. Aus diesen drei Gruenden, der praktischen Relevanz, der technischen Moeglichkeit und der funktionalen Notwendigkeit, ist eine Tendenz zur Beschaeftigung mit der Kommunikationssituation zu erwarten.

**THESEN ZUR SPRACHORIENTIERTEN KI-FORSCHUNG
BZW. ZUR KI-ORIENTIERTEN SPRACHFORSCHUNG**

B. Endres-Niggemeyer

FH Hannover
Fachbereich BID
Hannomagstr. 18
D-3000 Hannover

1. Sprachorientierte KI bzw. KI-orientierte Sprachverarbeitung beziehen ihre Durchschlagskraft zum großen Teil aus dem kognitiven Paradigma, das unter anderem weit in die Linguistik - von der Psycholinguistik bis zur Universalienforschung - und in interdisziplinäre Bereiche wie die Textwissenschaft hineinreicht.
2. Zum Vorteil der KI-orientierten Sprachverarbeitung und der Computerlinguistik integriert das kognitive Paradigma ein breites Feld von Forschungsanstrengungen so, daß die Ergebnisse - vom kindlichen Spracherwerb bis zur Argumentationstheorie - relativ leicht für eine Computersimulation sprachlicher kognitiver Prozesse herangezogen werden können. Der gemeinsame begriffliche Bezugsrahmen für empirische, abstrakt theoretische und auf die Systemrealisierung orientierte Forscher macht das kognitive Paradigma zu einem Schwerpunkt wissenschaftlicher Aktivität.
3. Sprachorientierte KI bzw. KI-orientierte Sprachverarbeitung werden personell weitgehend von Informatikern getragen. Diese bringen wichtige Konzepte in die Computerlinguistik ein, haben aber grobenteils eine ihrer Ausbildung entsprechend verkürzte Vorstellung von Sprache und linguistischer Theorie. Für die sprachliche Qualität natürlichsprachlicher Systeme ist es wichtig, Defizite im sprachbezogenen Bewußtseinsstand künftiger Computerlinguisten durch gezielte Ausbildungsmaßnahmen ("Linguistik für Informatiker") auszugleichen, wie dies teilweise schon geschieht. Linguisten und besonders Computerlinguisten sind gefordert, die konzeptionellen Beiträge aus KI, Cognitive Science und allgemeiner Informatik theoretisch zu integrieren und in der linguistischen Ausbildung weiterzugeben.

**STUDIENGANG
"ALLGEMEINE SPRACHWISSENSCHAFT UND NEBENFÄCHER"
AN DER UNIVERSITÄT TÜBINGEN**

F. Guenther, W. Kreitmair

Universität Tübingen
Englisches Seminar
Wilhelmstr. 50
D-7400 Tübingen

Seit WS 1985/86 gibt es an der Universität Tübingen den Studiengang "Allgemeine Sprachwissenschaft und Nebenfächer". Es handelt sich hierbei nicht um einen Studiengang der LDV, sondern um ein interdisziplinär angelegtes Curriculum im Stile der Cognitive Science.

Die moderne Sprachwissenschaft ist heute nicht mehr denkbar ohne bestimmte theoretische Grundlagen aus dem informatischen, dem mathematischen, dem psychologischen, aber auch dem naturwissenschaftlichen Bereich. Die Komplexität des Gegenstands "Sprache" macht interdisziplinäre Zusammenarbeit von Sprachwissenschaftlern, Informatikern, Mathematikern und Psychologen unabdingbar. Der Studiengang wird diesen Anforderungen dadurch gerecht, daß er das Fach Sprachwissenschaft mit den Nebenfächern Informatik, Mathematik bzw. Psychologie kombiniert.

Aufbau des Studiengangs

Der Studiengang Allgemeine Sprachwissenschaft und Nebenfächer ist ein Magisterstudiengang der Neuphilologischen Fakultät der Universität Tübingen. Er setzt sich zusammen aus einem Hauptfach (Allgemeine Sprachwissenschaft) und zwei Nebenfächern. Hierbei sind als Nebenfächer vorerst Psychologie und Informatik vorgesehen. Es ist geplant, weitere Nebenfächer für diesen Studiengang einzurichten. Damit ist es möglich, das Studium der Sprachwissenschaft in einen sinnvollen interdisziplinären Zusammenhang zu stellen; sprachwissenschaftliche Problemstellungen werden soweit wie möglich unter informationswissenschaftlichen und psychologischen Gesichtspunkten behandelt.

Das Studium im Rahmen dieses Studiengangs besteht aus einem Grundstudium (1. bis 4. Semester) und einem Hauptstudium (5. bis 9. Semester). Die Mindeststudienzeit beträgt also neun Semester. Im Anschluß an das 3. Semester ist ein Praktikum von 2-3 Monaten, nach dem Grundstudium ein ein- bis zweisemestriger Auslandsaufenthalt vorgesehen.

Neben den inhaltlichen Anforderungen im Hauptfach und in den Nebenfächern wird erwartet, daß sich die Studenten während des Grundstudiums ausreichende Kenntnisse in zwei Fremdsprachen aneignen.

Modellstudienplan

Als Orientierungshilfe für das Studium der Allgemeinen Sprachwissenschaft mit den Nebenfächern Informatik und Psychologie kann folgender Modellstudienplan verwendet werden. Je nach Lehrangebot

können sich hier natürlich Verschiebungen ergeben. Im übrigen sollte sich jeder Student - in Zusammenarbeit mit der Fachstudienberatung - einen "individuellen" Studienplan erstellen.

Hauptfach **Allgemeine Sprachwissenschaft**

Grundstudium

1. Semester:
Einführung in die allg. Sprachwissenschaft I (VL und PS)
2. Semester:
Einführung in die allg. Sprachwissenschaft II (VL und PS)
Einführung in die Phonetik und Phonologie (PS)
3. Semester:
Syntax I (VL und PS)
Einführung in die Lexikographie (PS)
4. Semester:
Formale Methoden in der Linguistik (VL und Übungen)
Semantik I (VL und Übungen)

Hauptstudium

5. Semester:
Einführung in die Computerlinguistik (VL und HS)
Syntax II (VL und HS)
6. Semester:
Pragmatik (VL und HS) oder Semantik II (VL und HS)
Sprachphilosophie (VL oder HS)
7. Semester:
Einführung in die Psycholinguistik (VL und HS)
Geschichte der Sprachwissenschaft (VL oder HS)
8. Semester:
Dialogtheorien (HS)
9. Semester:
Neuere Entwicklungen in der Semantik (OS)

Hinzu kommen jeweils Übungen in den beiden gewählten Fremdsprachen.

Für die Zwischenprüfung ist die erfolgreiche Teilnahme an folgenden Veranstaltungen nachzuweisen:

- Einführung in die allg. Sprachwissenschaft I (PS)
- Einführung in die Phonetik und Phonologie (PS)
- Mathematische Methoden in der Sprachwissenschaft (PS)
- Syntax I (PS)
- Semantik I (PS)

Der Nachweis der erfolgreichen Teilnahme an folgenden Veranstaltungen ist für die Magisterprüfung obligatorisch:

- Einführung in die Psycholinguistik (HS)

Einführung in die Computerlinguistik (HS)
Syntax II (HS)
Semantik II (HS) oder Pragmatik (HS)

Nebenfach Informatik

Grundstudium

1. Semester:
Einführung in die Informatik I (VL und Übungen)
2. Semester:
Einführung in das systematische Programmieren (VL und Übungen)
3. Semester:
Logik (VL und Übungen)
4. Semester:
Nichtnumerische Datenverarbeitung und Information Retrieval (VL und Übungen)

Hauptstudium

5. Semester:
Einführung in die Informatik II (VL und Übungen)
6. Semester:
Aufbau und Struktur von Betriebssystemen (VL)
7. Semester:
Computerarchitektur (VL)
8. Semester:
Mikroprozessoren und Mikrocomputer (VL)

Für die Zwischenprüfung sind folgende Scheine erforderlich:

Einführung in die Informatik I
Einführung in das systematische Programmieren

Im Hauptstudium sind zwei beliebige Scheine zu erbringen.

Nebenfach Psychologie

Grundstudium

1. Semester:
Einführung in die Kognitionspsychologie (VL)
Methodenlehre (VL)
2. Semester:
Motivation und Lernen (VL)
Gedächtnis und Wissen (S)
3. Semester:
Sprachpsychologie (S)

Spracherwerb (S)

4. Semester:
Denken und Problemlösen (S)
oder
Sprache und Kognition (S)

Hauptstudium

5. Semester:
Biokommunikation (S)
Interaktion und Kommunikation (S)
6. Semester:
Mathematische Psychologie (S)
Neurophysiologie und Neuropsychologie der kognitiven
Funktionen (S)
7. Semester:
Wahrnehmungstheorie
Bewußtsein und Aufmerksamkeit (S)
8. Semester:
Psychologie der Entscheidungsprozesse (S)
9. Semester:
Sprachpathologie (VL oder S)

Aus den hier angegebenen Veranstaltungen sind im Grundstudium und im Hauptstudium je 2 beliebige Scheine erforderlich.

Praktikum und Auslandsaufenthalt

Das erforderliche 2-3-monatige Praktikum sollte nach dem 3. Semester durchgeführt werden. Hierzu wurde mit einer Reihe von Verlagen, Fachinformationszentren, Software-Herstellern etc. Kontakt aufgenommen.

Für das Auslandsstudium nach dem 4. Semester gibt es eine Reihe von Zusagen amerikanischer, französischer und englischer Universitäten, die über einen ähnlichen Studiengang verfügen.

Mögliche Berufsfelder

Für folgende Berufsfelder qualifizieren sich die Absolventen dieses Studiengangs:

- 1) Informationsverarbeitung und Dokumentation (Fachinformations- oder Dokumentationsstellen, bibliographische Institute, Verlage, Verwaltungen usw.);
- 2) Natürlichsprachliche Schnittstellen zu Datenbanken, Information Retrieval, Aufbau linguistischer Datenbanken (Lexika, Thesauri etc.), Textverarbeitung, Maschinelle Übersetzung usw.);
- 3) Pädagogik, Computer-gestützter Unterricht;
- 4) Sprachtherapie.

KI als Pflichtfach im Fachhochschulbereich**W. Brecht**

TFH-Berlin
 Fachbereich Informatik
 Luxenburgerstraße 10
 D-1000 Berlin 65

Mit Beginn des Sommersemesters 1985 ist an der Technischen Fachhochschule Berlin neben den bereits vorhandenen informatikorientierten Studiengängen Technische Informatik und Mathematik ein neuer Studiengang Allgemeine Informatik eingerichtet worden. Das erste Semester startete mit ca. 40 Studenten.

Der Studiengang weist folgende grobe Fächerverteilung auf:

56 % : Software;
 20 % : Mathematik;
 17 % : Hardware;
 7 % : Allgemeine Wissenschaften (Englisch, Recht, u.ä.).

Das erste Semester beginnt mit einer Lehrveranstaltung "Grundbegriffe der Informatik", außerdem werden bereits das Fach "Algorithmen und Datenstrukturen" gelehrt und ein dreisemestriger Kurs "Strukturiertes Programmieren (PASCAL)" begonnen. Im zweiten Semester folgt das Fach "Betriebssysteme", in dem auch der Umgang mit der Systemsoftware der TFH-Rechner geübt wird. In diesem zweiten Semester beginnt dann auch ein dreisemestriger Kurs "Software Engineering", in den die Anfertigung einer Studienarbeit integriert ist. Im dritten bzw. vierten Semester sind die Fächer "Datenbanksysteme", "Operations Research" und "Maschinenorientiertes Programmieren" vorgesehen sowie die im technischen bzw. kommerziellen Bereich noch immer weit verbreiteten Sprachen FORTRAN und COBOL. Mit dem vierten Semester ist das Grundstudium abgeschlossen.

Im Hauptstudium (5-tes und 6-tes Semester) sind pflichtgemäß vorgesehen: LISP, Fortführung des "Maschinenorientierten Programmierens", "Systemprogrammierung", "Daten- und Rechnernetze", "Compilerbau", "Graphische Datenverarbeitung" und "Künstliche Intelligenz".

Zur Veranstaltung "Künstliche Intelligenz" (sie gehört in das 6-te Semester) ist zu bemerken, daß sie im Studienplan nicht isoliert ist, sondern durch die beiden Pflichtfächer "Prädikatenlogik" (im 4-ten Semester) und "Programmieren in LISP" (im 5-ten Semester) unterstützt und vorbereitet wird.

Durch die Verankerung als Pflichtfach im Studienplan hat die Ausbildung in Künstlicher Intelligenz an der TFH einen hohen Stellenwert erhalten. Damit soll der Entwicklung dieser Disziplin Rechnung getragen werden, die in den letzten Jahren mit zunehmender Tendenz zur praktischen Anwendung gelangt und damit für die Absolventen von Informatik-Studiengängen von Bedeutung ist.

Überlegungen an der TFH zur Relevanz der Informatik in den zukünftigen Arbeitsbereichen der Ingenieure hat dazu geführt, daß ein Vorschlag für einen Aufbaustudiengang Ingenieurinformatik

erarbeitet wurde, der zur Zeit diskutiert wird. Er soll in einer dreisemestrigen Tages- bzw. einer fünfsemestrigen Abendform (beide sollen angeboten werden) bereits fertige und eventuell schon im Beruf stehende Ingenieure in Informatik ausbilden.

Auch in diesem Studiengang ist eine Veranstaltung "Künstliche Intelligenz" als Pflichtfach vorgesehen, die durch eine zweisemestrige "Mathematische Logik" vorbereitet wird. Aus zeitlichen Gründen konnten Programmierübungen in LISP nicht eigens ausgewiesen werden, sondern mußten in die Lehrveranstaltung integriert werden.

Die inhaltliche Ausfüllung der "Künstlichen Intelligenz" wird ausgesprochen anwendungsbezogen sein. Das heißt, daß die Lehrveranstaltung (4 Semesterwochenstunden) zu einem sehr großen Teil aus Übungen im Umgang mit vorhandener (zugänglicher) ausgewählter KI-Software (eventuell auch KI-Hardware) besteht. Bei den Software-Übungen werden Bedienung, Anwendung, Implementierung, Änderung und Erweiterung behandelt. Theoretische Fragestellungen können nicht vertieft werden, jedoch wird ein Überblick über die Teilgebiete der Künstlichen Intelligenz, über ihre Methoden und Techniken sowie über ihre Problembereiche gegeben.

Es wäre nützlich, wenn für die konkrete Ausfüllung des Stoffplanes vorhandene Erfahrung mit ähnlichen Lehrveranstaltungen einbezogen werden könnte. Für diesbezügliche Vorschläge und Hinweise bin ich dankbar.

RINGVORLESUNG "GRAMMATIK UND KOMMUNIKATION" TU BERLIN

Im Sommersemester 1985 fand am Institut für Linguistik der Technischen Hochschule Berlin in Zusammenarbeit mit dem Semiotischen Arbeitskreis Berlin (SAB) eine Ringvorlesung über Grammatik und Kommunikation statt. Die Veranstaltungsreihe diente der Klärung von Problemen, die in der Ringvorlesung vom Sommersemester 1984 über "Zeichenökonomie in Sprache und Sprachtheorie" offengeblieben waren. Es sprachen u.a.:

- R.N. Srivastava (Neu-Delhi) über "Grammar and Nongrammatical Rules of Verbal Communication" (6.6.85).
 - Franz Hundnurscher (Münster) über "Dialog und Grammatik" (13.6.85).
 - Helmar Frank (Paderborn) über "Die Reduktion grammatischer Redundanz in der Sprachplanung" (4.7.85).
 - Manfred Krause (Berlin) über "Grammatik in der Mensch-Maschine-Kommunikation" (18.7.85).
- Die Vortragsreihe wird fortgesetzt.

- Robert de Baugrande (Gainesville/Florida) über "Die Entwicklung des Sprachsystems aus der Kommunikationserfahrung" (2.5.85).

(Semiotischer Arbeitskreis Berlin)

LINGUISTISCHE DATENVERARBEITUNG AN DER UNIV. MAINZ

An der Johannes Gutenberg-Universität Mainz wurden vom Institut für Allgemeine und vergleichende Sprachwissenschaft (FB 14-29) im Jahr 1985 die folgenden Lehrveranstaltungen im Bereich der maschinellen Analyse natürlicher Sprache angeboten:

- Prof. Dr. D. Seelbach: Computerlinguistik und Informationswissenschaft (Vorlesung).

- Prof. Dr. D. Seelbach: Methoden und Anwendung der Sprachorientierten KI (Hauptseminar).
- Prof. Dr. W. H. Veith: Rechnerunterstützte Sprachanalyse des Deutschen (Hauptseminar).

**RINGVORLESUNG "SPRACHE UND COMPUTER"
AN DER UNIV./GESAMTHOCHSCHULE DUISBURG**

Fachbereich Sprach- und Literaturwissenschaften
dienstags von 18 bis 20 Uhr im Raum LK 052

- 22.10. 85 Prof. Dr. René Dirven (Trier/Duisburg):
Auswertung der Rückübersetzung mittels Computer
- 29.10. 85 Prof. Dr. Walther von Hahn (Hamburg):
Zum gegenwärtigen Stand der Sprachsimulation und der Sprachgenerierung
- 5.11. 85 Prof. Dr. Bernd Spillner (Duisburg)
Möglichkeiten und Grenzen der automatischen Sprachübersetzung
- 12.11. 85 Prof. Dr. L. K. Engels (Leuven):
Didaktisches Lexikon im Fremdsprachenunterricht
- 19.11. 85 Dr. Ulrich Schmitz (Duisburg):
Computer als Hilfsmittel der Sprachwissenschaft
- 26.11. 85 Dr. Helmut Frick (Duisburg):
Rechnernetze und Rechnerkommunikation
- 3.12. 85 Dr. Rainer Rauch (Duisburg):
Mensch-Maschine-Kommunikation
- 10.12. 85 Prof. Dr. Udo L. Figge (Bochum):
Automatische Simulation menschlicher Sprachverarbeitung
auf kognitiver Basis
7. 1. 86 Prof. Dr. Walter Geisselhardt (Duisburg):
Expertensysteme
14. 1. 86 Prof. Dr. Wolfgang Wahlster (Hamburg):
Natürlich-sprachlicher Zugang zu Expertensystemen
21. 1. 86 Dr. Christine Klein-Braley (Duisburg):
Kann man mit dem Computer die Verständlichkeit von Texten vorhersagen?
28. 1. 86 Prof. Dr. Dieter Rumpel (Duisburg):
Erkennbarkeit von Abkürzungen
4. 2. 86 Dr. U. Kling (Duisburg):
Künstliche Intelligenz im Bildungsbereich

Information:Herr Schmitz: 0203/3792397

CLEARINGHOUSE ON NATURAL LANGUAGE PROCESSING SYSTEMS AND INTELLIGENT USER INTERFACES

Battelle is currently conducting an industry-sponsored analysis of intelligent user interfaces, in particular natural language systems. The sponsors of this project are interested in assessing the state-of-the-art of IUI/NLP and identifying ongoing research for possible funding.

An important part of this survey is the establishment of a Clearinghouse hosted on a PC in dBase II format. The Clearinghouse will allow researchers and companies alike to access information about ongoing projects.

If you are currently working on an NLP or IUI project, please let us know, and we will send you our questionnaire. In return, we will make

the result of our survey available to you. As a further benefit, your project will be included in upcoming publications (ACL, ACM, etc.) related to the Clearinghouse.

We hope you will be able to cooperate and help us to establish this needed network of peoples/projects in the area of IUI/NLP.

Please contact: Battelle Columbus Laboratories, Dr Klaus Obermeier, Artificial Intelligence/Knowledge-Based Systems, 505 King Avenue, Columbus, OH 43201-2693 (614) 424-5570

aus: SIGART Newsletter July 1985, Nr. 93

ARTIFICIAL INTELLIGENCE IN SWITZERLAND

To explode the myth that the interest in AI is at best lukewarm, a preliminary one-day conference on Artificial Intelligence was organised at short notice at the University of Geneva on 29th June 1984. Five papers were presented by Swiss authors and an audience of about 60 expressed interest in the formation of a Swiss association for AI.

At the general assembly of the Swiss Society of Informaticians (SI) on October 18, 1984, the Swiss Group for Artificial Intelligence and Cognitive Science (SGAICO) was founded as a special interest sub-group of SI. The principle and official goal of SGAICO is to promote the study of AI and CS in Switzerland by developing contacts inside and outside the country, and helping to establish a rapport between universities and industry.

To these ends, SGAICO aims to become a member of the ECCAI, and the procedure to establish this is now in the final stages. In addition, the activities of the SGAICO during 1985 will

include the publication of a newsletter, and the organisation of a five day tutorial on Knowledge Based Systems during October in Lugano.

The sub-group is being provisionally run by P. Shann, M. Rosner and R. Pfeifer. For further information, contact any of M. Rosner or P. Shann, Institut Dalle Molle ISSCO, 54 route des Acacias, CH 1227 Geneva
R. Pfeifer, Soziologisches Institut der Universität Zürich, Zürichbergstr. 43, CH 8044 Zürich

aus: ECCAI 2(1985), 1

FuE-PROJEKT FÜR ELEKTRONISCHES PUBLIZIEREN

Ende 1983 hat sich ein Konsortium aus Verlagen und Fachinformationseinrichtungen der BRD gebildet, um gemeinsam ein Forschungs- und Entwicklungsvorhaben in den Jahren 1984 und 1985 durchzuführen. Beteiligt sind:

- Verlag für Technische Regelwerke GmbH (Tochtergesellschaft des DIN, Deutsches Institut für Normung e.V., Berlin)
- Springer Verlag GmbH & Co.KG (Berlin, Heidelberg, New York, Tokyo)
- VDI-Verlag GmbH (Düsseldorf)
- C.E.Poeschel Verlag (Stuttgart)
- Fachinformationszentrum Energie Physik, Mathematik GmbH Karlsruhe
- Gesellschaft für Information und Dokumentation mbH (Frankfurt)

Bei diesem Vorhaben geht es, wie in der Broschüre "Forschungs- und Entwicklungsvorhaben >Elektronisches Publizieren technisch-wissenschaftlicher Texte<" formuliert wird, um den konsequenten Einsatz elektronischer Techniken. Der Einsatzbereich umfaßt die Manuskripterstellung, den Verlag und das Angebot über Datenbanken und auch die vollständige Auslieferung des Dokuments mit vollem Text Graphik auf das Gerät des Lesers bzw. der Setzerei für die Herstellung eines Druckfilms.

Mit dem Ziel, Möglichkeiten und Grenzen des "Elektronischen Publizierens" und seine u.U. weitreichenden Folgen aufzuzeigen, wird das Vorhaben von der Kommission der Europäischen Gemeinschaft und dem Bundesministerium für Forschung und Technologie mit öffentlichen Mittel gefördert. Der finanzielle Eigenaufwand der Konsorten ist erheblich.

Das Ziel des Vorhabens ist es, einen elektronischen Ablauf für das

- * Editieren,
- * Speichern in geräteunabhängigem Format,
- * Weiterverarbeiten und Liefern

technisch-wissenschaftlicher Texte zu entwickeln und für ein breites Anwendungsspektrum in der Praxis erproben.

Das Ergebnis soll sein, Aussagen hinsichtlich technischer Durchführbarkeit, wirtschaftlicher und organisatorischer Lebensfähigkeit der Verfahren und dem Grad der Benutzerakzeptanz zu gewinnen.

Die in den Versuchen zu bearbeitenden technisch-wissenschaftlichen Dokumente beinhalten alle komplexen Elemente einer Volltextspeicherung, wie Texte, Formeln, Tabellen und Bilder. Im einzelnen sind einbezogen:

- 1 Technische Normen (DIN) und Richtlinien (VDI) als Beispiel einer komplexen Dokumentart, bestehend aus Texten und Bildern;
- 2 Mathematische Texte und Formeln hohen Schwierigkeitsgrades als Beispiel für die Probleme beim Editieren und späteren Darstellen auf auf geeignete Ausgabemedien;
- 3 Unveröffentlichte wissenschaftliche Manuskripte als Beispiel für "Graue Literatur" und "Auslieferung Bestellung (Publishing-on-demand)";

Schlußfolgerung durch Konsorten, Förderer und Fachöffentlichkeit werden am Ende des Experiments zu ziehen sein,

- 1 wenn die Entwicklungen im Projekt bei den Konsorten eingesetzt und der interessierten Öffentlichkeit im GID-Labor zu Testzwecken zur Verfügung gestellt werden;
- 2 wenn die Ergebnisse der informationswissenschaftlichen Untersuchungen bei Nutzern und gesellschaftlichen Gruppen die Auswirkung Elektronischen Publizierens und Lieferens technisch-wissenschaftliche Dokumente erkennen lassen.

Angesichts der anspruchsvollen Ziele werden während der Laufzeit des Experiments bis Ende 1985 nur die grundlegenden Entwicklungen und Evaluierungen zu leisten sein. Ob und in welcher Form das Projekt weitergeführt wird, kann erst entschieden werden, wenn die Dokumentlieferung 1985 getestet wird.

Wer sich bereits jetzt näher über das Projekt "Elektronisches Publizieren technisch-wissenschaftlicher Texte"

informieren will, kann bei der Gesellschaft für Information und Dokumentation mbH, Sektion für Systementwicklung, Tiergartenstr. 17 D-6900 Heidelberg, eine gleichnamige Broschüre bestellen, die vom Konsortium herausgegeben wird.
(Quelle: die angegebene Broschüre).

SYNONYM FINDER TOOL

Synonym Finder, a new tool for use while word processing, consists of a database of 9000 keywords. It can be used as a thesaurus as well as a spelling verifier, and runs on the PC-DOS Wordstar and Multimate versions. Details from Writing Consultants, 11 Creek Bend Dr., Fairport NY 14450.

aus: SCOPE 3(1985),1

SOFTWARE ZUR TEXTANALYSE

ARRAS, ein System, um Konkordanzen, Indexe zu erstellen und ähnliche Werkzeuge zur Textanalyse benutzen eine einfache Englisch-basierte Kommandosprache. Bezug über: John B. Smith, Department of Computer science, University of North Carolina, Chapel Hill NC 27514; 919/967-0819 oder 814/422-8486.

aus: SCOPE 3(1985),1

PHILOSOPHICAL SPELLING CHECKER

Phildict, a spelling checker glossary, consists on proper names and philosophical terms organized in seven ASCII files. All entries of the Dictionary of Philosophy by Dagobert Rune are included. Cost is \$75. Details from Walter A. Coole, Sacramento, 1325 Shirley Pl., Mt. Vernon WA 98273; 206/ 428-0441.

aus: SCOPE 3(1985),1

TEXT FILE COMPRESSION AND BASIC

Textpress, a program for text file compression during word processing, costs \$52,95 and is available at The Alternate Source, 704 North Pennsylvania Ave., Lansing MI 48906<; 517/482-8270. Also available is Vivace, a Basic compiler that is 100% compatible with the host interpreter. It is currently available for Radio Shack's Model 4 and TRS-80 Model III at a cost of \$49,95.

aus: SCOPE 3(1985),1

KLASSIFIKATION ALS WERKZEUG DER FORSCHUNG

I. Dahlberg

Gesellschaft für Klassifikation
D-6000 Frankfurt/M 50

Zu diesem Thema waren Vortragsvorschläge für die 9. Jahrestagung der Gesellschaft für Klassifikation, Karlsruhe, 26.-29.6.1985, erbeten worden. Da die Tagungsleitung (Prof. Dr. Wolfgang Gaul, Institut für Entscheidungstheorie u. Unternehmensforschung der Universität Karlsruhe) erstmalig die Richtung der numerischen Klassifikation vertrat, war die Mehrzahl der Vorträge (89 von insgesamt 106) aus dem Bereich der Daten- und Clusteranalyse eingereicht worden. Unter den 17 Vorträgen der sog. nicht-numerischen Klassifikation, also der begriffsorientierten Richtung, fielen 10 in die Sektion "Concept Analysis". Über sie soll hier kurz berichtet werden.

Schon bei der 8. Jahrestagung in Hofgeismar hatte Rudolf WILLE, TH Darmstadt seine Konzeption einer graphentheoretischen Darstellung von Begriffssystemen erläutert (in der erschienen in Studien zur Klassifikation, Bd.15, 1984, in Englisch in Int.Classif.11(1984)No.2 - "Line Diagrams of Hierarchical Concept Systems"). Inzwischen kann man schon fast von einer Wille-Schule beim FB Mathematik in Darmstadt sprechen. Sie kam in Karlsruhe dadurch zum Ausdruck, daß fünf "Wille-Schüler" und er selbst neuere Arbeiten zu diesem Thema brachten. Im einzelnen sprachen M.SKORSKY, B.GANTER und K.RINDFREY ("Programs for Concept Analysis") über eine Reihe von Fortran Programmen (für VAX 11/780) und zeigten P.LUKSCH u.M.SKORSKY ("Drawing concept lattices with a computer"), wie man mit Computerhilfe, ausgehend von einem Satz von Standard-Liniendiagrammen ("graphic scales" genannt), einen Kontext in eine Auswahl von bekannten "graphic scales" zerlegt und diese wiederum in einem Liniendiagramm miteinander verbindet (das Programm verwendet verschiedene Konstruktionsverfahren der Gittertheorie). J.STAHL, B.GANTER, u.R.WILLE ("Ideas of conceptual measurement") verwenden eine auf Begriffsstrukturen basierende Theorie des Messens, die von der entwickelten, formalen Be-

griffsanalyse abgeleitet werden kann. In seinem, diese Serie abschließenden Vortrag ("From preconcepts to concepts") ging R.WILLE davon aus, daß ein Kontext eine Hierarchie von Begriffen bestimmt, womit also das Problem der Begriffsbildung von dem der Formulierung der Kontexte abhängt. Als mathematisches Hilfsmittel für die Erstellung von Modellen von Begriffsfeldern werden Relationalstrukturen herangezogen. Durch einige Beispiele (z.B. 'Raum' und 'Zeit') wurde die Methode veranschaulicht.

Rudolf UNGVARY, Budapest, berichts bei der 7. Jahrestagung in Königswinter die Möglichkeit einer beziehungsartenorientierten Representation der Begriffsbeziehungen in einem Thesaurus (siehe Int.Classif.10(1983)No.2, p.63-8 - "Application of the Thesaurus Method to the Communication of Knowledge") dargestellt hatte, befaßte sich diesmal mit einer "Empirischen Untersuchung begrifflicher Dualität". Er bezeichnete die elementare, begriffliche Erreichbarkeit innerhalb eines Begriffspaars (wie z.B. Raum-Zeit, Nichts-Etwas, Materie-Energie) als die Relation der intensionalen Toleranz. Er vermutet, daß diese eine entscheidende Rolle in der semantischen Tiefenstruktur einer Sprache habe.

Norbert MEDER, Pädagogisches Seminar der Universität zu Köln, versuchte in seinem Vortrag über die "Künstliche Intelligenz als Werkzeug der Klassifikation" die Prinzipien herauszuarbeiten, denen die KI folgt, sofern sie ihren medialen Grundsätzen gemäß eingesetzt wird, was hier konkret durch die Antwort auf die Frage geschah: Wie kann ich durch die Konstruktion und Programmierung intelligenter Operationen etwas Neues über die menschlichen Kognitionen in Erfahrung bringen? Er entwickelte seine Gedanken über Begriffsbildung und angemessene Begriffsverbindungen an Marc DEMEY's Werk "The Cognitive Paradigm".

Helmut MÖNKE, Traunstein, der bei den Jahrestagungen in Frankfurt-Höchst, 1978 und Salzburg 1980 über seine

wird. Die Vorträge von Ungvary und Meder erscheinen in Englisch jedoch bereits 1985 im Heft 3 von International Classification.

der linguistischen Definition" und verwendete als klassisches Beispiel für eine heuristische-produktive, analytische Definition einzelner Elemente eines Begriffsfeldes und deren sinnvolle Ordnung das phonologische System.

Dem Sprachwissenschaftler Mönke folgte der Terminologiewissenschaftler Wolfgang NEDOBITY, Wien ("Classification, Artificial Intelligence and Cognitive Psychology"). Er machte klar, daß sowohl KI-Forschung als auch die Klassifikationswissenschaft an der Entwicklung und Konstruktion von wissensbasierten Systemen arbeiten und zeigte aufgrund von vier Ansätzen von Wissensrepräsentation in der KI, die Don WALKER 1981 einmal aufgestellt hatte, welche Parallelen sich zu den Tätigkeiten und Ergebnissen der Klassifikationsforschung ziehen lassen.

Victor E. WEISSMANN, TU Berlin, bezog sich in seinen Ausführungen ("A Framework for a Theory of Statistical Applications") zur Theorie von Informationssystemen auf die Arbeiten von SNEED und STEGMÜLLER. Es zeigte sich jedoch - insbesondere auch in der

Diskussion - daß eine Theorie von statistischen Anwendungen vermutlich wenig hilfreich eingesetzt werden kann.

Ein weites Anwendungsfeld begriffstheoretischer Untersuchungen wurde dagegen aus dem Vortrag von Klaus M. SCHMIDT, Bowling Green, Ohio, ersichtlich, der seit 1975 an dem Projekt eines "Conceptual Dictionary for Medieval German Epics" arbeitet und in seinem Vortrag ("Problems of Classification and Application of Conceptual Systems to Historical Languages") die begriffsanalytische Arbeitsweise umriß, die an "tote" Sprachen angelegt werden muß. Es sei an dieser Stelle auch auf die Konferenzen hingewiesen, die von der von K. Schmidt gegründeten SCCAC (Society for Conceptual and Content Analysis by Computer) veranstaltet werden. (Über SCCAC siehe Int. Classif. 1985-1, p. 21).

Die Vorträge werden in einem Proceedingsband erscheinen, der diesmal nicht beim INDEKS Verlag sondern bei North-Holland 1986 herausgebracht werden wird. Die Arbeiten zur Definitionstheorie referiert hatte, befaßte sich diesmal mit der "klassifikatorischen Leistung

CSLI-SOMMERINSTITUT "LOGIK, LANGUAGE, COMPUTATION"

Vom 8. bis 20. Juli fand am Center for the Study of Language and Information (CSLI) der Stanford University ein Sommerinstitut über Logic, Language, and Computation statt, das von J. Barwise, S. Feferman, D. Israel und W. Marsh mit Unterstützung der Association for Symbolic Logic (ASL) organisiert wurde. Es wurden Kurse angeboten von J. Barwise über "Situation Semantics", M. van Emden über "Prolog", G. Plotkin über "Denotational Semantics", D. MacQueen über "Types and mathematical Logic", W. Maass über "Complexity Theory", J. Meseguer über "Abstract Data Types", Y. Moshovakis über "The Theory of

Algorithms", L. Moss über "Generalized Quantifiers", B. Smith über "LISP", und R. Tomason über "Foundations of Intensional Logic".

Während des Sommerinstituts gab es drei spezielle Symposien über "Types in the Study of Computer and Natural Languages" (mit Chierchia, Feferman, MacQueen, Partee), über "Role of Logic in Artificial Intelligence" (Israel, McCarthy, Rosenschein) und über "Possible Worlds" (Perry, Stalnaker).

Überlassen von der Arbeitsstelle für Semiotik

EINDRÜCKE VON DER IJCAI - 85

M. Schröder, B. Hellingrath

Forschungsstelle für KI und WI
 Universität Hamburg
 D-2000 Hamburg 13

Fraunhofer Institut für Transporttechnik
 und Wahrendistribution
 D-4600 Dortmund 50

Drei beherrschende Eindrücke gleich vorneweg:

1. Die IJCAI-85 war riesengross (ueber 6000 (!) Teilnehmer).
2. Die Industrie ist voll in den KI-Markt eingestiegen.
3. Die neuen wissenschaftlichen Erkenntnisse hielten sich (erwartungsgemaess) in Grenzen und standen damit in krassem Gegensatz zu den Hochglanzprospekten der kommerziellen Anbieter.

Die IJCAI-85 (International Joint Conference on Artificial Intelligence) fand vom 18.-23. Aug 85 in Los Angeles auf dem Campus der University of California Los Angeles (UCLA) statt. Die ersten beiden Tage waren 14 Tutorials und der Eroeffnung (Montagabend) vorbehalten, das wissenschaftliche Programm dauerte von Dienstag bis Freitag nebst begleitender Industrieausstellung (Mo-Fr). Das Campus der UCLA stellte mit unerschöpflich vielen Vortragssaalen, weiten Plaetzen und seiner mediterranen Atmosphaere einen idealen Rahmen fuer die Konferenz dar. Auch das kalifornische Wetter wurde seinem Ruf gerecht.

Fuer das wissenschaftliche Programm waren 726 Papiere eingereicht worden, von denen 245 akzeptiert wurden (fast 300 Begutachter). Die Sessions waren in folgende Themengebiete unterteilt:

Vortr. Sessions

Expert Systems	28	6
Theorem Proving	13	4
Natural Language	28	7
Perception	24	5
Knowledge Representation	21	4
Automated Reasoning	22	5
Learning and Knowl. Acquis.	31	6
Planning and Search	20	4
AI Architectures and Lang.	8	2
Logic Programming	12	2

AI and Education	6	1
Automatic Programming	7	2
Robotics	11	3
Cognitive Modelling	12	2
Philosophical Foundations	2	1

Dazu kamen noch 14 Panels, so dass regelmaessig 8 Sessions parallel liefen. Dass dieser Bericht einen sehr zufaelligen und subjektiven Eindruck darstellt, braucht wohl nicht besonders betont zu werden. Gemaess den Interessen der Autoren sollen die Gebiete Expert Systems (BH) und Natural Language (MS) etwas naeher beleuchtet werden.

Zunaechst zu Expert Systems: Die Probleme im Bereich Expert Systems sind im wesentlichen noch die gleichen wie vor 4 Jahren. Es ist bisher nicht gelungen, das Anwendungsgebiet von XPS durch Verfeinerungen und kleine Erweiterungen auf wesentlich komplexere Problemstellungen auszuweiten. Gut loesen lassen sich Diagnoseprobleme, die nicht unbedingt tiefes Wissen der dahinterliegenden Prozesse benoetigen. Dieses Wissen ist immer noch nicht in ausreichendem Masse repraesentierbar. Die etablierten und bisher benutzten Techniken sind zu schwach fuer sehr komplexe Probleme. Es gibt Versuche mit neuen Techniken wie Qualitative Simulation und Qualitative Reasoning, die aber bisher noch keinen Durchbruch geschafft haben. Wesentliche Fortschritte werden entscheidend von neuen Methoden der Wissensrepraesentation und des Reasoning abhaengen.

Die Vortraege zu XPSs hatten meist Verfeinerungen bestehender Techniken (z.B. Kontrollstrukturen,

Konsistenzueberpruefungen, Uncertain Reasoning) und bessere Unterstuetzung fuer den Benutzer zum Inhalt. Zwei der interessantesten Vortraege seien hier erwaehnt:
 "SEEK2: A Generalized Approach to Automatic Knowledge Base Refinement" von A.Ginsberg, S.Weiss (Rutgers) und P.Politakis (DEC). Vorgestellt wurde

ein Ansatz zur Verfeinerung von Wissensbasen, bei dem durch Hinzufügen, Löschen und Ändern von Regeln einer KB die Performanz des Systems erhöht werden soll. Diese Verfeinerung ist mit SEEK2 ohne Interaktion mit einem Experten möglich. Heuristische Suchalgorithmen und Metawissen über die Suche werden eingesetzt, um das Experimentieren mit verschiedenen Konzepten innerhalb der KB zu ermöglichen.

"Explainable (and Maintainable) Expert Systems" von R. Neches, W.R. Swartout, J. Moore (USC/ISI). Hier ging es um neue Entwicklungsmethoden für Expertensysteme. Durch neue Methoden sollen vor allem die Erklärungsfähigkeiten und die Wartung wesentlich verbessert werden. Durch "explanation strategies" sind erheblich detailliertere Aussagen über die Arbeit und das Wissen des Systems möglich. Wissen über den Aufbau, die Terminologie, die problem-solving Strategien etc liegt ebenfalls deklarativ in der Wissensbasis vor.

Nun zu Natural Language: In diesem Gebiet lässt sich ähnliches konstatieren wie bei XPSs: Die Entwicklungstendenz geht vor allem in die Breite, d.h. bestehende Techniken werden oft angewendet, verknüpft und weiter verfeinert. In den Vorträgen wurde die zigste Abwandlung einer Parsingmethode vorgestellt und ohne Ende disambiguiert. Vom Versuch einer integrierenden Theorie war nichts zu spüren. Etliche Vortragende mussten in der Diskussion zugeben, ihre neue Methode nur an einem oder sehr wenigen Beispielen entwickelt zu haben. So liess denn auch Barbara Grosz bei ihrem Special Topics Talk durchblicken, dass es seit ihrem letzten Talk 1979 keine gravierenden Veränderungen gegeben habe. Sie stellte fest, dass es eine leichte Verschiebung weg von der Betrachtung einzelner Äusserungen hin zu einem Verständnis für den Zusammenhang sprachlicher Äusserungen gegeben habe. Fortschritte seien auf den Gebieten beliefs, goals and intentions, generation und discourse structure zu verzeichnen. Weitere Impulse schienen auch am ehesten von der "weiblichen Intuition" mit Kathleen McCoy, Kathleen McKeown, Barbara Grosz und Candace Sidner auszugehen.

Zwei der interessantesten Vorträge waren:

"Tense, Aspect and the Cognitive Representation of Time" von Kenneth Man-kam Yip (MIT). Ausgehend von einer "computational theory of temporal representation" (nach James

Allen) und einer "formal linguistic theory of tense" (nach Norbert Hornstein) untersuchte Yip "important relationships between abstract constraints governing linguistic behavior and a computational scheme to reason about temporal relationships". Ein Kernpunkt sind die Beziehungen zwischen S (moment of speech), R (reference point) und E (moment of event). Heftige Diskussionen löste Yips Behauptung aus, seine Ergebnisse, am Englischen entwickelt, hätten für alle Sprachen Gültigkeit.

"An Efficient Context-free Parsing Algorithm For Natural Languages" von Masaru Tomita (CMU). Tomita stellte einen erweiterten LR Parsingalgorithmus vor, der seine Stärken besonders bei natürlichsprachlichen Grammatiken zeigt. Er soll dann angeblich 5 - 10 mal (!) schneller sein als der Earley-Algorithmus.

Noch ein Wort zu Knowledge Representation: "An Essential Hybrid Reasoning System: Knowledge and Symbol Level Accounts of KRYPTON" von R. Brachman, V. Gilbert und H. Levesque. KRYPTON ist eine Weiterentwicklung von KL-ONE. Dieser Vortrag verdiente die volle Aufmerksamkeit der immens zahlreichen Zuhörerenschaft, hatte man doch andeutungsweise das höchst seltene Gefühl, hier könne sich ein qualitativer Durchbruch andeuten. Das war sicherlich auch der gekonnten Vortragsweise von Ron Brachman zu verdanken. Der Schlüssel zu wesentlichen Fortschritten in vielen KI-Gebieten liegt bei der Wissensrepräsentation. Ist KRYPTON die Antwort?

Preise wurden auch wieder einige verliehen. Die IJCAI hat einen neuen Preis aus der Taufe gehoben, den "IJCAI Award for Research Excellence to honour sustained excellence in AI Research". Erster Preisträger war (natürlich) John McCarthy. Seine Festrede geriet allerdings mehr zu einer Plauderei aus dem Nähkästchen.

Die Computers and Thought Lecture wurde von Hector Levesque (Toronto) gehalten. Sein hervorragender Vortrag ("Making Believers Out of Computers") wusste die Zuhörer zu fesseln. Leider nicht im Tagungsband erschienen.

Der MIT Press Award für den besten Vortrag ging bei der Eröffnungsveranstaltung etwas unter. Er wurde Robert Fagin und Joseph Halpern vom IBM San Jose Research Center zugesprochen. Leider verlor das Programmkomitee bei der Verleihung kein Wort darüber, nach welchen Kriterien dieser Preis vergeben wurde und warum gerade

dieser Vortrag ihn erhielt. Kein grosser Anreiz fuer die Zukunft.

Vier Programmpfeiler in der Form von Invited Talks zogen sich durch das Programm:

"Discovery and Reasoning in Mathematics" (Alan Bundy),
 "Building a Bridge Between AI and Robotics" (Hirochika Inoue),
 "Self-Knowledge and Self-Representation" (John Perry),
 "Understanding and Automating Algorithm Design" (Elaine Kant).

Interessant verlief das Panel "Expert Systems: How Far Can They Go?", das in einem 'ausverkauften' riesigen Theatersaal stattfand. Terry Winograd warnte mit Blick auf die Industrieausstellung vor einem 'sell-off' der Artificial Intelligence. Er hat damit wohl vielen Wissenschaftlern aus dem Herzen gesprochen. Den Gegenpol der Diskussion bildete Randall Davis, der dem Einsatz von XPS in weiten Bereichen sehr positiv gegenueberstand. Weitere Diskussionspartner: Stuart Dreyfus aus Berkeley ("Before you ask: No, I'm his brother") und Brian Smith (Xerox PARC).

Zur Industrieausstellung: Die Ausstellung war keineswegs nur ein kleines Anhaengsel zum wissenschaftlichen Programm, sondern nahm im Bild der Tagung einen bedeutenden Platz ein (Wenn der Trend anhelt, wird das Vortragsprogramm als Ergaenzung zur Ausstellung stattfinden). Der Tagungsteilnehmer wurde von einer Flut von Hochglanzmaterialien ueberschwemmt, die Innovation war schier unerschoepflich. Das Angebot der ausstellenden Firmen wurde von drei Bereichen gepraeagt: Hardware, Programmiersprachen und XPS Tools. Hardware: Allen voran natuerlich die 'KI-Maschinen' von Symbolics, Xerox, Texas, LMI, aber auch DEC, SUN, Apollo etc mischen kraeftig mit (Die Aufzaehlung ist nicht vollstaendig, sondern gibt nur subjektive Eindruecke wieder). Der Trend allgemein: kleiner, schneller, billiger. Ein Ende ist noch nicht in Sicht. Auch gute Systemsoftware wird angeboten, z.B. ein beeindruckender 3-D Editor von Symbolics. In Sachen Farb-Graphik und Miniaturisierung ist noch einiges zu erwarten. Erstaunlicherweise waren auch Macintosh und IBM-PC stark vertreten. Programmiersprachen: Fuer LISP scheint man sich wirklich auf Common-LISP geeinigt zu haben, es wird inzwischen ueberall angeboten.

Bei Prolog ist das Bild eher diffus: viele neue Interpreter kommen auf den Markt, alle beziehen sich auf Clocksin/Mellish, alle sind im Detail etwas verschieden und alle sind schneller als alle anderen. Erwaehnungswert noch Smalltalk.

Tools: Spezial-Programmiersprachen wie OPS5 und LOOPS sowie ganze Knowledge Engineering-Umgebungen wie KEE, ART, ESE, Knowledge Craft praegen das Bild. Hier erhoffen sich die Anbieter das erste grosse Geschaeft mit 'echten' KI-Produkten. Im NL-Bereich ist das Angebot noch sehr duenn (z.B. Language Craft von der Carnegie Group).

Ein weiterer Trend ist sichtbar: hin zu PCs! Beispielsweise wurden fuer den Macintosh angeboten: (mindestens) zwei LISPs, Prolog, OPS5 und ein Systemprogramm zur Kopplung von Mac und Symbolics. Slogan: Make a Symbolics out of your Macintosh! Auf dem IBM-PC laeuft eine ganze Expertensystemumgebung (ESE).

Die Industrieausstellung nahm Ausmasse an, die wohl niemand vorher fuer moeglich gehalten hatte. Der Ausstellungsraum auf dem Campus reichte bei weitem nicht aus, so dass viele Firmen in einigen Hotels wie dem Holiday Inn oder dem noblen Beverly Hilton Suiten und Zimmer angemietet hatten. Intellicorp versorgte seine potentiellen Kunden in einer riesigen Suite im Erdgeschoss mit kostenlosem Eiscreme, waehrend LMI seine Interessenten in einer Nobelkarosse mit Bar, Fernseher und PC zwischen Campus und Beverly Hilton hin und her kutschierte. Das gab der IJCAI ein bisher nicht gekanntes Flair. Weniger finanzstarke Firmen hatten es da schon schwerer, ihre Interessenten in den 3. oder 4. Stock des Hotels zu locken.

Weniger angenehm war die Beobachtung, dass die Industrie offensichtlich mit der Ausstrahlung des "Zauberwortes" 'Artificial Intelligence' auf Kundenfang geht. Man braucht einem Produkt nur den Stempel 'AI' aufdruecken, und schon geht alles wie von selbst. So war es fast woertlich zu lesen. (Ganz witzig noch der beliebte Button von LMI: "Better Artificial Intelligence than none!".) Hier besteht die grosse Gefahr, dass beim Kunden auch mit guten Produkten ueberhoehnte Erwartungen geweckt werden, die diese Produkte einfach nicht halten koennen. Es bleibt nur zu hoffen, dass sich die grossen Werbesprueche nicht eines Tages als Boomerang erweisen und der ganze KI-Markt zu einem Flop wird. Die Industrie taete

gut daran, die Einsatzmöglichkeiten und Fähigkeiten ihrer Produkte etwas realistischer darzustellen und Artificial Intelligence nicht als Wunderwaffe zu preisen. In diese Richtung ging auch die Warnung von Terry Winograd.

Die grosse Teilnehmerzahl stellte auch die Funktion der Konferenz als 'Kontaktecke' in Frage. Bei Zuhörerzahlen von 100 - 500 (geschätzt) pro Session war es kaum möglich, jemanden 'gezielt

zufällig' zu treffen, da man so viele Namensschilder gar nicht lesen kann. Wenn man einen Vortragenden nicht direkt nach seinem Vortrag erwischte, war es wenig wahrscheinlich, ihn jemals wiederzutreffen. Die kalifornische Sonne hat jedenfalls nicht enttäuscht.

Fazit: siehe oben.

Artificial Intelligence, where are you going?

L.A.U.T. - SYMPOSIUM, TRIER

Das L.A.U.T.-Symposium, das vom 25. bis 28. März 1985 an der Universität Trier stattfand, war die zehnte in einer Reihe von Konferenzen, zu denen Prof. R. Dirven jeweils bekannte Linguisten eingeladen hatte. In einer Folge von acht jeweils zweistündigen Vorlesungen stellte R.W. Langacker von der University of California in San Diego seine neueren Untersuchungen zur kognitiven Grammatik vor. So kann das diesjährige Symposium als Fortführung des 1983 veranstalteten betrachtet werden, auf dem George Lakoff seine Arbeiten zu einer experimentell abgesicherten linguistischen Theorie vorstellte, die in weiten Teilen auf ähnlichen Annahmen basieren, wie die Positionen Ronald Langackers. Eine zunehmende Zahl von Linguisten (unter denen die früheren Vertreter der Generativen Semantik - Langacker, Lakoff und Fillmore - sind) hat in den letzten Jahren versucht eine Reorientierung der linguistischen Forschung auf ein kognitionsbasiertes Modell zu erreichen. Der grundlegende Zug dieses Modells ist seine semantische Orientierung, ein Gesichtspunkt, der in den verschiedenen Hinsichten weiter spezifiziert wird. (1) Große Aufmerksamkeit schenkt Lakoff dem Problem der konzeptuellen Struktur. In Übereinstimmung mit dem prototypischen Modell der konzeptuellen Struktur, wie es in der Psycholinguistik von Eleanor Rosch vertreten wird, werden kognitive Kategorien als nicht-rigide Entitäten mit flexiblen Grenzen aufgefaßt, innerhalb derer ein mehr oder weniger zentral gelagerter Kern von

peripheren Momenter der Kategorie umgeben ist. (2) Es wird keine explizite Unterscheidung zwischen semantischen und enzyklopädischem Wissen vorgenommen; die konzeptuellen Kategorien und Strukturen, wie sie durch sprachliche Ausdrücke repräsentiert werden, konstituieren keine autonome Ebene semantischen Wissens, sondern Teil eines größeren kognitiven Systems. Aus dieser Position folgt, daß die Linguistik ein experimentelles Fundament hat: Es kann bestimmt werden, wie die semantischen Strukturen auf unsere menschlichen Erfahrungen basieren; Lakoffs und Johnsons Forschungen über die Metapher sind in dieser Hinsicht exemplarische Versuche. Darüber hinaus wird deutlich, daß die Linguistik keine autonome Disziplin ist: Sprache ist ein symbolisches System, dessen Beherrschung integriert ist in einen größeren Bereich symbolischer und anderer kognitiver Fähigkeiten; in diesem Zusammenhang muß auch Sprache untersucht werden. (3) Die Syntaxtheorie wird durch diese eben beschriebene Konzeption der Semantik beeinflusst. Syntaktische wie konzeptuelle Kategorien werden als prototypisch organisiert gedacht, das heißt, daß syntaktische Phänomene gradierbar bzw. nur partiell motiviert sein könne. Auch die Syntax der natürlichen Sprach kann nicht als autonome untersucht werden, da sie nur einen Aspekt des sprachlichen Zeichensystems darstellt, das wiederum in das gesamte kognitive System eingebettet ist. - In der sich gerade entwickelnden Tradition der kogniti-

ven Grammatik, wie es gerade entwickelt worden ist, sind Langackers Arbeiten wichtige Beiträge, sowohl was ihre Gegenstände als auch was die Detailanalyse angeht. Mit größter Sensibilität in der Wahrnehmung von Details und einem bewundernswerten Sinn für systematische Darstellung hat Langacker ein theoretisches Netzwerk entwickelt, innerhalb dessen viele wichtige linguistische Fragen neu formuliert und erhellert werden können.- Während des Langacker-Symposiums in Trier wurde eine größere Zahl weiterer Vorträge gehalten. Es sprachen U. Claudi und W. Heine ("Metaphor and Grammar in Ewe"), D. Geeraerts ("Where does Prototypicality come from?"), J. Handke ("Pragmatic and Cognitive Aspects in the Acquisition of Adverbial Subordinate Clauses"), H. Basbol ("Two Recent Scandinavian Contributions on Aspect and Aktionsarten"), E. Steiner ("From Action through Cognition to Language

age"), T.P. Krzeszowski ("Ethic Aspects of Semantics"), C. Vandeloise ("The Preposition 'in' and the Relationship Container/Content"), M.G. de Boer ("Prototypes and Prepositions"), R.T. King ("Spatial Metaphor in German Causative Constructions"), F. Wande ("Do Finns and Swedes Experience different cognitive worlds?") C. Zelinsky-Wibbelt ("Verstehensprozesse von Metaphern"). Alle diese Vorträge werden als L.A.U.T.-Papiere veröffentlicht werden; eine größere Zahl mit einem Beitrag von R. Langacker, wird von B. Rudzka unter dem Titel "Progress in Cognitive Grammar" herausgegeben. Wichtige Kapitel aus Langackers Buch "Foundations of Cognitive Grammar" sind bereits veröffentlicht als L.A.U.T.-Papiere A99 und A100.

D. Geeraerts, Leiden
(überlassen vom Semiotischen Arbeitskreis Berlin)

JAHRESTAGUNG DER KANADISCHEN SEMIOTIK-GESELLSCHAFT

Vom 2. bis 5. Juni 1985 fand an der Université de Montreal in Quebec Kanada, die Jahrestagung der Kanadischen Semiotik-Gesellschaft statt. Sie war in vier Schwerpunkte gegliedert und wurde mit einer Podiumsdiskussion abgeschlossen. Im Schwerpunkt Semiotik des Texts gab es Vorträge über "Les instances du destinataire et du destinataire et l'articulation du sens" von Jose-Maria Nadal (Université Laval), "Pragmatics, Subjective Criticism and the Classroom" von Stephen Bonycastle (Royal Military Collage), "Pour une pragmatization de la textualite mimetique" von Marie Brisson (Queen's University), "Context and Convention: The Possibility of a Literary Pragmatics" von Marilyn Randall (Université du Quebec a Montreal), "The Semiotic Shape of Things to Come: Lautreamonts Tourbillon" von Gloria D'Ambrosio Griffith (University of Toronto), "The Relationship of Contradiction to one Thematic Structure of The Importance

of Beeing Earnest" von Deborah Du Bartell (Buffalo), "Langue, signification et sexuation: A propos de la nouvelle de Djuna Barnes, Le Valer" von Denise Marchand (Université de Quebec a Montreal), "Vers une approche de l'imaginaire quebecois" von Christiane Kegel (Université du Quebec a Montreal), "Ins and Outs" de la semantique musicale" von Roland Bermingham (Université du Quebec a Montreal) und "Semiotique musicale, taxinomie et pragmatique des techniques musicales" von Gian Franco Arlandi Como, (Italien). Die weiteren Schwerpunkte waren Theoretische Semiotik, Semiotik des Visuellen, und Theatersemiotik.

Die abschließende Podiumsdiskussion stellte die Semiotik in den Rahmen der Wissenschaftstheorie und behandelte die gemeinsamen Probleme semiotischer Analyseverfahren in den verschiedenen Wissenschaften.

(Semiotischer Arbeitskreis Berlin)

Wissenschaftliches Symposium der IBM Deutschland GmbH

NATÜRLICHE SPRACHE UND WISSENSBASIERTE SYSTEME

G. Knorz

Technische Hochschule Darmstadt
Fachbereich Informatik
D-6100 Darmstadt

Vom 27. bis 29. November 1985 fand in Bad Neuenahr ein von IBM Deutschland unter der wissenschaftlichen Leitung von Dr. O. Herzog, (IBM Stuttgart) veranstaltetes Symposium statt. Thema: NATÜRLICHE SPRACHE UND WISSENSBASIERTE SYSTEME. Teilnehmer: (im wesentlichen aus dem Inland, u.zw. aus den IBM- und Hochschul-Bereich) mit Schwerpunkt Linguistik, Sprachorientierte KI und Informatik, Logik, Philosophie, ...

Um es vorwegzunehmen: Der Besuch hat sich sehr gelohnt!

Zum Tagungsprogramm und -ablauf

Es gab insgesamt 4 halbtägige Sitzungen:

- Gegenwärtiger Stand und Überblick (Kanngießler, Habel, Uszkoreit)
- Syntaktische Analyse (McCord, Heidorn, Jensen)
- Semantische Aspekte (Wahlster, Biebel, Guenther)
- Projekte (Herzog, Lehmann, Fargues, Perrault)

Mein (globaler) Eindruck:

- Das 45-Minuten-Raster bot eine gelungene Alternative zu vielen Tagungen, bei denen immer mehr Vorträge in immer kürzerer Zeit abgewickelt werden sollen.
- Auch die für Diskussion vorgesehenen Zeiten erschienen durchaus angemessen kalkuliert.
- Die Qualität der Vorträge (Fachlich und "didaktisch") war naturgemäß differenziert, aber insgesamt auf wünschenswertem Niveau.

Bemerkungen zum Rahmenprogramm:

- Die Möglichkeit zu Kontakten "am Rande" war anhand des recht "dichten" Teilnehmerkreises besonders unterstützt
- Für das leibliche Wohl der Tagungsteilnehmer war in hervorragender Form gesorgt. Insbesondere das Dinner am Donnerstag Abend war so, wie man es von dem Einladenden gemeinhin erwartet. Prof. Schnelle, der als "dienstältester" Vertreter der Linguistik eine kleine Dankesrede hielt, erinnerte sich recht plastisch an die beiden vorausgegangenen CL-Booms. Sosehr er das in dem Symposium zum Ausdruck gebrachte neuwachte Industrie-Interesse an der Linguistik begrüßte (und mit all den anderen Teilnehmern genoß), sosehr warnte er vor überzogenen Erwartungen und der Gefahr, diese gerade im kommerziellen Bereich zu wecken.

In eigener Sache:

Ich hätte mir gewünscht, mein "Tagungs-Skript" mit etwas Ruhe und Sorgfalt zu einem mehr konzentrierten Tagungsbericht ausarbeiten zu können. Leider stand die dazu notwendige Zeit nicht zur Verfügung, weil die Endredaktion des LDV-FORUM jede freie Kapazität mühelos ausschöpft. So kommt der folgende Text im wesentlichen durch Zusammenstreichen meiner Aufzeichnungen über die einzelnen Sitzungen zustande. Selbstverständlich: alle Bewertungen stellen nichts anderes als meine subjektive Meinung dar. Es mag durchaus sein, daß manche Kritik mehr über mich, als über den kritisierten Beitrag aussagt. Dennoch mag es sein, daß dieser Bericht bei dem einen oder anderen Leser Interesse und Initiative (sich

nämlich weiter zu informieren) weckt, womit der Zweck bereits weitgehend erreicht wäre.

Zur Sitzung 1: "Gegenwärtiger Stand und Überblick"

Aus meiner Sicht war insbesondere der Vortrag von H. Uszkoreit bemerkenswert. Zunächst jedoch sprach Prof. Dr. S. Kanngießer über Grundlagen der Sprachsimulation und der Sprachinterpretation. Ein fiktiver Dialog zwischen den Kontrahenden "Zweifluß" und "McSimon" sollte zeigen, daß die Kritik von Zweifluß am Simulationsanspruch der KI nur bei Verzicht auf weitgehend allgemein akzeptierten Voraussetzungen haltbar ist. Seine Schlußfolgerungen müssen m.E. aber solange weitgehend an der Oberfläche bleiben, bis definiert ist, was eigentlich eine "Erklärung" ist. Anschließend ging Kanngießer auf die unterschiedlichen Wissensbegriffe in KI und Linguistik ein, motivierte die Notwendigkeit einer Zusammenarbeit und stellte als Exempel eigene Arbeiten zur Nominalkomposition vor.

Anschließend stellte Dr. CH. Habel in seinem Vortrag über "Referenzialität" heraus, daß Weltmodelle (Wissensbasen) Strukturen über Referenzobjekten und Beziehungen zwischen Referenzobjekten darstellen. Das Problem der "Referenz" und deren Repräsentation ist deshalb für Verstehens-, Inferenz- und Generierungsprozesse zentral. Nach einer Reihe von Beispielen für Referenzphänomene und -probleme wurden dann sogenannte "referenzielle Netze" zur Repräsentation vorgeschlagen, die in der Lage sind, heterogenes Wissen zu integrieren.

Als einen der wichtigsten Vorträge des Symposiums überhaupt schätze ich den von Dr. H. Uszkoreit ein: "New Tools for Syntactic and Semantic Analysis". Er stellte fest, daß eine Reihe zunächst unterschiedlicher neuerer Grammatikformalismen in gewissem Sinn "konvergieren", und zwar im engeren Sinn FUG (Functional Unifikation Grammar), DCG (Definite Clause Grammar), PAIR II (speziell damit hatte der Autor gearbeitet) sowie (als Vertreter theoriegebundener Formalismen), GPSG (Generalized Phrase Structure Grammar), LFG (Lexical Functional Grammar), HG (Head Grammar). Aber auch andere Formalismen lassen sich neu (und damit verwandt) interpretieren, u.a. Kategorial-Grammatik (damit hat der Autor speziell gearbeitet), Dependenz-Grammatik, Kasus-Grammatik. Zentrale Beobachtung: Die von den genannten Formalismen beschriebenen Strukturen lassen sich stets als gerichtete Graphen interpretieren. Wichtige Konsequenz: Damit läßt sich stets auch die "Graphen-Unifikation" (analog zur "Term-Unifikation" etwa in PROLOG) anwenden, die mittels Patternmatching umfassendere Graphen aus Teilgraphen compiliert. Das Prinzip der Graphen-Unifikation ist somit als effizienter und eleganter Mechanismus für Feature-Test und -Propagierung und für syntaktische und semantische Bindungen nutzbar. Bemerkenswerte Beobachtung: Die Frage nach "richtiger" Trennung in Lexikon und Grammatik löst sich weitgehend in nichts auf: es macht kaum mehr einen Unterschied aus. Das grammatische Wissen kann weitestgehend dem Lexikon zugeordnet werden, ohne daß es offensichtlich zu Redundanzen (wie z.B. bei den Word Expert Parsern) kommen muß. Mittels seines "Projektionsfolien-Computers", der Graph unification in konstanter Zeit (und höchst anschaulich) durchzuführen gestattet, demonstriert Uszkoreit sowohl das allgemeine Prinzip, als auch seine speziellen Arbeiten zur Synthese von Kategorial-Grammatik und Unifikation Grammar. Ein vorbildlicher Vortrag über eine Arbeitsrichtung, die sowohl vom theoretischen Standpunkt wie auch aus praktischer Sicht vielversprechend und wichtig scheint.

Zur Sitzung 2: Syntaktische Analyse

Ein vorgesehener Vortrag von Prof. Dr. C. Rohrer fiel aus (Deutsche Wortstellung, LFG). Bereits von der engagierten und souveränen Art der Präsentation mußte der Vortrag von Karen Jensen in dieser Sektion herausragen (Im Übrigen die einzige Referentin, leider). Während ihr Vordröner und Kollege, G.E. Heidorn, die von beiden vertretene "IBM-Tradition" der Sprachverarbeitung ebenfalls gut "verkauft", konnte ihr zweiter Kollege vom IBM Research, Yorktown Heights (zumindest mich) nicht überzeugen: Die große Linie im Vortrag fehlte. Ich wurde an einen Informatiker erinnert, der die Statements seines Programms vorstellen will.

Während ich auf den Vortrag von Dr. M.C. McCord "Syntax in modular Logic Grammars", in dem es um "Modular Logic Grammars" (MLG), also um eine Modifikation der DCG (bzw. der ursprünglichen Colmerauer'schen metamorphis grammar) ging sowie auf den Vortrag von Dr. G.E. Heidorn "PLNLP: The Programming Language for Natural Language Processing" nicht näher eingehen will, zumindest einige Bemerkungen zum Beitrag von Kareen Jensen. Sie berichtete über "The PLNLP English Grammar", eine "Breitband-Grammatik" des modernen Englischen im System PLNLP. Die Grammatik steht für verschiedene Anwendungen zur Verfügung:

- als Hilfsmittel zur Text-Überarbeitung: CRITIQUE (früher: EPISTEL).
- zur maschinellen Übersetzung.
- für Sprachsynthese.
- für die Inhaltsanalyse im Bereich Information Retrieval (bei Prof. Salton)

Ziel der Grammatikentwicklung war die Parserkonstruktion für das CRITIQUE-System (Beginn der Arbeiten: 1980). Nach 2 Jahren lag die erste Prototyp-Grammatik vor, die z.Z. etwa 70.000 Stammformen (100.000 Wortformen) umfaßt. Kompetent und engagiert untermauerte sie ihr Plädoyer

- für eine strikte Trennung von Ableitungsstruktur und berechneter Struktur,
- für Lösungen mit wenigen Regeln,
- für flache Strukturen, in denen auch Linguisten "gerne denken",
- für einen "approximate parse", der ohne kombinatorische Explosion von Regelanwendungen einen Ableitungsbaum liefert. Dieser kann dann mittels semantischer Transformationen korrekt motiviert werden.
- Gegen eine unangemessene Betonung von Semantik im strukturellen Bereich (die bekannten Beispiele semantisch unsinniger Sätze sind für Jensen durchaus "wonderful sentences"). Breite Systeme, so Jensen, erlauben kaum sinnvolle semantische Beschränkungen.

Sitzung 3: Semantische Aspekte

Was die Präsentation betrifft, war die Sitzung 3 ein echtes Kontrastprogramm: Wahlster in gewohnter Manier mit professionellen Folien und ebensolchem (Eil-)Vortrag, W. Bibel mit einem didaktischen Beitrag in (verglichen mit seinem Vorredner) Zeitlupe und dann F. Guenther mit einer für viele völlig unverständlichen Art, Zwischenrufe aus dem Zuhörerkreis und das Bemühen der Technik nach Lautsprecherverstärkung ungerührt zu ignorieren. Ich will im folgenden kurz auf den Beitrag von Prof. Dr. W. Wahlster "Forschungsstand und Perspektiven von Dialogsystemen" (oder ähnlich) eingehen, und zwar stellte Wahlster das Jahr 1985 als das Jahr der natürlichsprachigen Interfaces vor, nachdem 3 Systeme sind auf dem Markt gekommen sind:

- NL Menue (Texas Instruments, Tennant): grammatik-generierte Menues stellen eine stets korrekte Eingabe sicher; komfortable Window-Technik
- Language Craft (DANET, Carbonell)
- Q&A (Symantec, Hendrix): sehr robust und leicht vom Benutzer selbst einzurichten; allerdings Beschränkung auf nur eine Relation (das System ist übrigens den GWA-Teilnehmern aus einer Vorführung bekannt).

Anschließend ging Wahlster auf die Behandlung von parasprachlichen Zeigehandlungen ein, die in zwei seiner Projekten behandelt wird. Seine Aussage "die Zeit ist reif" darf wohl unwidersprochen akzeptiert werden. Weniger dagegen seine wenig differenzierte Einstellung gegenüber dem Problem empirischer Fundierung. Mir schien in diesem Punkt sein Bemühen um klare Statements ("förderungsgeeignet") etwas übertrieben.

Die beiden anderen Vorträge waren: Dr. W. Bibel: Inferentielle Methoden in wissensbasierten Systemen. Prof. Dr. F. Guenther: Zwei Typen von semantischen Relationen

Sitzung 4: Projekte

Diese Sitzung hinterließ, mit Ausnahme des Lehmann-Beitrages, einen schwachen Eindruck: O. Herzog stellte eine Projekt-Planung vor, die kaum über die Gliederungspunkte der Folien hinaus kommentiert wurde, und J. Fargues war einerseits durch sein "Französisch-Englisch" gehandicapt, wobei man andererseits hoffen darf, daß damit tatsächlich seine Hilflosigkeit Fragen gegenüber erklärt ist. Allgemeiner Eindruck aus Sitzung und Diskussion: am Freitag Vormittag war "die Luft raus".

Das Projekt LILOG, über das DR. O. Herzog in "LILOG - Auf dem Weg zu sprachbeherrschenden Systemen" vortrug, ist nicht nur wegen der Größenordnung von ca. 50 Wissenschaftler (oberhalb der "kritischen Masse") interessant. Wer sich die Namen der Beteiligten (u.a. Studer, Rollinger,...) anschaut, wird sicher mit großem Interesse die Entwicklung verfolgen wollen. Der Beitrag von DR. H. Lehmann "Juristisches Expertensystem auf der Grundlage von Linguistik und Logik" sprach nicht nur die Wissenschaftler, sondern auch die Autofahrer(innen) an. Das Projekt "LEX" bearbeitet nämlich, wie sich inzwischen herumgesprochen hat, im wesentlichen 3 Paragraphen der Straßenverkehrsordnung. Interessant war auch hier die Abstützung auf die Diskursrepräsentationstheorie, um die man sich tatsächlich mittlerweile näher bemühen muß, wenn man dies noch nicht getan haben sollte.

Nicht eingehen will ich hier auf J. Fargues: Semantic Information Processing using the Conceptual Graph Model.

Ausblick

"Große Namen" aus USA auf Europa-Trip: oft eine Enttäuschung. Nicht so hier. Kein unverbindliches Plaudern und aufgesetztes Selbstgespräch nach dem Motto: "Augenblick, ich muß mal schnell selbst sehen, welche Folien ich im Flugzeug eigentlich zusammengepackt habe". Stattdessen prägnante Beispiele, kompetente Kommentierungen und klare Aussagen. Es sprach:

C. R. Perrault: Semantics and Natural Language Interfaces

Nachdem er allgemeine Beschränkungen gegenwärtiger NL-Systeme systematisiert und an Beispielen offengelegt hatte, kam er zur Moral: Nach Perrault gibt es zwei Arten von Fragestellungen: eine falsche: "Was kann die Datenbank für das NLI tun?" und eine richtige: "Was kann NL für die Datenbank tun?". Seine Strategie zur Entwicklung von NLI läßt sich einfach fassen:

- (1) Man entwickle zunächst ein generelles System mit weitgefaßter Semantik. Das heißt, man implementiere die "bestmögliche" Semantik.
- (2) Dann definiere man die "Target-Language" in NL-Ausdrücken
- (3) Anschließend schränke man das NLI möglichst weit ein, um Mehrdeutigkeiten zu eliminieren und um aus den Beschränkungen allgemein Vorteil zu ziehen.

Abschließendes Resümee: Wenn es um eine Neuauflage des Symposiums geht, die sich unbedingt empfohlen hat, möchte man sicher wieder eingeladen sein, schon um zu beobachten ob sich Linguisten bis dahin generell darauf eingestellt haben, ihre Vorträge in Barocksälen zu präsentieren, wie es sich Kann gießer als neuen Standard für Linguisten wünschen wollte.

GWAI-TAGUNGSBERICHT

S. Busemann

Technische Universität Berlin
 Fachbereich Informatik
 Fachgebiet Computergestützte Informationssysteme
 Franklinstr 28/29
 D-1000 Berlin 10

Die jaehrliche Fachtagung fuer Kuenstliche Intelligenz, German Workshop on Artificial Intelligence (GWAI), fand von 23.-27. September wieder im CVJM-Heim 'Haus Solling' in Dassel statt. Mit weit ueber 200 Teilnehmern war das Haus tuechtig ueberfullt; aber ich glaube, wir traten uns nicht allzu oft auf die Zehenspitzen. Die landschaftlich schoene Lage ermunterte auch zu teilweise ausgedehnten Waldspaziergaengen, was nicht nur Politiker inspirieren soll. Das Wetter spielte ganz gut mit; am Abreisetag schliesslich lud strahlender Sonnenschein zum Bleiben ein. Unterkunft und Verpflegung stellten wieder die schon gewohnte eigenartige Mischung zwischen Jugendherberge und Hotel dar.

Das offizielle Programm umfasste die Vortraege, Vollversammlungen von Fachgruppen, Fachausschussitzungen, Systemvorfuehrungen und Projektberichte. Die persoenlichen Gespraechе in den Pausen auf der Terrasse oder abends (oft beim Einbecker Urbock) bildeten einen ganz wichtigen Aspekt der Tagung.

Das Programmkomitee hatte mehr Vortraege akzeptiert als vorgesehen, so dass nun auch hier das in meinen Augen unerfreulichste Phaenomen der internationalen Konferenzen auftrat: die Zeitnot von Vortragenden, die nicht nur an der Oberflaeche bleiben wollen, denen es aber nicht gelingt, die tieferen Gedanken in zwouelf Minuten und dann noch gemeinverstaendlich darzustellen. Unausbleibliche Folge bei den Zuoerern: nach einem halben Dutzend verschiedener Themen rauchte der Kopf, und frische Luft (s.o.) wurde lebenswichtig.

Die Themen der fuenf eingeladenen Vortraege boten einigen Zuendstoff: 'Kann die "Kuenstliche Intelligenz" Fragen der Philosophie beantworten?' von A. Beckermann (Goettingen); 'What is common sense and how to formalize it?' von J. McCarthy (Stanford); 'Logic

Programming Development' von H. Gallaire (M(n)chen); 'Assumption based truth maintenance' von J. de Kleer (Palo Alto); 'Expert systems for CAD/CAM' von J.P. Tsang (Grenoble). Das von Manfred Pinkal geleitete Tutorium der Fachgruppe Natuerlichsprachliche Systeme zum Thema Situationssemantik und Diskursrepraesentationstheorie fand ich wichtig und gelungen. Nach sehr guten Einfuehrungen in beide Gebiete von Manfred Pinkal, Michael Herweg und Uwe Reyle verspuerte ich das dringende Beduerfnis, die einschlaegigen Arbeiten erneut und diesmal richtig durchzuarbeiten. Bei Alice ter Meulens Vortrag ueber Anaphorisierung blieben fuer mich allerdings die Bezuege zum vorher Gesagten im Dunkeln. Soweit ich gehoert habe, empfanden alle die Einrichtung von Tutorien auf der GWAI als wesentliche Bereicherung.

Oft wurde der 'Workshop-Charakter' der GWAI beschworen; etwas, was jeder wollte, aber keiner exakt beschreiben mochte. Ich stelle mir darunter vor allem das Arbeiten in kleinen Gruppen (10 bis 12 Leute etwa) an einem speziellen Thema vor. Auf einer Fachtagung mit Teilnehmern unterschiedlicher Neigungen moegen die Gruppen groesser und die Themen nicht ganz so speziell sein. Mir scheint, gerade Tutorien koennen die bestmoegliche Workshop-atmosphaere schaffen.

Das Schlagwort vom interdisziplinaeren Charakter der GWAI wurde in Dassel mit Leben erfuehrt, indem diverse 'fachfremde' Vortraege, z.B. aus Philosophie und Mathematik, aufgenommen wurden. Dass sich die Betreffenden eingangs durchweg, fast entschuldigend, als Exoten bezeichneten, zeigt, dass der Kontakt zwischen der KI als abgegrenztem Gebiet und ihren zahlreichen Nachbardisziplinen keineswegs einfach ist. Vielleicht hat es doch auch Nachteile, die Interdisziplin KI als (selbstaendiges) 'Teilgebiet' oder

auch als 'Paradigma' zu bezeichnen. KI laesst sich damit zwar leichter erklaren, aber ein intensiverer Austausch mit anderen Gebieten wird dadurch eher schwieriger.

Am 23.9. fand eine Besprechung ueber die Veroeffentlichung des seit geraumer Zeit in KI-Kreisen kursierenden 'Memorandums zur Situation der KI in der Bundesrepublik Deutschland' statt. Ueber Ergebnisse und Aktionen berichtet Christa Hauenschild an anderer Stelle. Auf der Fachausschusssitzung 1.2 wurde in diesem Zusammenhang ein Arbeitskreis fuer 'Strukturfragen der KI' gegruendet mit Thomas Christaller als Sprecher.

Auf dem von Christopher Habel geleiteten Panel ueber die Zukunft der GWAI sahen die meisten ein, dass kuenftig die umstrittenen Parallelveranstaltungen nicht mehr zu vermeiden seien. Die sollten sich jedoch auf die (dann nicht mehr gar so kurzen) Kurzvortraege beschraenken, damit die fuer die

meisten Zuhoerer fruchtbareren laengeren Vortraege von allen gehoert werden koennten. Den Vorschlag, Postersessions einzufuehren, halte ich fuer eine gute Idee, ebenso die Anregung, die Moeglichkeiten zu Systemvorfuehrungen besser zu nutzen. Inhaltlich war man offenbar mit dem gebotenen breiten Spektrum zufrieden; leider gibt es mangels Foerderung von Benutzer- und Wirkungsforschung zuwenig Beitrage ueber soziale Folgen der KI. Claus Rollinger, mit der ebenso ehrenwie muehevollen Aufgabe betraut, die naechste GWAI zu organisieren (irgendwo zwischen Flensburg und Wien), lauschte mit grossen Ohren und mag manche Anregung bekommen haben. Die Qual der Wahl unter den Parallelsitzungen soll durch eine Abstractliste erleichtert werden, wenn nicht der Tagungsband bereits zu Beginn der GWAI 86 fix und fertig vorliegen sollte.

Symposium Automatic Compilation of Dictionaries.

Vom 25. bis 27. November fand in Tallin ein vom Institut für Sprache und Literatur veranstaltetes Symposium über "Automatic Compilation of Dictionaries" statt. Den Schwerpunkt bildeten Probleme der automatischen Editierung von Wörterbüchern und der Erstellung neuer Wörterbücher aus vorhandenen (gegenüber einer Erstellung von Wörterbüchern auf der Basis einer Textsammlung).

Das Symposium umfaßte eingeladene Vorträge, Übersichts-Veranstaltungen und Plenums-Diskussionen. Der Tagungsband (in Englisch) kann angefordert werden von:

KEELE JA
KIRJANDUSE INSTITUUT
200103 Tallinn, Lauristini 6

RETROSPEKTIVE RECHERCHEN (Literatur oder FuE-Projekte) zum Thema

..... DM 20,--
" 10,-- für Auszub.

STANDARDPROFIL (6x/Jahr) zum Thema

Linguistische Datenverarbeitung und Information DM 30,--/Jahr
 Informationsnetze " " "
 Betriebswirtschaftliche Fragen in IuD " " "
 Terminologiefragen " " "
 Klassifikationsforschung/Thesauruskunde " " "
 Neue Medien/Telekommunikation " " "
 Bildschirmtext/Videotext " " "
 Büroorganisation/Büroautomation " " "
 Künstliche Intelligenz/Expertensysteme " " "

INDIVIDUALPROFIL (6x/Jahr) zum Thema

..... DM 50,--/Jahr

CURRENT CONTENTS "Neue Medien" (12x/Jahr)

..... DM 30,--/Jahr

ONLINE-ANSCHLUSS AN INFODATA

..... DM 65,--/Anschaltstunde
+ DM 0,10/ausgedruckte

INFODATA. INFORMATIONEN FÜR ONLINE-BENUTZER (Stand: Aug. '85)

..... DM 10,--

DESKRIPTORENLISTE ZUM BEREICH INFORMATIONSWISSENSCHAFT UND -PRAXIS

deutsche Fassung englische Fassung DM 30,--

RECHERCHESTRATEGIE (1982-1984)

(enthält 200 Literaturnachweise) DM 25,--

ONLINE FÜR NEULINGE

(180 Literaturnachweise zur Einführung in die Praxis
des Information Retrieval) DM 25,--

... zu bestellen bei:

Gesellschaft für Information und Dokumentation mbH (GID)
 Informationszentrum für Informationswissenschaft und -praxis (GID-IZ)
 Lyoner Straße 44-48; D-6000 Frankfurt 71 (Niederrad)
 Telefon (0 69) 66 87-360; Telex 4 14 351; Telefax (0 69) 66 87-336
 Bildschirmtextteilnehmer erreichen uns mit: *69371 #

ARBEITSKREIS "MASCHINELLE SPRACHÜBERSETZUNG"

U. Klenk

Treffen vom 5.7.1985 in Saarbrücken

Das Treffen wurde von Heinz Josef Weber organisiert und geleitet. Die Programmpunkte waren:

- A. Vorträge von Mitarbeitern der Teilprojekte A2 und C des SFB 100 an der Universität Saarbrücken, mit anschließenden Vorführungen am Rechner.
- B. Vorstellung des Projekts EUROTRA-D.

Zu A.: Manfred Thiel berichtete über das Grammatiken-, Prozeß- und Strategienkonzept des MÜ-Systems SUSY II (Teilprojekt A2). Erwin Stegentritt und Johannes Ritzke stellten das System ASCOF (zur Analyse und Synthese des Französischen, Teilprojekt C) vor. In den beiden Vorträgen ging es hauptsächlich um die Analysekomponente, für welche - abgesehen von der lexikalisch-morphologischen Ebene - jeweils verschiedene Vorgehensweisen verwendet werden. SUSY II arbeitet mit einem System von Teilgrammatiken einer Sprache, die durch aufeinander abgestimmte Prozesse aktiviert werden. Den aufzubauenden Datenstrukturen liegt das Konzept der Chart zugrunde. ASCOF wendet in verschiedenen Teilphasen der Analyse kontextfreie Syntaxen, ATN sowie zur Darstellung von Weltwissen semantische Netze an. In letzter Zeit hat man sich hier auch der Synthesekomponente stärker gewidmet.

Zu B.: Das Projekt EUROTRA-D, das seit Mai 1985 am Saarbrücker Institut für Angewandte Informationswissenschaft durchgeführt wird, hat als Aufgabe, innerhalb des EG-Gesamtprojekts EUROTRA die Analyse- und Synthesekomponente des Deutschen sowie Transfer-Komponenten von und zum Deutschen zu entwickeln. Johann Haller hielt ein einführendes Referat über die Organisation von EUROTRA auf EG-Ebene und erläuterte die Funktion der einzelnen Gremien und Arbeitsgruppen. Zur MÜ-Konzeption von EUROTRA sprach Heinz-Dieter Maas: Da die nationalen Gruppen in der Analyse- und Synthesekomponente nur jeweils ihre eigene Sprache bearbeiten, ergibt sich als erste Aufgabe für ein derart angelegtes multilinguales MÜ-System, geeignete Zwischenstrukturen zu definieren, die die Ausgabe der Analyse der Quellsprache und nach Durchlaufen des Transfers die Eingabe in die Synthese der Zielsprache sind. Das hier anliegende Problem ist, sicherzustellen, daß die Zwischenstrukturen von verschiedenen Arbeitsgruppen, die sie benutzen, in gleicher Weise interpretiert werden, damit ein Satz (bzw. Text) richtig übersetzt werden kann. EUROTRA arbeitet vorwiegend mit dependentiellen Strukturen und Kasusrahmen. Die Analyse läuft in mehreren Teilphasen ab, die durch die Struktur ihrer Ausgaben definiert sind und "Ebenen" bilden. Für die Transformation der Strukturen von einer Ebene zur nächsten wurde in letzter Zeit ein neuer formaler Ansatz entwickelt, den H.-D. Maas vorstellte.

Kommentar: Zu EUROTRA stellen sich m. E. Fragen hinsichtlich der Zwischenstrukturen und des verwendeten grammatischen Ansatzes. Die Festlegung der Zwischenstrukturen scheint nur in bestimmten Bereichen genügend vorangetrieben zu sein, z.B. bzgl. der Valenzen und Kasusrahmen von Verben. Hier erscheint der dependentielle Ansatz durchaus sinn-

voll. Man fragt sich aber, wie andere Bereiche der Syntax und Semantik behandelt werden sollen, etwa komplexe Nominalphrasen mit mehreren untereinander abhängigen Modifizierern oder der Skopus von Negationsausdrücken und Quantifizierern, um nur einige wichtige Problemkreise zu nennen. Es ist die Frage, ob eine Darstellung der hier auftretenden Erscheinungen durch Merkmale, wie es wohl vorgesehen ist, eine glückliche Lösung ist. Wünschenswert wäre es, wenn im Hinblick auf die Vielsprachigkeit des Systems eine umfassendere abstrakte Interlingua definiert würde. Unter Ausklammerung des Lexikons erscheinen mir auf der Grundlage neuer Grammatiktypen (z.B. GPSG, kategoriale Ansätze), die eine geeignete semantische Komponente haben (intensionale Semantik), Bemühungen um die Konstruktion einer solchen Interlingua durchaus erfolgversprechend. Ein Ziel muß dabei sein, den interlingualen Strukturen Interpretationsregeln zuzuordnen, die eindeutige Interpretationen liefern.

Der EUROTRA-D-Gruppe, - die bereits mit Vorgaben der zentralen Gruppen arbeitet -, sei an dieser Stelle gewünscht, daß sie mit allen anstehenden Problemen gut fertig werde.

Dank an Heinz Josef Weber für Organisation und Leitung des Treffens!

ARBEITSKREIS "Maschinelle Lexikographie und Lexikologie"

G. Frackenhohl

Lughauser Str. 102
D-5064 Reivoth 1

Mit 21 Teilnehmern gut besucht war die erste Sitzung des Arbeitskreises, die am 19.11.85 im Institut für deutsche Sprache in Mannheim stattfand. Nach einem Überblicksreferat von Gert Frackenhohl, das die Teilnehmer für die Inhalte und Probleme des Fachgebietes sensibilisierte, informierten alle Anwesenden über aktuelle Arbeiten und Vorhaben (PAIN-D, EUROTRA-D, MOLEX, LEDA, COMPULEX, MADES, COTEX, PROTEXT, LEX, MKWB, Verbesserung des Retrievals, Terminologische Datenbank, Bedeutungsdarstellung, graphische Ausgabe von semantischen Beziehungen).

Die Themenschwerpunkte 'Gestaltung eines Lexikographischen Arbeitsplatzes' und 'Bedeutungsbeschreibungen in Lexika' wurden am Nachmittag ausführlicher diskutiert.

Der Arbeitskreis unterstützt die Absicht, auf der GLDV-Jahrestagung 86 außer den Vorträgen auch eine Panel-

sitzung zur praktischen Arbeit von Lexikographen zu veranstalten.

Kurzberichte und Gruppenarbeit zu speziellen Themen wurden als Vorbereitung für zukünftige Sitzungen des Arbeitskreises bestätigt, zu denen auch Interessenten aus Verlagen und Firmen eingeladen werden. Die Themenschwerpunkte der nächsten Sitzungen sind

- Bericht aus dem DIN-Ausschuß,
- Repräsentation verbaler und nominaler Valenz und
- Arbeitsplatzgestaltung.

Die nächste Zusammenkunft des Arbeitskreises findet am 7. Januar 1986 im Institut für Kommunikationswissenschaft und Phonetik der Universität Bonn, Poppelsdorfer Allée 47, von 17 Uhr statt.

Alle Interessenten sind aufgefordert an dem Treffen teilzunehmen.

**NEUE ANALYSE- UND SYNTHESEVERFAHREN
ZUR MASCHINELLEN ÜBERSETZUNG
Begleitforschung zu EUROTRA - D**

J. Kilbury, D. Metzling, U. Reyle, J. Wedekind

Projektleiter: Prof.C. Rohrer (Uni Stuttgart)
in Zusammenarbeit mit:

- TU Berlin (Prof.H.J. Schneider)
Thema: Einsatz von wissensbasierten Verfahren (Parsing und Generierung) bei der maschinellen Uebersetzung
- Uni Bielefeld (Prof.D. Metzling)
Thema: Flexible Grammatik-Komponenten-Interaktion fuer Parser des Deutschen (INTERPARS)

Projekttraeger: BMFT

Laufzeit: 1.5.85 - 30.4.87 (vorlaeufig)

Mitarbeiter: Stuttgart Prof. C. Rohrer/U. Reyle (Leitung)
U. Heid
S. Momma
K. Netter
J. Wedekind

Berlin Dr. J. Kilbury (Leitung)
S. Busemann
Dr. C. Hauenschild

Bielefeld Prof. D. Metzling (Leitung)
J. Kindermann
J. Meier

Kurzbeschreibung:

Ziel des Forschungsvorhabens ist es, die Ergebnisse im Grenzbereich von theoretischer Linguistik, Informatik und KI, die in den letzten Jahren erzielt worden sind, fuer die maschinelle Uebersetzung nutzbar zu machen. In einer ersten Phase soll anhand der auf Unifikationsmechanismen basierenden Syntaxtheorien LFG (Lexical Functional Grammar) und GPSG (Generalized Phrase Structure Grammar) untersucht werden, welche Analysemethoden und welche Repraesentationsebenen sich am besten fuer ein Uebersetzungssystem eignen.

Dabei geht die Stuttgarter Gruppe (LFG) von der Hypothese aus, dass F-Strukturen die geeignete Eingabe fuer die semantische Komponente sind, waehrend die Berliner Gruppe (GPSG) die semantischen Interpretationen ohne eine solche zusaetzliche Repraesentationsebene aufbaut. Fuer die Untersuchung wird in beiden Syntaxtheorien eine Grammatik fuer ein grosses Fragment des Deutschen entwickelt und ein Parser geschrieben.

Ueber die Untersuchung dieser beiden Formalismen hinausgehend soll erforscht werden, wie man die Interaktion zwischen Syntax und Semantik besser steuern kann, um noch effizientere Parser zu entwickeln (Schwerpunkt Bielefeld). Die Entwicklung von Transfer- und Generierungskomponenten in das Franzoesische (Stuttgart) und Englische (Berlin) wird in dieser Phase nur prototypisch sein. Erst in einer zweiten Phase werden diese Komponenten naeher erforscht und weiter ausgebaut. Es wird ferner untersucht werden, welche Art von semantischer Information fuer den Uebersetzungsprozess notwendig ist, und wie sie sich einbeziehen laesst. Aus dieser Aufgabenstellung ergeben sich in der ersten Arbeitsphase (bis April 1987) fuer die drei Forschergruppen folgende Arbeitsziele:

(a) Stuttgart

- Entwicklung einer umfangreichen LFG-Grammatik des Deutschen.
- Implementierung einer LFG-Software-Umgebung zur Weiterentwicklung des am ILR geschriebenen Compilers fuer Grammatiken des Deutschen und eventuell notwendige Erweiterungen des LFG-Formalismus.
- Auswertung von Texten des Deutschen und Franzoesischen unter kontrastiven Kriterien.
- Implementierung eines Parsers des Deutschen.
- Entwicklung eines Interpretationsmechanismus fuer die Generierung.
- Entwurf fuer die Gestaltung einer Transferkomponente Deutsch-Franzoesisch.
- Vergleich von LFG und GPSG in Hinblick auf die Interaktion und Repraesentation von Syntax und Semantik und die Generierung.

(b) Berlin

- Entwurf und Implementierung einer Software-Umgebung fuer die Entwicklung von GPSG-Grammatiken.
- Entwurf und Implementierung eines Parsers.
- Entwicklung einer Grammatik fuer ein repraesentatives Fragment des Deutschen.
- Entwurf und Implementierung eines Generators auf GPSG-Basis.
- Entwurf und Implementierung einer exemplarischen Transferkomponente Deutsch-Englisch.
- Beruecksichtigung von KI-Methoden fuer den Parser, den Generator und die Transferkomponente, besonders in Hinblick auf den Kontrollfluss und auf die Erzeugung und Manipulation von semantischen Repraesentationen.

(c) Bielefeld

- Entwicklung effektiver Verarbeitungsstrategien (Erkennung/Generierung) fuer (i) PROLOG-Unifikations-Parser und (ii) LFG- bzw. GPSG-Unifikations-Parser.
- Pruefung der Moeglichkeiten, die Analyse durch Interaktion syntaktischer, lexikalischer und semantischer Information zu steuern.
- Untersuchung der im Projektverbund verwendeten Parser.
- Effektivierung der Parser durch geeignete interaktionsgesteuerte Analysestrategien.

SONDERFORSCHUNGSBEREICH 100
UNIVERSITÄT DES SAARLANDES
370 ELEKTRONISCHE SPRACHFORSCHUNG



Teilprojekt A2

Die laufenden Arbeiten des TP A2 des SFB 100 betreffen 3 Aspekte sprachverarbeitender Systeme:

- theoretische Grundlagen inkl. Softwarekonzeption
- Realisierung eines Modells dieser Konzeption (SAFRAN)
- Implementierung eines linguistischen Modells (CAP: controlled active procedures)

Grundlagen:

Es wird unterschieden zwischen Objekt- (Wissen über die Sprache) und Metawissen (Wissen über das Objektwissen und das System selbst). Die Metawissensbasis steht sowohl dem Linguisten (dadurch wird das System selbsterklärend) als auch dem System selbst (z.B. dem Parsergenerator) zur Verfügung. Beide Wissensarten können deklarativ oder prozedural repräsentiert werden. Die Einheiten deklarativer Wissensquellen werden in Frames geschrieben, die prozeduraler in einer funktionalen Metasprache (FUSL). CAP's werden also in FUSL geschrieben. In dieses Raster lassen sich die einzelnen Wissensquellen des Modells einordnen. Methodologisch wird eine linguistische Aufgabe (z.B. maschinelle Übersetzung) nach der Problemreduktionsmethode in immer kleinere Probleme aufgeteilt.

Das Modell SAFRAN:

Die erwähnten Komponenten des Modells sind als erste Entwicklungsziele zu betrachten, d.h. weitere spezielle Wissensquellen können hinzugefügt werden.

Objektwissen

- prozedural: - Die Wissensquelle der REGELN enthält die linguistischen Regeln (Syntax und Semantik). Sie sind als Produktionen geschrieben.
- Die Wissensquelle der GRAMMATIKEN beschreibt die linguistisch motivierte Paketierung der Regeln.
- deklarativ: - Die Wissensquelle der SYMBOLE enthält die Definition der Attribut-Wert-Paare als Beschreibungen der Objekte.

Metawissen

- prozedural: - Die Wissensquelle der PROZESSE enthält die zu lösenden Teilprobleme und die Definition der Bedingungen der Organisation der Teilprobleme.
- deklarativ: - Die Wissensquelle KONTROLLSTRUKTUR enthält den Graphen, der die Zusammenhänge zwischen allen Prozessen, Grammatiken und Regeln repräsentiert.
- Die Wissensquelle METASPRACHDEFINITION enthält die Definition der Syntax und Semantik der Metasprache FUSL, in der CAP's (Prozesse, Grammatiken und Regeln) geschrieben werden können.

Dem Benutzer steht ein systemgenerierender Werkzeugsatz (EDICAP) zur Verfügung, der eine strenge Kontrolle des Zugangs zu den linguistischen Daten und deren Modifikation gestattet. Dieses Softwarepaket erlaubt auch die Spezifikation eines speziellen Systems (z.B. eines Partiellen Parsers) durch Aussagen über das vorhandene Wissen, also auf Metaebene. Der Zugang zu dem linguistischen und strategischen Wissen erfolgt ausschließlich über die Kontrollstruktur. EDICAP umfaßt folgende Komponenten:

Benutzerinformation: Z.B. an welchen Pfaden (vom Startknoten ausgehend) eine Regel hängt, alphabetische Ausgabe aller Prozesse, statistische Angaben, ...

Grapheditor: Umhängen, löschen, hinzufügen von Teilgraphen mit selbständiger paralleler Veränderung der entsprechenden CAP's.

CAP-Editor: Editieren von Regeln, Grammatiken und Prozessen.

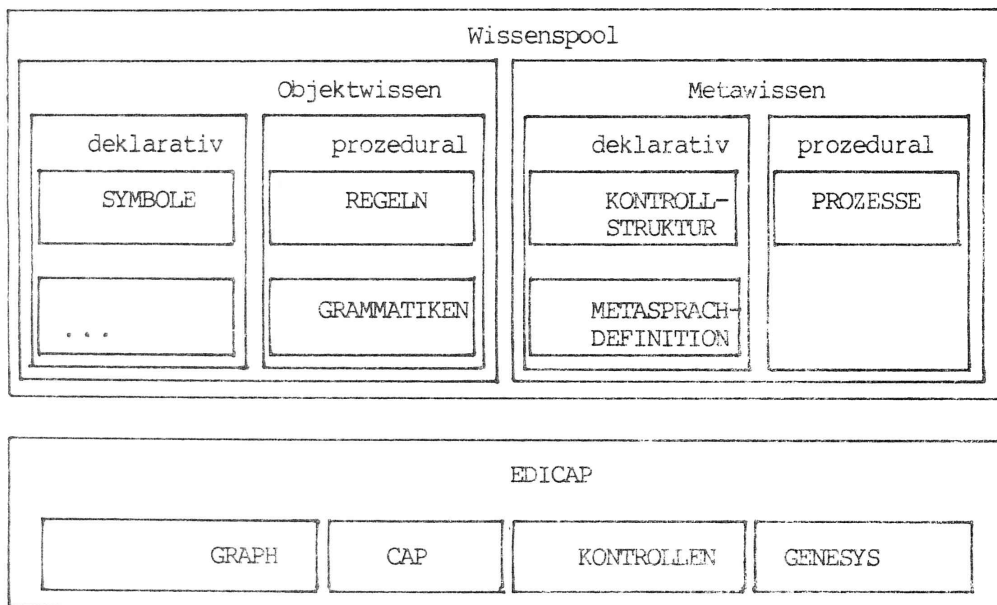
Kontrollen: Sie laufen für den Benutzer unsichtbar ab,
 - z.B. Syntax- und Semantikprüfung (unter Benutzung der Wissensquelle METASPRACHDEFINITION),
 - das Führen einer Agenda, in der noch zu leistende "Editierhandlungen" als Konsequenz aus bisherigen vermerkt sind, ...

GENESYS: Ein Softwarepaket, das CAP's unter Berücksichtigung der METASPRACHDEFINITION compiliert. Die Compilierung erfolgt ebenfalls für den Linguisten unsichtbar .

Im Rahmen von EDICAP wird das lauffähige System auch gestartet. Auf oberster Benutzerebene stehen also lediglich zwei Kommandos zur Verfügung: 'edit' und 'run'.

Zur Implementierung: Durch die Verwendung der Wissensquelle METASPRACHDEFINITION durch den Compiler für die Sprache FUSL wird gewährleistet, daß FUSL relativ leicht durch Neudefinition von (höheren) Funktionen (z.B. Komplexe Vererbungsmechanismen, Unifikation, ...) an verschiedene linguistische Theorien angepaßt werden kann. Sobald die Semantik der Funktionen definiert ist, kann sie von GENESYS verarbeitet werden.

Das folgende Diagramm gibt einen Überblick über SAFRAN:



DIE ULMER TEXTDATENBANK - PROJEKTBSCHREIBUNG**S. Goeser**Sonderforschungsbereich 129
Universität Ulm
D-7900 Ulm

Das Teilprojekt B2 ist Teil des seit 1980 geförderten Sonderforschungsbereichs 129 'Psychotherapeutische Prozesse' an der Universität Ulm. Das Projekt hat zum Ziel **"Computergestützte Methoden zur Handhabung, Beschreibung und vergleichenden Untersuchung eines psychotherapeutischen Textkorpus"** zu entwickeln und der Psychotherapieforschung zugänglich zu machen. Die derzeitige Arbeit konzentriert sich auf den weiteren Ausbau des maschinenlesbaren Textkorpus (z.Z. ca. 10 Mill. laufende Wörter, vorwiegend sog. Verbatimprotokolle) sowie des zugehörigen **Textaufnahme- und Verwaltungssystems** einschließlich einer **Worddatenbank**, sowie auf die Entwicklung einer vereinheitlichten **Methodenbank** für die Textanalyse. Soweit möglich, wird hierfür Textverarbeitungssoftware aus dem Bereich der LDV herangezogen und integriert, wobei eine modulare Systemarchitektur, das **Textbankverwaltungssystem (TBMS)**, zugrundeliegt. Nach Abschluß der Arbeiten am Textbanksystem ist für die Förderungsphase 1/86 - 12/88 die Entwicklung eines **Expertensystems** für die Subdomäne Kindheitserinnerungen vorgesehen, anhand dessen therapeutische Schlußfolgerungsprozesse simuliert werden sollen.

AUFBAU DER TRANSFERLEXIKA IN EUROTRA-D**J. Brustkern**
Institut für Kommunikationsforschung
und Phonetik, Bonn

Im Rahmen des EG-Projekts zur maschinellen Übersetzung (EUROTRA) werden seit dem 1.9.85 am Institut für Kommunikationsforschung und Phonetik (IKP), Bonn in Zusammenarbeit mit dem Institut für Angewandte Informationsforschung (IAI), Saarbrücken die Transferlexika für EUROTRA-D erstellt. Für die derzeit sieben EG-Amtssprachen sind insgesamt 42 Transferlexika erforderlich. Jeweils 6 Transferlexika werden von der

nationalen Sprachgruppe bearbeitet. Die deutsche Sprachgruppe (EUROTRA-D) ist in Saarbrücken am IAI eingerichtet. Das in Bonn laufende Teilprojekt erstellt die Transferlexika: deutsch - griechisch, deutsch - niederländisch, deutsch - dänisch und deutsch - italienisch. Durch kontrastive und experimentelle Untersuchungen sollen die inhaltlichen Anforderungen an die Transferlexika bestimmt werden. Im Anschluß daran werden die Lexika in der festgelegten Struktur für ein Testkorpus mit ca. 2.500 verschiedenen Wortformen aufgebaut. Bei dem Testkorpus handelt es sich um den Beschluß des EG-Rats zum ESPRIT-Programm. Das Bonner Teilprojekt besteht aus 1 ¹/₂ wiss. Mitarbeitern und 2 stud. Hilfskräften und läuft zusammen mit der in Saarbrücken abgewickelten 1.Hauptphase von EUROTRA-D im April 1987 aus. EUROTRA-D wird anteilmäßig von der EG und durch das BMFT finanziert. Projektträger ist die GID.

- Veranstaltungen**
- 17.-19.1: Darmstadt
 Arbeitstagung (Gesellschaft für Klassifikation), "Begriffsanalyse"
 Prof. Wille, Technische Hochschule Darmstadt, Fachbereich
 Mathematik, Karolinenplatz 5
- 8.-10.1.86 Paris
 Business Applications of Approximate Reasoning.
 Martine Bachman 25, Bd de l'Amiral Bruix-75016 Paris- France
- 13.-15.1.86 St. Petersburg, Fla.
 Thirteenth Annual ACM Sioact-Sigplan Symposium on Principals
 of Programming Languages.
 C. Wasilewski. Computer Science Dept., University of Wis-
 consin-Madison, Madison WI 53706
- 4.- 6.2.86 Cincinnati, Ohio
 CSC 86: ACM Computer Science Conference.
 Lawrence A. Jehn, Computer Science Dept., University of
 Dayton, Dayton, OH 45469. Dayton, OH 45435.
- 6.- 8.2.86 Las Cruces, N.M.
 Workshop on the Foundation of Artificial Intelligence.
 Derek Partridge, Computing Research Laboratory, New Mexico
 State University, Las Cruces, NM 88003.
- 17.-21.2.86 Christchurch, New Zealand
 International Conference on Future Advances in Computing (a
 Workshop on Alternative Technologies for Artificial Intelli-
 gence).
 John G. Cleary. Dept. of Computer Science. The University of
 Calgary, Alberta, Canada
- 19.-21.2.86 Sankt Augustin
 Erster Workshop der GI-FG 1.1.2 (Expertensysteme): Anforder-
 ungen an wirkliche Expertensystem-Entwicklungsumgebungen.
 H.-W. Hein, FG Expertensysteme, GMD Schloß Birlinghoven, 5205
 Sankt Augustin
- 25.-27.2.86 Göttingen
 11. Jahrestagung der Gesellschaft für Linguistische Datenver-
 arbeitung (GLDV): Linguistische Datenverarbeitung und Geis-
 teswissenschaften.
 M. Thaller, Max-Planck-Institut für Geschichte, Herrmann-
 Föge-Weg 11, D-3400 Göttingen
- 26.-28.2.86 Phoenix, Ariz.
 Fifth Annual Phoenix Conference on Computers and Communi-
 cations
 IEEE Computer Society, 1730 Massachusetts Ave., Washington,
 DC 20036-1903;
- 8.-16.3.86 Dassel/Solling
 4. Frühjahrsschule Künstliche Intelligenz.
 Frau Harms (ccHä-c/o GMD), PF 1240, D-5205 Augustin 1.

- 17.-19.3.86 Saarbrücken
ESOP 86: European Symposium on Programming. Bernard Robinet,
Laboratoire Informatique Theoretique et Programmation, Uni-
versite
Paris VI, 4, Place Jussieu, F-75230 Paris Cedex 05, France.
- 17.-21.3.86 Baden-Baden
2nd International Conference on the Application of Micro-Com-
puters in Information, Documentation and Libraries (CAMCI).
Deutsche Gesellschaft für Dokumentation e.V.(DGD) Westend-
str.19, 6000 Frankfurt/Main 1
- 19.-22.3.86 Monterey, Calif.
Conference of Theoretical Aspects of Reasoning About Know-
ledges.
J. Halpern, IBM Research, K51/281, 5600 Cottle Rd., San Jose,
CA 95193.
- 24.-26.3.86 Houston, Texas
OAC 86 : Office Automation Conference
OAC 86, AFIPS, 1899 Preston White Dr., Reston, VA 22091
- 31.3.-4.4.86 Orlando, Fla
Third Annual Conference on Applications of Artificial Intelli-
gence.
John F. Gilmore, Artificial Intelligence Branch, Georgia Tech
Research Institute, Atlanta, GA 30332.
- 1.- 4.4.86 Charleston, S.C.
First International Conference on Expert Database Systems.
Larry Kerschberg, Program Chairman, College of Business
Administration, University of South Carolina, Columbia, SC
29208.
- 1.- 4.4.86 WIEN
8th European Meeting on Cybernetics and Systems Research.
R. Trappl, Department of Medical Cybernetics and Artificial
Intelligence, University of Vienna, Freyung 6/2 A-1090 Wien
- 2.-4.4.86 Chicago
Conference on Management and Information Technologies
Gordan B. Davis, School of Management, University of Minne-
sota, 271 19th Ave. South, Minneapolis, MN 55455
- 6.-10.4.86 Miami, Fla.
IEEE Infocom 86: Fifth Annual Joint Conference of the IEEE
Computer and Communication Societies.
Bezalel Gavish, P.O.Box 639, Silver Spring, MD 20901.
- 16.-18.4.86 Nottingham, England
International Conference on Text Processing and Document
Manipulation Peter R. King, University of Manitoba, Dep. of
Computer Science, Winnipeg, Manitoba R3T 2N2
- 13.-17.4.86 Boston, Mass.
CHI 86: Conference on Human Factors in Computing.
Weitere Information in: C.ACM 28(1985)5
- 28.-30.4.86 München
NTG-Fachtagung: Speech Communication.
Prof. Dr.-Ing. E. Poulens, Institut für Nachrichtentechnik,
TU Braunschweig, Schleinitzstr. 23, 3300 Braunschweig

- 20.-23.5.86 Enschede, The Netherlands
Eurit 86: European Conference on Information Technology in Education.
Pieter Sinnighe Damste, c/o Twente University of Technology, C.O.J., P.O.Box 217, NL-7500 AE Enschede, The Netherlands.
- 26.-28.5.86 Blacksburg
1986 International Symposium on Multiple-Valued Logic
Joseph G. Tront, Virginia Polytechnic Institute and State University, Dep. of Electrical Engineering, Blacksburg, VA 24061
- 5.- 6.6.86 Bochum
Kolloquium über Semiotik und Wissenschaftstheorie.
Prof.Dr.W.Koch, Englisches Seminar, Ruhr-Universität, Postfach 102148, D-4630 Bochum 1.
- 5.- 6.6.86 Washington, D.C.
Workshop on Empirical Studies of Programmers.
S.Sitharama Iyengar, Dept. of Computer Science, Louisiana State University, Baton Rouge, LA 70803-4020.
- 11.-14.6.85 New York City
Annual Meeting of the Association for Computational Linguistics (ACL).
D. Terence Langendoen, Director, 1986 LSA Linguistic Institute, CUNY Graduate Center, 33 W 42nd Street, New York, NY 10036 USA
- 18.-20.6.86 Münster/W
Die Klassifikation und ihr Umfeld. (10. Jahrestagung der Gesellschaft für Klassifikation).
Gesellschaft für Klassifikation e.V., Woogstr. 36a, D-6000 Frankfurt 50
- 20.-30.6.86 Cerisy, France
Colloque sur approches de la cognition en intelligence artificielle, sciences humaines, neurosciences.
Centre Culturel International de Cerisy la Salle, 27, rue de Boulainvilliers, F-75016 Paris.
- 18.-19.7.86 New York
International Conference on Literacy, City University.
D. Terence Langendoen, Director, 1986 LSA Linguistic Institute, CUNY Graduate Center, 33 W 42nd Street, New York, NY 10036 USA
- 23.6.-20.7.86 Chicago
8th International Summer Institute for Semiotic and Structural Studies.
Martin Müller, Department of English, Northwestern University, Evanston, IL 60201.
- 11.-27.7.86 Urbino
Semiotisches Sommerinstitut.
Christina Catani, Centro Internazionale di Semiotica e di Linguistica, Piazza del Rinascimento 7, I-61029 Urbino.
- 21.-25.7.86 London, England
Third International Conference on Logic Programming Keith Clark, Imperial College of Science and Technology, London, U.K.

- 21.-25.7.86 Brighton, U.K.
European Conference on Artificial Intelligence
Rebedict du Boulay, ECAI, The University of Sussex, Cognitive
Studies Programme, Brighton BN1 9QN, U.K.
- 25.-26.7.86 New York
Linguistic Summerschool Meeting.
D. Terence Langendoen, Director, 1986 LSA Linguistic Insti-
tute, CUNY Graduate Center, 33 W 42nd Street, New York, NY
10036 USA
- 29.7.-10.8.86 Marktobendorf/Allgäu
International Summer School.
Institut für Informatik der TU München, H. Kus, Arcisstr.21,
8000 München 2.
- 11.-15.8.86 Philadelphia, Pa.
Fifth National Conference on Artificial Intelligence.
Lorraine Cooper, AAAT, 445 Burgess, Menlo Park, CA 94025.
- 25.-29.8.86 Bonn
11th International Conference on Computational Linguistics.
Prof.Dr.W.Lenders, Institut für Kommunikationsforschung und
Phonetik, Poppelsdorfer Allee 47, D-5300 Bonn 1.
- 27.-29.8.86 Noordwijkerhout
Conference on Information Systems Assessment
Gordon Davis, Dept. of Management Sciences, School of Manage-
ment and Economics, University of Minnesota, 668 Management
and Economics, 271 19th Ave. South, Minneapolis, MN 55455;
oder Theo Bemelmans, University of Technology, Dept. of
Industrial Engineering, Eindhoven, The Netherlands
- 1.- 5.9.86 Dublin
IFIP Cogress 86: 10th World Computer Congress.
Congress Secretariat, IFIP Congress 86, 44 Northumberland
Rd., Dublin 4, Ireland.
- 2.-4.9.86 Aix-en-Provence, France
1 st International Conference on Economics and Artificial
Intelligence
AFGET, 156, Boulevard Pereire, F. 75017 Paris, France
- 8.-10.-9.86 Pisa, Italien
1986 ACM Conference on Research Development in Inforamtion
Retrieval Systems
L. Rossi Bernardi, C.N.R. P/zza A. Moro 7, Roma, Italy
- 14.-18.9.86 Montreal, Canada
43rd Conference and Congress of the International Federation
for Documentation.
43rd FID Conference and Congress, C.P. 1144, Succursale
Place Desjardins, Montreal, Quebec, Canada H5B 1B3.
- 15.-19.9.86 München
8th International Conference on Computer Communication ICC86
Conference Secretary ICC86 '86, H. Heyder, Verband Deutscher
Elektroniker (VDE) e.V., Zentralstelle Tagungen, Stresemann-
allee 15, 6000 Frankfurt/Main 70.

- 16.-18.9.86 Linz, Österreich
Die Zukunft der Informationssysteme.
Prof. Dr. Arno Schulz, Institut für Informatik, Johannes
Kepler Universität Linz, A-4040 Linz.
- 22.-26.9.86 Ottenstein bei Zwettl, Österreich
10. GWAI und 2. Österreichische AI-Tagung.
Dr. C.-R. Rollinger, IBM Deutschland GmbH, Abt. 3504 LILOG
Pf. 800880, D-7000 Stuttgart 80
- 2.-6.11.86 Dallas, Texas
ACM-IEEE/CS Fall Joint Computer Conference.
ACM-IEEE/CS FJCC, P.O. Box 639, Silver Spring, MD 20901.
- 15.-19.10.86 San Franzisko
11th Annual Meeting of the Semiotic Society of America.
Semiotic Society Secretariat, P.O. Box 10, Bloomington, IN
47402, U.S.A.
- 26.-28.11.86 Paris
3rd International Exhibit and Symposium on Artificial In-
telligence and Productivity.
SIMTEC Communications, 211, rue St. Honore, 75001 Paris,
France.
- 17.-19.2.87 St.Louis
CSC 87: ACM Computer Science Conference
Arjan DeKock, Computer Science Dept., University of Missouri-
Rolla, Rolla, MO 65401
- 23.-27.2.87 Darmstadt
Datenbanken in Büro, Technik und Wissenschaft
Prof. Dr. H.-J. Schek, Fachbereich Informatik, Technische
Hochschule Darmstadt, Karolinenplatz 5, D-6100 Darmstadt.
- 3.-11.3.86 Dallas, Texas
OAC 87: Office Automation Conference
AFIPS, 1899 Preston White Dr., Reston, VA 22091
- 1.-3.4.87 Houston, Texas
Conference on Management and Information Technologies
Conferences Dept., ACM Headquarters, 11 W. 42nd St., New
York, NY 10036
- 10.-15.8.87 Berlin (DDR)
XIV Internationaler Linguistenkongress.
Akademie der Wissenschaften der DDR, Otto-Nuschke-Straße
22/23, PF Linguistenkongreß, DDR-1086 Berlin
- Oktober 87 Essen
5. Kongress der Deutschen Gesellschaft für Semiotik (DGS).
Prof. Dr. Werner Enninger, FB 3, Universität GHS Essen,
Universitätsstr. 2, D-4300 Essen 1
- 9.-14.4.89 Perpignon
4. Kongress der Internationalen Vereinigung für Semiotik
(IASS/AIS).
Prof. Dr. Gerard Deledalle, Place de l'Eglise, F-34140
Montbazin.

11 th International Conference on Computational Linguistics

PRAGMATICS

Deadline
for submission of papers
1st Dec. 1985
send papers to:
Macoto Nagao
Dept. of Electrical Engineering
Kyoto University
Sakyo-Ku, Kyoto, 606 Japan

SYNTAX

Tutorial Program
Aug. 20.-22 1986
Wolfgang Hoepfner
Heinz Marburger
INCA e.V. P.O. Box 302781
D-2000 Hamburg 36

SEMANTICS

Invited Papers
Paper Sessions
Short Communications
Panel Discussions
Demonstrations
Hardware and book exhibition
Tutorials

**bo
nn**

COLING '86

August, 25th to 29th, 1986
Winfried Lenders, Institut für
und Phonetik der Universität Bonn, Poppelsdorfer Allee 47
D-5300 Bonn 1

for further information:
Kommunikationsforschung



GLDV-Jahrestagung

vom 25.-27.2.1986 in Göttingen

Die Vortragsanmeldungen mit Abstracts sind eingetroffen. Hiernach zu urteilen, werden wir ein vielseitiges Programm haben.

Hauptsächliche Themen: Geisteswissenschaftliche Analysesysteme (u.a. Textdatenverarbeitung)
 Automatische Analyse und Synthese natürlicher Sprache
 Maschinelle Übersetzung
 Formale Grammatiken
 Lexikographie
 Informationslinguistik

Vorläufiger Zeitplan:

Di. 25. 2.

vorm.: Begrüßungsansprachen, anschließend Vorträge
 nachm.: Vorträge
 später Nachm.
 - Abend: Round-Table-Gespräch
 Arbeitstitel: Ist die philologisch-literarische DV unter die linguistische DV subsumierbar?

Mi. 26. 2.

vorm.: Vorträge
 nachm.: Informationsveranstaltung zu "Berufsperspektiven in der LDV"
 später Nachm.: Mitgliederversammlung der GLDV

Do. 27. 2.

vorm.: Vorträge
 nachm.: Sitzungen der Arbeitskreise,
 Vorführungen am Rechner

Begleitend wird es eine Posterausstellung geben.

Verschiedene gesellschaftliche Aktivitäten sind ebenfalls geplant.

Weitere Informationen sind im Dez./Jan. nach Sitzung des Programmkomitees zu erwarten.

Ursula Klenk
 (im Org.-Kommittee)

Gesellschaft für Klassifikation e.V.

VEREINIGUNG ZUR FÖRDERUNG THEORETISCHER UND ANGEWANDTER WISSENSORDNUNG UND SYSTEMATIK



TAGUNGSANKÜNDIGUNG

und Einladung zur Anmeldung von Vorträgen für die 10. Jahrestagung der Gesellschaft

Thema: -

Die Klassifikation und ihr Umfeld

Ort: Universität Münster/W
Zeit: 18.-20. Juni 1986

Traditionsgemäß soll die Tagung ein Forum sein, auf dem Wissenschaftler und Praktiker in Vortrag und Gespräch Fragen aus dem Gesamtgebiet der begrifflichen, formalen oder numerischen Klassifikation (mit zahlreichen Randgebieten) erörtern. Vorträge über neue theoretische Ansätze sind ebenso erwünscht wie die Darstellung spezieller Anwendungen bzw. die Präsentation praktischer Erfahrungsberichte. Die Behandlung folgender Schwerpunktthemen ist vorgesehen:

Begrifflich-inhaltlicher Teil

Grundfragen der Wissensordnung und Geschichte der Sachkatalogisierung

Definitionstheorie und -methoden; ihre Anwendung auf Begriffserkenntnis und -analyse und die Erstellung von Begriffs- und Klassifikationssystemen

Untersuchung und Anwendung von quantifizierenden/messenden Methoden bei der Begriffsanalyse

Begriffsbeziehungen und grammatische, insbesondere syntaktische Theorien

Erschließung von Texten für das visuelle oder maschinelle Retrieval

Einsatz neuer Technologien für Wissensordnung und strukturierte Wissensrepräsentation

Bewertung von Klassifikationssystemen und Thesauri nach dem ihnen zugeordneten Zweck

Kompatibilität von Klassifikationssystemen und Thesauri oder Integration dieser Systeme

Unterricht und Ausbildung im Fach Klassifikation und Indexierung, auch Einsatzmöglichkeiten des Computers im Fach "Inhaltliche Erschließung"

Zukunftsperspektiven klassifikatorischer/wissensordnender Tätigkeit

Mathematisch-numerischer Teil

Hierarchische und nichthierarchische Klassifikationsverfahren

Analyse und Aggregation diskreter Strukturen

Vergleich und Bewertung von Klassifikationen

Mustererkennung und Diskriminationsverfahren

Multivariate Methoden zur Datenanalyse, z.B. Skalierungsmethoden, Korrespondenzanalyse

Phylogenetische Stammbäume und DNA-Analyse

Graphische Darstellung von Strukturen und Klassifikationsergebnissen

Probabilistische und statistische Modelle

Rechenprogramme und Software-Aspekte

Praktische Anwendungen, z.B. in Biologie, Information Retrieval und Dokumentation, Psychologie, Wirtschaftswissenschaften und Marketing etc.

Das zehnjährige Bestehen der Gesellschaft für Klassifikation ist Anlaß, nicht nur die bisherige Arbeit der Gesellschaft für Klassifikation in ihren Jahrestagungen und in ihrer Gremienarbeit zu reflektieren, sondern auch zu versuchen, das Gesamtgebiet auf die gegenwärtige Situation klassifikatorischer Tätigkeit und Forschung sowie deren Zukunftsperspektiven zu beziehen.

Die Gesellschaft lädt hiermit freundlich zur Anmeldung von Vorträgen ein. Senden Sie bitte eine Zusammenfassung Ihres geplanten Vortrags (1 Seite) bis zum 15. Januar 1986 an die untenstehende Anschrift. Die Vorträge werden nach üblicher Begutachtung in einem Tagungsband veröffentlicht.

Die Teilnahmegebühr (einschließlich Tagungsband) beträgt DM 100.- für Mitglieder und DM 150.- für Nichtmitglieder. Studierende der Universität Münster haben freien Einlaß zu den Vorträgen.

Hier abtrennen und zurücksenden an:

Gesellschaft für Klassifikation e.V.

Wooogstr.36a
D-6000 Frankfurt 50

Name: _____

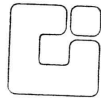
Titel: _____

Institution: _____

Adresse: _____

- Ich werde an der Tagung teilnehmen
- Ich habe mich noch nicht entschlossen, möchte aber weitere Tagungsinformationen erhalten
- Ich möchte einen Vortrag halten
- Der Titel meines Vortrages lautet: _____

- Abstrakt ist beigelegt
- Abstrakt wird bis zum 15. Januar 1986 zugesandt

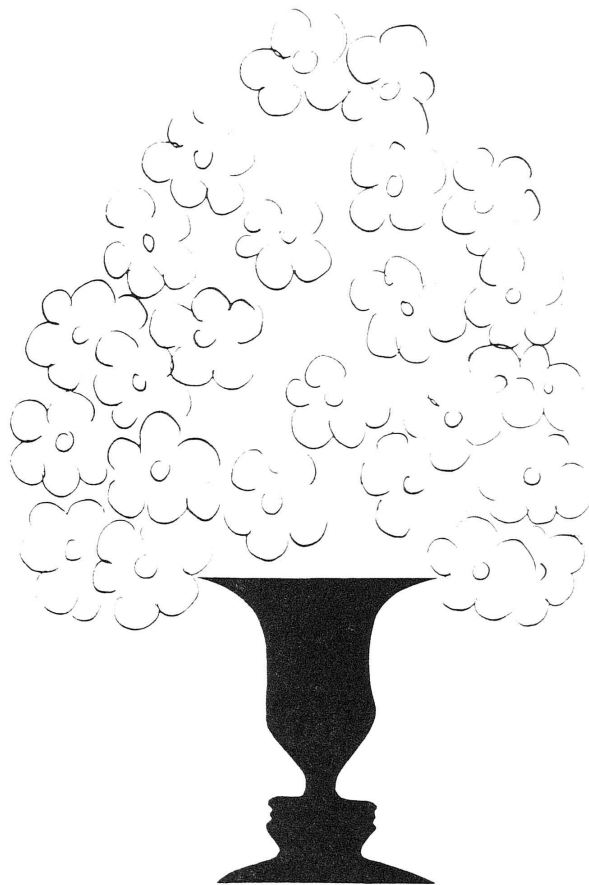


10. GWAI 1986

und



2. Österreichische
AI-Tagung 1986



Die 10. Fachtagung des Fachausschusses 1.2 "Künstliche Intelligenz und Mustererkennung" im Fachbereich 1 der Gesellschaft für Informatik e.V. und die 2. Österreichische Artificial Intelligence Tagung der Österreichischen Gesellschaft für Artificial Intelligence finden gemeinsam vom

**22.-26. September 1986
in Ottenstein bei Zwettl
Niederösterreich**

statt.

Programmkomitee

Ernst Buchberger, Wien
Werner Dilger, Karlsruhe
Alfred Kobsa, Saarbrücken
Katharina Morik, Berlin
Bernd Neumann, Hamburg
Franco di Primio, St. Augustin
Johannes Retti, Wien
Herbert Stoyan, Erlangen
Robert Trappl, Wien
Harald Trost, Wien
Wolfgang Wahlster, Saarbrücken

Chairman und Vorsitzender des Programmkomitees:

Dr. C.-R. Rollinger
IBM Deutschland GmbH
Abt. LILOG 3504
Postfach 80 08 80
D-7000 Stuttgart 80
Tel. +49 / 7031 / 17-6011

Cochairman:

Dr. W. Horn
Österreichische Gesellschaft für
Artificial Intelligence
Postfach 177
A-1014 Wien
Tel. +43 / 222 / 663281-0

Es sind Beiträge aus allen Teilgebieten der KÜNSTLICHEN INTELLIGENZ und MUSTERERKENNUNG vorgesehen, z.B.:

- Theoretische und formale Grundlagen
- Problemlösen und heuristische Verfahren
- Repräsentation von Wissen
- Automatische Deduktion
- Programmverstehen, -verifikation und -synthese
- Expertensysteme
- Natürlichsprachliche Systeme
- Bildverstehen
- KI-Programmierung
- Maschinelles Lernen
- Simulation kognitiver Prozesse
- Robotik

Es sind eingeladene Hauptvorträge, referierte Fach- und Kurzvorträge, Systemvorführungen und Projektvorstellungen geplant. Die Veröffentlichung der Haupt-, Fach- und Kurzvorträge in den Informatik Fachberichten des Springer-Verlages ist vorgesehen.

Die Tagungssprachen sind Deutsch und Englisch.

Prozedur

Lang-Papiere haben eine Länge von max. 5000 Wörtern. Kurz-Papiere haben eine Länge von max. 2500 Wörtern. Bei Überschreitung der Länge wird das eingereichte Papier umgehend zurückgeschickt. Die Manuskripte sind mit einem zwei-zeiligen Zeilenabstand zu schreiben. Auf jedem Manuskript muß Name, Anschrift und Titel vermerkt sein. Das Manuskript ist in vier Exemplaren an den Vorsitzenden des Programmkomitees bis zum 15.04.86 zu schicken.

Die eingereichten Papiere werden von den verantwortlichen Programmkomiteemitgliedern rezensiert und zur Rezension weitergegeben. Die Endverantwortung trägt der Leiter des Programmkomitees.

Die Benachrichtigung über Annahme oder Ablehnung des Papiers erfolgt ab dem 15.6.86.

Von den akzeptierten Papieren muß danach durch die Autoren eine Druckvorlage erstellt werden, in die die Anmerkungen der Rezensenten eingearbeitet wurden und die bis zum 15.7.86 an den Leiter des Programmkomitees geschickt werden sein muß. Druckvorlagen, die danach eingehen können nicht mehr in den Tagungsband aufgenommen werden, da dieser zur Tagung ausgeliefert werden soll!

Einem akzeptierten Lang-Papier entspricht ein Fach-Vortrag von 45 Minuten Dauer, incl. 15 Minuten Diskussion. Einem akzeptierten Kurz-Papier entspricht ein Kurz-Vortrag von 30 Minuten Dauer, incl. 10 Minuten Diskussion.

Parallel zu dem normalen Tagungsablauf sollen Systemvorführungen und Projektvorstellungen durchgeführt werden. Letztere in Form einer Postersection. Ausgearbeitete Vorschläge sind bis zu dem 15.7.86 an den Leiter des Programmkomitees in dreifacher Ausfertigung zu senden. Das Programmkomitee behält sich die Annahme oder Ablehnung dieser Vorschläge vor.

Es besteht die Möglichkeit parallel zueinander vier halbtägige Tutorials bzw. Kurz-Workshops durchzuführen. Ausgearbeitete Vorschläge sind in dreifacher Ausfertigung bis zum 15.4.86 an den Vorsitzenden des Programmkomitees zu richten. Das Programmkomitee behält sich die Annahme oder Ablehnung dieser Vorschläge vor.

Termine

Einreichen der Kurz- bzw. Lang-Papiere in vier Exemplaren an den Vorsitzenden des Programmkomitees bis 15. April 1986.

Benachrichtigung über Annahme oder Ablehnung des Vortrages ab 15. Juni 1986.

Anmeldungen von Systemvorführungen und Projektvorstellungen bis 15. Juli 1986.

Abgabe der Druckvorlagen der akzeptierten Vorträge bis 15. Juli 1986.

Tagung in Ottenstein bei Zwettl, Niederösterreich 22.-26. September 1986.

ANMELDEFORMULAR

NAME:

ANSCHRIFT:

TELEFON:

AN
 Dr. C.-R. Rollinger
 IBM Deutschland GmbH
 Abt. 3504 LILOG
 - ÖGAI/GWAI 86 -
 Postfach 80 08 80
 D-7000 Stuttgart 80

Ich beabsichtige, an der Fachtagung teilzunehmen.

Den Tagungsbeitrag in Höhe von DM / ÖS habe ich überwiesen.

Ich beabsichtige, einen Vortrag einzureichen.

Ich bin Mitglied bei:

Meine Mitglieds-Nr. ist:

Datum, Unterschrift

Tagungsbeitrag

Bezahlt	bis		nach	
	1.8.86		1.8.86	
Mitglieder der GI (AFCET, AISE, AICA, BCS, NGI, NTG) bzw. der ÖGAI	DM	90	DM	110
	ÖS	640	ÖS	780
Nichtmitglieder	DM	130	DM	150
	ÖS	930	ÖS	1060
Studenten	DM	50	DM	70
	ÖS	360	ÖS	500

Die Anmeldung erfolgt durch Einsendung des Anmeldeformulars und durch Bezahlung des Tagungsbeitrages (in der BRD) auf das Konto GWAI-86, Sonderkonto-GI (Dr. Rollinger), Konto-Nr. 5780614 bei der Berliner Commerzbank (BLZ 100 400 00) bzw. (in Österreich) auf das Konto 004-62683 bei der Ersten Österreichischen. Nur eine frühzeitige Anmeldung garantiert Ihnen die Teilnahme.

Tagungsort

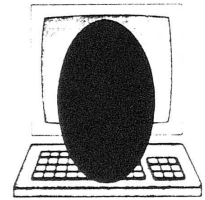
Die Tagung findet im Bungalowhotel Ottenstein statt. Ottenstein liegt in der Nähe von Zwettl in Niederösterreich. Es stehen ca 180 Betten in Zwei- und Dreibettzimmern im Hotel und ca. 100 Betten in Achtbettzimmern im Jugendgästehaus zur Verfügung. Damit ist die Teilnehmerzahl auf 280 Personen beschränkt. Die Vollpension beträgt ca. ÖS 340,-/DM 49,- pro Tag im Hotel und ca. ÖS 200,-/DM 29,- pro Tag im Jugendgästehaus. Zur An-/Abreise wird ein Busservice von St.Pölten zur Verfügung stehen. Außerdem wird eine kostengünstige Anreisemöglichkeit insbesondere für Studenten aus der BRD organisiert werden.

Industrierausstellung

Für eine Industrierausstellung steht begrenzt Platz zur Verfügung. Interessenten wenden sich bitte an Dr. Johannes Retti, Österreichische Gesellschaft für Artificial Intelligence, Postfach 177, A-1014 Wien, Tel.: +43-222-7293-5030.

SECOND INTERNATIONAL CONFERENCE
ON THE APPLICATION OF
MICRO-COMPUTERS IN
INFORMATION, DOCUMENTATION AND LIBRARIES

BADEN-BADEN, FEDERAL REPUBLIC OF GERMANY · MARCH 17 TO 21, 1986



PRESSEMITTEILUNG

4/85

Das Programm der

Second International Conference on the Application of Micro-Computers in Information, Documentation and Libraries,

die vom 17. bis 21. März 1986 im Kongresshaus Baden-Baden stattfinden wird, liegt nun vor.

Der Ministerpräsident des Landes Baden-Württemberg, Herr Dr. Lothar Späth, und Herr Dr. Hans-Peter Geh, Präsident der IFLA, haben bereits Ihr Erscheinen zugesagt.

In 20 Veranstaltungen werden mehr als 100 Vorträge gehalten. Die Themen reichen von der Anwendung in Bibliotheken und Informationszentren bestimmter Fachgebiete bis hin zu Expertensystemen und elektronischem Publizieren.

Besondere Aufmerksamkeit wird der Behandlung von

**Komplexen Systemen,
Verbundsystemen,
Arbeitsplatzrechnern und
Bildplatten als Informationsträger
in Verbindung mit Micro-Computern**

gewidmet sein.

Im Rahmen des Kongresses wird ein gesonderter Workshop angeboten zum Thema

Integrierung von Datenbank Management Systemen auf Micro-Computern in Bibliotheken.

Eine umfangreiche Fachausstellung mit internationalen Angeboten ergänzt die Konferenz und Fachbesichtigungen zu Informationszentren und Bibliotheken nach Freiburg, Heidelberg, Karlsruhe, Stuttgart und Wehr bieten zusätzliche Informationen.

Weitere Auskünfte erhalten Sie bei:

**Sekretariat Deutsche Gesellschaft
für Dokumentation e.V.
Westendstrasse 19
D-6000 Frankfurt am Main 1
Telefon: (069) 747761**

Diese Seite enthält "Mitteilungen aus der GLDV", die erst unmittelbar vor Drucklegung eingingen. Die "reguläre" Rubrik findet sich ab Seite 157.

Neuaufnahmen der GLDV:

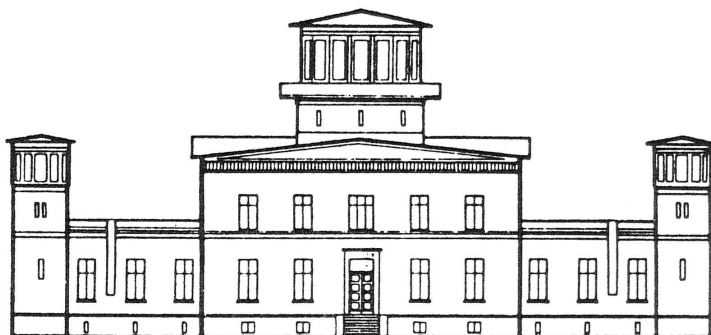
Elumenthal, Andreas M.A./Germanistisches Seminar der Universität
Heidelberg/Hauptstr. 207-209/6900 Heidelberg
Dr. Boas, Hans Ulrich/Kattenbühl 16/3510 Hann.-Münden
Dr. Fröhlich, Gerhard/LG Soziologie und empirische Sozialforschung
der Universität Hannover/FBE I/Bismarckstr. 2/3000 Hannover
Dr. Hacker, Robert/Kreuzgasse 31/213/A-1180 Wien
Dr. Hoepelman, Jakob Ph./Fraunhofer Gesellschaft IAO/Holzgarten-
str. 17/ 7000 Stuttgart
Kreyhs, Jutta/Porta-Nigra-Platz 6/5500 Trier
Kunta, Tamerlan George/Am Mühlbach 9/8100 Garmisch-Partenkirchen
Luckhardt, Heinz-Dirk/Petrusstr. 12/6602 Dudweiler
Müller, Albert/Im Bachlfeld 5/A-4040 Linz
Dr. Panyr, Jiri/Siemens AG, Abt. ZTI INFL/Otto-Hahn-Ring 6/
8000 München 83
Dr. Pütz, Horst P./Germanistisches Seminar der Universität
Kiel/Olshausenstr. 40/2300 Kiel 1
Rieck, Harald/Fichtenweg 12/206/7400 Tübingen
Ripp, Volker/Birkenstr. 12, 1000 Berlin 21
Schwanke, Martina/Germanistisches Seminar der Universität Kiel/
Olshausenstr. 40/2300 Kiel 1
Steffens, Petra M. Phil./Taunusstr. 72/7030 Böblingen
Dr. Wegstein, Werner/Franz-Liszt-Str. 16/8700 Würzburg
Wesche, Birgit M.A./Schönauerstr. 12a/7000 Stuttgart 80
Zäch, Olivier/Dufourstr. 24/CH-9008 St. Gallen
Dr. Zock, Michael/1, rue A. Guilpin/F-94250 Gentilly

Austritt aus der GLDV:

Dr. Ehlers, J./Postfach 809/7000 Stuttgart

Anschriftenänderungen

Dr. Johann Haller/FR 5.5 Informationswissenschaft/Universität des Saarlandes/
Postfach/6600 Saarbrücken 11
Prof. Dr. Klaus Mudersbach/Gerbodoweg 17/6900 Heidelberg
Franc Wagner/Postfach 106151/6900 Heidelberg
Hartmut Warlich/Ringstr. 44/1000 Berlin 45
Christa Womser-Hacker/Weihmühlstr. 6/8052 Moosburg a.d.Isar



Institut für Kommunikationsforschung und Phonetik (IKP) der Universität Bonn

Poppelsdorfer Allee 47, 5300 Bonn 1, Telefon (02 28) 73 56 38

Abt.: Linguistische Datenverarbeitung - Leiter: Prof. Dr. Winfried Lenders

ACHTUNG - ACHTUNG - ACHTUNG - ACHTUNG - ACHTUNG

Der IKP-Arbeitsbericht (Abt. LDV) Nr. 6 ist erschienen.

Er enthält folgende Beiträge:

G.Willée: Das Bonner Lexikonsystem BONNLEX1

G.Heinze-J.Brustkern: Das Lexikonsystem ALEXSYS

Der Bericht kann angefordert werden bei

G. Willée

Institut für Kommunikationsforschung
und Phonetik der Universität Bonn
Poppelsdorfer Allee 47

5300 Bonn 1

IKP - Arbeitsberichte

Abteilung LDV

Neu erschienen im Buske-Verlag:

Christine Schneider

Automatische Indexierung und Syntexanalyse.

Zur Entwicklung sprachanalytischer Komponenten von Informationssystemen auf empirischer Grundlage.

Grundlage der Arbeit bildet ein methodisches Konzept, das die Evaluierung von bestehenden (Modell-)Systemen als Forschungsinstrument zur Weiterentwicklung auf dem Gebiet des Information Retrieval versteht. Während in der Arbeit nur ein Teilaspekt zum Bereich Nominalgruppenanalyse evaluiert werden konnte, liegt inzwischen die umfassendere Realisierung des Evaluierungskonzepts in dem Projekt PADOK (Test und Vergleich von Texterschließungssystemen für das Deutsche Patent- und Fachinformationssystem) vor. Die Arbeit wurde mit dem Kulturpreis Ostbayern ausgezeichnet.

RMM

Regensburger Microfiche Materialien

Texte und Arbeitsmittel zur Sprach- und Literaturwissenschaft

Herausgeber: Herbert E. Brekle, Ludwig Hitzenberger,
Jürgen Krause, Klaus Traede, Manfred Wacht
in Verbindung mit: Hans-Werner Eroms, Ernst Heitsch,
Manfred Hellmann, Klaus Matzel, Helmut Rix

Bestimmte wissenschaftliche Arbeitsergebnisse und Materialien sprengen oft allein aufgrund ihres Umfanges den Rahmen herkömmlicher Verlagsprogramme. Sie bleiben somit trotz ihrer Bedeutung unzugänglich. Hier schließt die Mikrofilm-Publikation eine Lücke. Z.B. finden Wortregister zu Texten deutscher Gegenwartssprache bei einem vorausgesetzten Corpus von rund einer Million laufender Text-Wörter, nach herkömmlicher Publikationsart etwa 100.000 Seiten, auf 400 Microfiche Platz.

In der nachfolgenden wissenschaftlichen Reihe erscheinen Arbeiten zur Sprach- und Literaturwissenschaft sowie Arbeitsmittel für Forschung und Lehre, die aufgrund ihres Inhalts, ihres Umfangs oder ihres Verwendungszwecks für die Publikationsform Microfiche besonders geeignet sind.

Bisher sind erschienen:

RMM 02

Alfred Hoppe, Harald Zimmermann, Ludwig Hitzenberger

LIMAS CORPUS

500 Texte mit zusammen 1,4 Mio. fortlaufender Wörter aus allen Bereichen der deutschen Gegenwartssprache.

Teil a) Klartexte, alphabetische Indices

(vorläufig und rückläufig, Frequenzranglisten)

74 Fiches mit Erläuterungen

Teil b) KWOC-Konkordanz mit Satzkontext, 293 Fiches

RMM 03

Jürgen Krause, Peter Rosenbeck, Harald Zimmermann

DONALD

Dokumentation zu Forschungsprojekten zur Computer-Linguistik zusammengestellt aus Projektanzeigen aus Computers and the Humanities u.a.

Autoren-, Schlagwort-, Sprachen- und Corporakataloge, 3 Fiches

RMM 04

Manfred Wacht

Lemmatisierter Index zu Vergil mit statistischen Anhängen zu Sprache und Metrik, 5 Fiches

RMM 05

Manfred Wacht

Lemmatisierter Index zu Lukan mit statistischen Anhängen zu Sprache und Metrik, 3 Fiches

RMM 06

Margot Schmidt

Lemmatisierte Indices, Konkordanzen und Sonderlisten zum mittelhochdeutschen Text "Die sibben strassen zu got" von Rudolf von Biberach, 13 Fiches

RMM 07

Manfred W. Hellmann

BONNER ZEITUNGSCORPUS

Konkordanz zum Bonner Zeitungscorpus. Zusammengesetzt aus Tageszeitungen der DDR und der Bundesrepublik Deutschland. Zeitlicher Umfang: 1949-1974.

KWIC-Konkordanzen, 427 Fiches

RMM 08

Wolfgang Schneider

Materialien zur deutschen Rechtschreibung

Textsammlung von Probearbeiten, Diktaten und Schulbüchern, alphabetische Indices, Frequenzranglisten, Konkordanzen, 12 Fiches

RMM 09

Manfred Wacht

Lemmatisierter Index zu den "Carmina maiora" Claudians mit statistischen Anhängen zu Sprache und Metrik, 3 Fiches

RMM 10

Manfred Wacht

Lemmatisierter Index zu Silius Italicus mit statistischen Anhängen zu Sprache und Metrik, 3 Fiches

RMM 12

Herbert-Ernst Brekle (Hrsg.)

Nominalkomposita,

Kommunikative und pragmatisch-semantische Bedingungen der Aktualgenese, der Verwendung und des Verstehens von Nominalkomposita (im Deutschen), 5 Fiches

RMM 13

Walter von Hahn

HAM-RPM

Projektdokumentation des Hamburger Rede-Partner-Modells zur künstlichen Intelligenz. Arbeitspapiere, Programm, Demonstrationsbeispiele, Daten, 8 Fiches

MCS - Verlagsabteilung
Postfach 31 54
8500 Nürnberg 80

Die folgende Übersicht über aktuelle Literatur im Bereich LDV und etwas allgemeiner "Mensch-Maschine-Kommunikation" (insbesondere "sprachorientierte KI-Forschung" setzt die in Heft 1/85 begonnene Zusammenarbeit mit dem GID-IZ fort. Es besteht inzwischen die "bereits praktizierte" Vereinbarung, daß die Redaktion des LDV-FORUM den laufenden einschlägigen Profildienst des GID-IZ maschinenlesbar zur Verfügung gestellt bekommt. Nicht zuletzt um den etwa 2-3 cm dicken Stapel von Referenzen auf wenige Seiten zu reduzieren, schließt sich dann eine Selektion nach formalen und (subjektiv) inhaltlichen Kriterien an. Außerdem wird die vollständige Dokumentbeschreibung (wie anschließend demonstriert) auf wenige Angaben komprimiert.

Im einzelnen kann die Literatur-Übersicht wie folgt charakterisiert werden:

- (1) Quelle: Datenbasis INFODATA des GID-IZ
- (2) Auswahl: Standardprofil "Linguistische Datenverarbeitung und Information" mit anschließender Selektion. Insbesondere sind Beiträge über Expertensysteme ohne spezielle Behandlung der Benutzerschnittstelle weitgehend nicht übernommen, im Gegensatz zu Beiträgen über Wissensrepräsentation (Zugegebenermaßen nur schwer objektivierbar)
- (3) Sprache: ausschließlich Deutsch/Englisch (Selektion)

Selbstverständlich können die zitierten Arbeiten bei der GID-IZ bestellt werden!

Beispiel für eine "Original-Dokumentationseinheit" aus dem SDI-Profil:

Copyright: INFODATA. GID-Informationszentrum, Frankfurt/M.

In: UC1 84-0909.1

Bergeijk, D. van <International Translations Centre, Delft, NL>
The language as information barrier

FID, Den Haag, NL: FID congress 42: The use of information in a changing world, Den Haag, NL: 1984.09.24-1984.09.27

In: Laan, A. van der (Hrsg.); FID, Den Haag, NL: The use of information in a changing world. Proceedings of the forty-second FID congress held in The Hague, The Netherlands, 24-27 September, 1984. Amsterdam, NL: North Holland (1984) S. 399-407, 9 Lit., en = FID 631, ISBN 0-444-87554-9

The language barrier is an obstacle in the free flow of information. For years on end it has exerted a major influence on man's society. The application of modern techniques in the translating process is making headway. Translations, with or without the aid of computers, are indispensable means to follow the research activities carried out in the world. Preparing translations is and will be a costly and a time consuming matter. Even with the use of MT. Therefore existing translations should be used again. The international translations centre, together with CNRS, has developed a database on existing translations. This database is a unique tool to search for translations. Short- and long term projects are outlined. (Autor)

Übersetzungswesen; Datenbank; International; Sprachbarriere;
Informationsfluss

Z 8504 0785

Beispiel für das Äquivalent der LDV-FORUM-Literaturübersicht:

Bergeijk, D.:

The language as information barrier

In: Laan, A. van der (Hrsg.); FID, Den Haag, NL: The use of information in a changing world. Proceedings of the forty-second FID congress held in The Hague, The Netherlands, 24-27 September, 1984. Amsterdam, NL: North Holland (1984) S. 399-407

Ref.-Nr.: Z 8504 0785

140

Zeigler, B. P.; Rada, R.:

Abstraction in methodology.

A framework for computer support

Information processing and management, Oxford, GB 20(1984)1/2, S. 63-79

Systemforschung; Modell; Simulation; Software; Kuenstliche_Intelligenz;
Kostenbewertung Rechnergestuetztes Problemloesen

Ref.-Nr.: Z 8504 0848

Flood, B. (Hrsg.); Witiak, J. (Hrsg.); Hogan, T. H. (Hrsg.); Cooper, L. (Bearb.); Heller, P. (Bearb.); Salz, A. (Bearb.):

Nineteen hundred and eighty four: Challenges to an information society

Proceedings of the 47th annual meeting, Philadelphia, Pennsylvania, October 21-25, 1984. White Plains, NY, US: Knowledge Industry Publ. (1984)

Informationswissenschaft; Gesellschaft; Systemforschung; Informationstechnologie; Informationsverhalten; Mensch-Maschine-Kommunikation; Expertensystem; Entwicklungstendenz

Ref.-Nr.: Z 8504 0940

mit folgenden Beiträgen (u.a.):

Roberts, C. G.:

The wariness of the work force. Growing concerns with office automation causes push for safeguards through collective bargaining. S. 13-15.

Bueroorganisation; Mensch-Maschine-Kommunikation; Humanfaktor

Ref.-Nr.: Z 8504 0943

Obermeier, K. K.; Cooper, L. E.:

Information network facility organizing system (INFOS). An expert system for information retrieval. S. 95-98.

Expertensystem; Systemforschung; Entwurf; Informationswissenschaft

Ref.-Nr.: Z 8504 0955

Smith, P. J.; Chignell, M.:

Development of an expert system to aid in searches of the Chemical Abstracts. S. 99-102.

Expertensystem; Mensch-Maschine-Kommunikation; Information Retrieval; Pharmakologie; Chemie

Ref.-Nr.: Z 8504 0956

Walker, G.; Janes, J. W.:

Expert systems as search intermediaries. S. 103-105.

Expertensystem; Mensch-Maschine-Kommunikation; Systemumgebung; Information Retrieval; Chemie; Medizin; Geologie; Informatik

Ref.-Nr.: Z 8504 0957

Clark, W. B.:

Supplantation of cognitive processes as a user aid in online systems. S. 106-109.

Systemforschung; Mensch-Maschine-Kommunikation; Humanfaktor; Recherche; Modell; Algorithmus; Test

Ref.-Nr.: Z 8504 0958

Tolle, J. E.:

Monitoring usage of online information systems NLM CATLINE database. S. 126-129.

Benutzerforschung; Informationsverhalten; Mensch-Maschine-Kommunikation; Information_Retrieval; Medizin; Empirische_Untersuchung

Ref.-Nr.: Z 8504 0962

Fidel, R.:

Request-related criteria for the selection of search keys. S. 141-143.

Information_Retrieval; Mensch-Maschine-Kommunikation; Recherchestrategie; Informationsverhalten

Ref.-Nr.: Z 8504 0965

Cole, E.; Trauth, E. M.:

User friendly and the user. Applying the information system model to the design of user interfaces. S. 148-152.

Informationssystem; Modell; Mensch-Maschine-Kommunikation; Benutzerfreundlich; Systemforschung

Ref.-Nr.: Z 8504 0967

Algon, J.; Peterson, A. C.; Tao, A.

PUBS: An integrated word processing/data processing system to manage the rapidly expanding volume of technical manuscripts submitted for in-house clearance. S. 179-182.

Dialogsystem; Textverarbeitung; Literaturdokumentation; Minicomputer; Software

Ref.-Nr.: Z 8504 0972

Palmquist, R. A.; Eisenberg, M.:

Testing a text analysis metric using magnitude estimation. S. 231-236.

Informationswissenschaft; Mensch-Maschine-Kommunikation; Informationsverhalten; Messung; Methode; Kommunikationsprozess

Ref.-Nr.: Z 8504 0979

Bock, H. H.:

Anwendungen der Klassifikation: Datenanalyse und numerische Klassifikation.

Proceedings, 8. Jahrestagung der Gesellschaft fuer Klassifikation, Hofgeismar, 10.-13. April 1984, Teil 2, Frankfurt, DE: Indeks (1984)

Klassifikationsforschung; Klassifikationsmethode; Mathematische_Methode; Clustering; Systemforschung; Datenanalyse; Maschinelle_Klassifikation; Information_Retrieval

Ref.-Nr.: Z 8504 0980

Poetzsch, E.:

Faktografische Informationssysteme. Probleme der Systemgestaltung und der Anwendung von Informationssprachen.

Informatik, Berlin, DD 31(1984)4, S. 20-23.

Informationssystem; Fakten; Faktendokumentation; Automation; Thesaurus; Sprache; Informationsnetz; DDR

Ref.-Nr.: Z 8504 1241

Dierich, E.:

Untersuchungen zur Formalisierung des Referierens

Informatik, Berlin, DD 31(1984)4, S. 23-25

Textanalyse; Maschinelles_Referieren; Maschinelles_Indexierungsverfahren; Test; Bauwesen

Powers, D. M. W.:

Natural language the natural way

Computer compacts, Amsterdam, NL 2(1984)3/4, S. 100-109.

Sprache; Kuenstliche_Intelligenz; Psychologie; Linguistik; Grammatik; Wortschatz; Test; Spracherwerb

Ref.-Nr.: Z 8504 1258

Wilensky, R.; Arens, Y.; Chin, D.:

Talking to UNIX in English. An overview of UC

Communications of the ACM, Baltimore, MD, US 27(1984)6, S. 574-593.

Mensch-Maschine-Kommunikation; Kommandosprache; Dialogsystem; Betriebssystem; Software

Ref.-Nr.: Z 8504 1285

Look, H. E.:

Monitor survey of the information industry. Natural language processing.

Monitor, Abingdon, GB (1984)44, S. 7-10.

Mensch-Maschine-Kommunikation; Sprache; Software; Kommandosprache; Bueeroorganisa-
tion; Entwicklungstendenz

Ref.-Nr.: Z 8504 1315

Moelle, K. (Bearb.):

Forschungs- und Entwicklungsprojekte in Informationswissenschaft und -praxis.

8. Ausg. Stand: Mai 1984

Muenchen, DE: Saur (1984)

Projekt; Informationswesen; Informationswissenschaft; Bibliothekswissenschaft;
Informationsdienst; Erstellung; Veroeffentlichungswesen; Linguistik; Benutzer-
forschung; Bundesrepublik_Deutschland

Ref.-Nr.: Z 8504 1319

Wolff, S.:

The use of morphosemantic regularities in the medical vocabulary for automatic
lexical coding.

Methods of information in medicine, Stuttgart, DE 23(1984)4, S. 195-203.

Textanalyse; Terminologie; Woerterbuecherstellung; Maschinell; Englisch; Medi-
zin

Ref.-Nr.: Z 8504 1332

Wittmann, A.:

Planungen zum Aufbau des deutschen Patent- und Fachinformationssystems. Re-
cherchieren in deutscher Sprache.

In: Deutsche Gesellschaft fuer Dokumentation (DGD) Vereinigung fuer Informa-
tionswissenschaft und -praxis, Frankfurt, DE: 6. Fruehjahrstagung der Online-
Benutzergruppe der DGD in Neu-Isenburg. Vortraege, 8. bis 10. Mai 1984.
Frankfurt, DE (1984) S. 44-58

Patentdokumentation; Sprachbarriere; Datenbank; Deutsch; Graphische Datenver-
arbeitung; Volltextdatenbank; Elektronisches_Publizieren; Pilotanwendung; Bun-
desrepublik_Deutschland

Ref.-Nr.: Z 8504 1346

Johnson, D.; Bryant, B. R.:

Formal syntax methods for natural language.

Information processing letters, Amsterdam, NL 19(1984)3, S. 135-143

Textanalyse; Syntaktisches_Netz; Grammatik; Syntax; Methode; Kuenstliche Spra-
che; Sprache; Englisch

Ref.-Nr.: Z 8504 1366

Grishman, R.:

Natural language processing.

Journal of the American Society for Information Science, Washington, DC, US
35(1984)5, S. 291-296

Verarbeitung_Gesprochener_Sprache; Frage-Antwort-System; Textanalyse; Informa-
tion_Retrieval

Ref.-Nr.: Z 8504 1370

Sparck Jones, K.:

Proposals for R and D in intelligent knowledge based systems (IKBS).

Journal of information science, Amsterdam, NL 8(1984)4, S. 139-147.

Kuenstliche_Intelligenz; Expertensystem; Systemforschung; Forschungsplanung;
Entwurf; UK

Ref.-Nr.: Z 8504 1425

Fischer, G.:

Wie intelligent koennen und sollen Computersysteme sein. Bedeutung fuer die
Gestaltung wissensbasierter Systeme in der Bueeroautomatisierung.

Datenverarbeitung. Steuer. Wirtschaft. Recht, Muenchen, DE 13(1984)12, S.
259-264.

Kuenstliche_Intelligenz; Mensch-Maschine-Kommunikation; Bueroorganisation;
Expertensystem
Ref.-Nr.: Z 8504 1496

Lutes-Schaab, B.:
MIKROPLIS - ein Textbank-Management-System.
Nachrichten fuer Dokumentation, Muenchen, DE 35(1984)6, S. 254-259.
Software; Textverarbeitung; Datenbank; Mikrocomputer
Ref.-Nr.: Z 8504 1522

Spine, T. M.; Williges, B. H.; Maynard, J. F.:
An economical approach to modeling speech recognition accuracy.
International journal of man-machine studies, London, GB 21(1984)3, S.191-202.
Verarbeitung_Gesprochener_Sprache; Bewertung; Modell; Effektivitaet; Test
Ref.-Nr.: Z 8504 1583

Dutta, S.; Sinha, P. K.:
Pragmatic approach to subject indexing. A new concept.
Journal of the American Society for Information Science, Washington, DC, US
35(1984)6, S. 325-331.
Maschinelles_Indexierungsverfahren; Syntaktisches_Indexieren; Permutation
Ref.-Nr.: Z 8504 1637

Frailey, D. J. (Hrsg.):
National Computer Conference 1984. July 9-12, 1984, Las Vegas, Nevada.
American Federation of Information Processing Societies (AFIPS), Arlington,
VA, US: National Computer Conference, Las Vegas, NV, US: 1984.07.09-1984.07.12
Reston, VA, US: AFIPS Press (1984)
Bueroorganisation; Hardware; Mikrocomputer; Software; Softwaretechnologie;
Datenbank; Datenstruktur; Gesellschaft; Kuenstliche_Intelligenz; Datenueber-
tragung
Ref.-Nr.: Z 8504 1671

mit (u.a.) folgenden Beiträgen:

Wasserman, A. I.:
Specification and implementations of interactive information systems. S.
259-266.
Softwaretechnologie; Dialogsystem; Mensch-Maschine-Kommunikation; Benutzer-
partizipation; Prototyp; Programmiersprache; Implementierung
Ref.-Nr.: Z 8504 1675

Larson, J. A.; Wallick, J. B.:
An interface for novice and infrequent database management system users.
S. 523-529.
Datenbank; Mensch-Maschine-Kommunikation; Benutzerfreundlich; Prototyp
Ref.-Nr.: Z 8504 1679

Tennant, H.:
Menu-based natural language understanding. S. 629-635.
Mensch-Maschine-Kommunikation; Frage-Antwort-System; Suchfrage; Computer-
linguistik
Ref.-Nr.: Z 8504 1684

Ballard, B. W.; Lusth, J. C.; Tinkham, N. L.:
Transportable English-language processing for office environments. S.643-
649.
Bueroorganisation; Computerlinguistik; Englisch; Information_Retrieval.
Ref.-Nr.: Z 8504 1685

Flowers, M.; Dyer, M. G.:
Really arguing with your computer in natural language. S. 651-659.
Kuenstliche_Intelligenz; Computerlinguistik
Ref.-Nr.: Z 8504 1686

Biermann, A. W.; Gilbert, K. C.; Fineman, L. S.:
Introducing VIPS: A voice-interactive processing system for document management. S. 661-666.
Verarbeitung_Gesprochener_Sprache; Textverarbeitung; Bueroorganisation
Ref.-Nr.: Z 8504 1687

Sprowl, J.; Balasubramanian, P.; Chinwalla, T.; Evens, M.; Klawans, H.:
An expert system for drafting legal documents. S. 667-673.
Expertensystem; Textverarbeitung; Prototyp; Mikrocomputer; Jura
Ref.-Nr.: Z 8504 1688

Wittmann, A.; Schikarski, L.:
Fuer eine deutsche Patentdatenbank ist es nicht zu frueh.
Blick durch die Wirtschaft, Frankfurt, DE 27(1984)223
Datenbank; Sprachbarriere; Englisch; Kommandosprache; Volltextsuche; Deutsch; Woerterbuch; Technik; Patent; Graphische_Datenverarbeitung; Volltextdatenbank; Bundesrepublik_Deutschland; USA
Ref.-Nr.: Z 8506 1739

Oppermann, R.:
Aktive Akzeptanzunterstuetzung durch Betroffenenbeteiligung.
Office Management, Baden-Baden, DE 32(1984)11, S. 1084-1086.
Systemforschung; Mensch-Maschine-Kommunikation; Informationsverhalten; Benutzerpartizipation
Ref.-Nr.: Z 8506 1795

Wallace, M.
Communicating with database in natural language.
Chichester, GB: Horwood (1984)
Kuenstliche_Intelligenz; Sprache; Computerlinguistik; Implementierung
Ref.-Nr.: Z 8506 1820

Balzert, H.; Fauser, A.:
Hardware- und Softwaretechnologien fuer vernetzte multifunktionale Bueroarbeitsplaetze.
In: Krueckeberg, F. (Hrsg.); Schindler, S. (Hrsg.); Spaniol, V. (Hrsg.); Gesellschaft fuer Informatik (GI), Bonn, DE; Nachrichtentechnische Gesellschaft im VDE (NTG), Frankfurt, DE: Offene Multifunktionale Bueroarbeitsplaetze und Bildschirmtext. Proceedings, Berlin, 25.-29. Juni 1984. Berlin, DE: Springer (1985) S. 90-104.
Mensch-Maschine-Schnittstelle. Bueroorganisation; Mensch-Maschine-Kommunikation; Informationsnetz; Integriert, Bueroinformationssystem.
Ref.-Nr.: Z 8506 1954

im gleichen Tagungsband (u.a.):

Bullinger, H. J.; Faehnrich, K. P.; Hanne, K. H.; Ziegler, J.:
Benutzerschnittstellen an multifunktionalen Bueroarbeitsplaetzen. Schnittstellenmodelle und Evaluation.
S. 141-159.
Mensch-Maschine-Kommunikation; Bueroorganisation; Systemforschung; Schnittstelle; Modell; Bewertung
Ref.-Nr.: Z 8506 1957

Haug, H.; Mohnhaupt, M.:

Kuenstliche Intelligenz. Die Avantgarde etabliert sich
Computer Magazin und Software Magazin, Stuttgart, DE 13(1984)10, S. 20-36.
Kuenstliche_Intelligenz; Mensch-Maschine-Kommunikation; Expertensystem; Ent-
wicklungstendenz
Ref.-Nr.: Z 8506 2062

Wieland, U.

Micro-computers based integration of computerized translation aids.
Terminologie bulletin, Luxembourg, LU 21(1984)46, S. 19-25.
Mikrocomputer; Verteiltes_System; Hardware; Software; Mensch-Maschine-Kommuni-
kation; Anforderung; Textverarbeitung; Uebersetzungswesen; Maschinelle_Ueber-
setzung; Datei; Terminologie Uebersetzungshilfe
Ref.-Nr.: Z 8506 2094

Luckhardt, H. D.:

Erste Ueberlegungen zur Verwendung des Sublanguage-Konzepts in SUSY.
Multilingua, Amsterdam, NL 3(1984)3, S. 135-142.
Maschinelle Uebersetzung; Textanalyse; Software; Entwicklungstendenz
Ref.-Nr.: Z 8506 2106

Vickery, A.:

An intelligent interface for online interaction.
Journal of information science, Amsterdam, NL 9(1984)1, S. 7-18.
Mensch-Maschine-Kommunikation; Software; Benutzerbedarf; Benutzerfreundlich;
Kuenstliche_Intelligenz; Expertensystem
Ref.-Nr.: Z 8506 2108

Ofori-Dwumfuo, G. O.:

Using a cognitive model of dialogue for reference retrieval.
Journal of information science, Amsterdam, NL 9(1984)1, S. 19-28.
Mensch-Maschine-Kommunikation; Recherchestrategie; Humanfaktor; Modell; Soft-
ware; Test; Gewichtung; Clustering
Ref.-Nr.: Z 8506 2109

Otten, K. W.:

Information technology and the potential for office automation and management
support. A systems perspective.
The journal of information and image management, Silver Spring, MD, US
17(1984)2, S. 13-20.
Bueroorganisation; Informationstechnologie; Integriert; Mensch-Maschine-Kommuni-
kation; Benutzerbedarf; Entwicklungstendenz
Ref.-Nr.: Z 8506 2110

Salvendy, G. (Hrsg.)

Human-computer interaction. Proceedings of the first U.S.A.-Japan conference,
Honolulu, Hawaii, August 18-20, 1984. Amsterdam, NL: Elsevier (1984)
Mensch-Maschine-Kommunikation; Humanfaktor; Kuenstliche_Intelligenz; Experten-
system; Arbeitswissenschaften; Arbeitszufriedenheit; Bueroorganisation; Ar-
beitsplatzgestaltung; Software; Benutzer; Befragung; USA; Japan
Ref.-Nr.: Z 8506 2120

mit u.a. folgenden Beiträgen:

Kobayashi, K.:

Man-machine interaction in 'C and C' age.
Mensch-Maschine-Kommunikation; Humanfaktor; Benutzerfreundlich; Hardware;
Software
Ref.-Nr.: Z 8506 2121

- Norman, D. A.:
Cognitive engineering principles in the design of human-computer interfaces.
Mensch-Maschine-Komm.; Benutzer; Benutzerforschung; Softwaretechnologie
Ref.-Nr.: Z 8506 2122
- Bullinger, H. J.; Faehnrich, K. P.:
Symbiotic man-computer interfaces and the user assistant concept.
Mensch-Maschine-Kommunikation; Frage-Antwort-System; Kuenstliche_Intelligenz; Humanfaktor
Ref.-Nr.: Z 8506 2123
- Badre, A. N.:
Designing transitionality into the user-computer interface.
Mensch-Maschine-Kommunikation; Informationssystem; MIS; Benutzer
Ref.-Nr.: Z 8506 2124
- Williges, R. C.:
Design of human-computer dialogues.
Mensch-Maschine-Kommunikation; Dialogsystem; Vergleich; Softwaretechnologie
Ref.-Nr.: Z 8506 2125
- Meister, D.:
New opportunities in the human factors engineering of computerized systems
Mensch-Maschine-Kommunikation; Humanfaktor; Softwaretechnologie
Ref.-Nr.: Z 8506 2126
- Roberts, T. L.:
Perspectives of a modern user- interface designer.
Mensch-Maschine-Kommunikation; Benutzer; Programmiersprache
Ref.-Nr.: Z 8506 2127
- Hirsch, R. S.:
National standards and the practice of human factors.
Mensch-Maschine-Kommunikation; Terminal; Normung; Hardware; Humanfaktor
Ref.-Nr.: Z 8506 2128
- Shneiderman, B.:
Correct, complete operations and other principles of interaction.
Mensch-Maschine-Kommunikation; Dialogsystem; Fehler; Kommandosprache; Software
Ref.-Nr.: Z 8506 2129
- Rouse, W. B.:
Design and evaluation of computer-based decision support systems.
Expertensystem; Entscheidungshilfe; Softwaretechnologie; Mensch-Maschine-Kommunikation; Kuenstliche_Intelligenz
Ref.-Nr.: Z 8506 2133
- Boose, J.:
A framework for transferring human expertise.
Expertensystem; Kuenstliche_Intelligenz; Mensch-Maschine-Kommunikation
Ref.-Nr.: Z 8506 2134
- Johansson, G.:
Computer technology. Stress and health relevant transformations of psychosocial work environments.
Mensch-Maschine-Kommunikation; Arbeitsplatzgestaltung; Bueeroorganisation; Humanfaktor
Ref.-Nr.: Z 8506 2138

Dainoff, M.:

A model for human efficiency. Relating health, comfort and performance in the automated office workstation.

Bueroorganisation; Arbeitsplatzgestaltung; Terminal; Humanfaktor; Mensch-Maschine-Kommunikation

Ref.-Nr.: Z 8506 2139

Mishra, B.; Trojan, B.; Burke, R.; Douglas, S. A.:

A quasi-natural language interface for UNIX.

Mensch-Maschine-Kommunikation; Kuenstliche_Intelligenz; Sprache; Kommando-sprache; Benutzer

Ref.-Nr.: Z 8506 2143

Sekiya, T.; Watanabe, A.; Kikuchi, M.:

Medical decision making by means of man-machine dialogue.

Mensch-Maschine-Kommunikation; Medizin; Gesundheitswesen; Befunddokumentation; Graphische_Darstellung

Ref.-Nr.: Z 8506 2145

Joost, M. G.:

Problems in vocal man-machine communication.

Mensch-Maschine-Kommunikation; Verarbeitung_Gesprochener_Sprache; Befragung; Benutzer

Ref.-Nr.: Z 8506 2147

Hahn, R.:

Expertensysteme als intelligente Informationssysteme. Konzepte fuer die funktionale Erweiterung des Information Retrieval.

Nachrichten fuer Dokumentation, Muenchen, DE 36(1985)1, S. 2-12.

Expertensystem; Information_Retrieval; Recherche; Dialogsystem; Faktendokumentation; Jura

Ref.-Nr.: Z 8506 2178

Appelrath, H. J.:

Die Erweiterung von DB- und IR-Systemen zu Wissensbasierten Systemen.

Nachrichten fuer Dokumentation, Muenchen, DE 36(1985)1, S. 13-21.

Expertensystem; Frage-Antwort-System; Datenbank; Systemumgebung; Systemforschung; Programmierung; Entwicklungstendenz Bueroinformationssystem

Ref.-Nr.: Z 8506 2179

Hast, C. D.:

Von der Industrie- zur Informationsgesellschaft. Maerkte und Konzepte der Informationsindustrie.

Geschaeftsbericht: VDRZ, Hannover, DE (1984) S. 27-71.

Informationstechnologie; EDV; Kuenstliche_Intelligenz; Informationsverhalten; Gesellschaft; Entwicklungstendenz; Prognose

Ref.-Nr.: Z 8506 2204

Online Review, Medford, NJ, US: International online information meeting 8, London. Oxford, GB: Learned Information (1984)

Dialogdienst; Wirtschaftsinformation; Medizin; Marketing; Informationsbedarf; Mensch-Maschine-Kommunikation; Mikrocomputer; Expertensystem; Literaturversorgung; Online

Ref.-Nr.: Z 8506 2206

unter anderem mit folgenden Beiträgen:

Jewitt, C.:

A subject indexing engine. S. 151-160.

Indexierungsverfahren; Rechnerunterstuetzt; Expertensystem; DK; Facettenklassifikation; Authority_File

Ref.-Nr.: Z 8506 2224

- Armstrong, C. J.:
Command languages and the intelligence factor. S. 161-169.
Information_Retrieval; Kommandosprache; Expertensystem; Mensch-Maschine-Kommunikation; Entwicklungstendenz
Ref.-Nr.: Z 8506 2225
- Burns, A.; Robinson, J.:
Tools and techniques for adaptable user interface development. S. 255-262.
Mensch-Maschine-Kommunikation; Benutzerfreundlich; Dialogsystem; Entwurf; Methode
Ref.-Nr.: Z 8506 2232
- Jones, K. P.; Bell, C. L. M.:
MORPHS. New developments in software to make MORPHS more widely available.
S. 263-273.
Dialogsystem; Software; Maschinelles_Indexierungsverfahren; Minicomputer; Eingabe; Synonym; Name; Abkuerzung; Heuristik
Ref.-Nr.: Z 8506 2233
- Plander, I. (Hrsg.):
Artificial intelligence and information-control systems of robots. Proceedings of the third international conference, Smolenice, Czechoslovakia, June 11-15, 1984 Amsterdam, NL: North-Holland (1984)
Kuenstliche_Intelligenz; Expertensystem; Mustererkennung; Frage-Antwort-Systemen-Computerlinguistik; Datenbank Robotertechnologie
Ref.-Nr.: Z 8506 2301
- u.a. mit folgenden Beiträgen:
- Aiello, L.:
Artificial intelligence and office automation. S 1-8.
Bueroorganisation; Expertensystem; Mensch-Maschine-Kommunikation; Modell
Ref.-Nr.: Z 8506 2302
- Bolc, L.; Kochut, K.; Rychlik, P.; Strzalkowski, T.:
Deductive question answering system DIALOG. S. 17-24.
Frage-Antwort-System; Deduktion; Test; Medizin
Ref.-Nr.: Z 8506 2304
- Adorni, G.; DiManzo, M.:
Adaptive natural language generation S. 77-80.
Expertensystem; Computerlinguistik; Semantisches_Netz; Mensch-Maschine-Kommunikation
Ref.-Nr.: Z 8506 2306
- Cudazzo, R.; Lesmo, L.; Randi, C.:
Interpretation of natural language queries via pattern-action rules S. 119-122.
Datenbank; Sprache; Computerlinguistik; Mensch-Maschine-Kommunikation; Mustererkennung Platzbuchungssystem
Ref.-Nr.: Z 8506 2308
- Gorodetzki, V. I.; Drozhzhin, V. V.; Yusupov, R. M.:
Application of attributed grammar and algorithmic sensitivity model for knowledge representation and estimation. S. 157-162.
Kuenstliche_Intelligenz; Mathematisches_Modell Wissensrepraesentation
Ref.-Nr.: Z 8506 2309
- Hajicova, E.; Hnatkova, M.:
Linguistically motivated representation of knowledge as a basis for inference procedures. S. 167-170.
Frage-Antwort-System; Semantisches_Netz; Inferenz; Test
Ref.-Nr.: Z 8506 2310

Helbig, H.:
 Natural language access to the data base of the AIDOS/VS information retrieval system. S. 171-174
 Information_Retrieval_System; Sprache; Kuenstliche_Intelligenz; Computer-linguistik
 Ref.-Nr.: Z 8506 2311

Khoroshevsky, V. F.:
 The ATNL-based macroprocessor. A software tool of communication modules implementation. S. 201-206.
 Mensch-Maschine-Kommunikation; Kuenstliche_Intelligenz; Software
 Ref.-Nr.: Z 8506 2313

Koch, D.:
 German language questioning of relational databases. S. 221-224.
 Datenbank; Datenstruktur; Deutsch; Sprache; Mensch-Maschine-Kommunikation; Computerlinguistik; Kuenstliche_Intelligenz
 Ref.-Nr.: Z 8506 2315

Panevova, J.:
 Natural language interface to an expert system. S. 277-280.
 Expertensystem; Computerlinguistik; Semantik; Mensch-Maschine-Kommunikation
 Ref.-Nr.: Z 8506 2317

Ruske, G.:
 Halbsilben als Verarbeitungseinheiten bei der automatischen Spracherkennung. Sprache und Datenverarbeitung, Tuebingen, DE 8(1984)1/2, S. 5-16.
 Verarbeitung_Gesprochener_Sprache; Phonetik; Syntax; Wort; Wortschatz; Mustererkennung; Methode; Deutsch
 Ref.-Nr.: Z 8506 2353

Ney, H.; Geppert, R.; Mergel, D.; Noll, A.; Piotrowski, H.; Schwartau, P.; Tomaschewski, H.:
 Statistical modelling and dynamic programming in speech recognition. Sprache und Datenverarbeitung, Tuebingen, DE 8(1984)1/2, S. 17-33.
 Verarbeitung_Gesprochener_Sprache; Mustererkennung; Phonetik; Wort; Mathematische_Methode; Programmierung; Statistik; Modell
 Ref.-Nr.: Z 8506 2354

Hoegel, H.; Lang, M.:
 Automatische Spracherkennung fuer den Dialog zwischen Mensch und Maschine. Sprache und Datenverarbeitung, Tuebingen, DE 8(1984)1/2, S. 34-43.
 Verarbeitung_Gesprochener_Sprache; Wort; Mustererkennung; Methode; Vergleich; Phonetik

Lenders, W.; Stock, D.:
 Zur Entwicklung eines akustischen Terminals zur automatischen Spracherkennung und Sprachproduktion. Sprache und Datenverarbeitung, Tuebingen, DE 8(1984)1/2, S. 44-53.
 Verarbeitung_Gesprochener_Sprache; Mustererkennung; Wort; Phonetik; Klassifikationsmethode; Transkription; Maschinell
 Ref.-Nr.: Z 8506 2356

Brietzmann, A.:
 Eine Netzwerk-Grammatik des Deutschen fuer die Automatische Spracherkennung. Sprache und Datenverarbeitung, Tuebingen, DE 8(1984)1/2, S. 54-63.
 Verarbeitung_Gesprochener_Sprache; Syntaktisches_Netz; Grammatik; Deutsch; Mustererkennung
 Ref.-Nr.: Z 8506 2357

Mudler, J.; Paulus, E.:
Entwicklung und Erprobung der erwartungsorientierten Analyse fuer die automatische Spacherkennung.
Sprache und Datenverarbeitung, Tuebingen, DE 8(1984)1/2, S. 64-71.
Verarbeitung_Gesprochener_Sprache; Phonetik; Woerterbuch; Methode; Test; Mustererkennung
Ref.-Nr.: Z 8506 2358

Smetacek, V.:
The semantic analyser. A processing tool for text data bases.
International forum on information and documentation, Moskva, SU 9(1984)1, S. 16-19.
Computerlinguistik; Textanalyse; Semantik; Woerterbuch; Volltextsuche; Kommandosprache
Ref.-Nr.: Z 8506 2388

im der gleichen Zeitschrift:

Tikhomirov, B. D.:
Some specific features of software and technology in the AMPAR and NERPA systems of machine translation. S. 9-11.
Maschinelle_Uebersetzung; Software; Englisch; Russisch; UdSSR
Ref.-Nr.: Z 8506 2395

Nagao, M.:
Outline of the machine translation project of the Japanese Government. S.12-17.
Maschinelle_Uebersetzung; Referat; Englisch; Japanisch; Software; Woerterbucherstellung; Japan
Ref.-Nr.: Z 8506 2396

Rozhdestvensky, Yu. V.; Sychev, O. A.:
General scientific lexicon in automated translation.
International forum on information and documentation, Moskva, SU 9(1984)2, S. 23-27.
Lexikologie; Terminologie; Wissenschaft_und_Technik; Maschinelle_Uebersetzung; Entwicklungstendenz; UdSSR
Ref.-Nr.: Z 8506 2398

Kiselev, A. N.:
The utilisation of the structural type-based routines as a homograph resolution method in the AMPAR system.
International forum on information and documentation, Moskva, SU 9(1984)2, S. 31-35.
Masch._Uebersetzung; Semantik; Algorithmus; Englisch; Russisch Disambiguierung
Ref.-Nr.: Z 8506 2399

Anmerkungen des Benutzers zum SDI-Dienst: US: Conference: Online '84, San Francisco, CA, (1984)
Dialogdienst; Volltextdatenbank; Bibliotheksautomation; 1 Bueroorganisation; Information_Retrieval; Mensch-Maschine-Kommunikation; Mikrocomputer; Downloading; Benutzerbedarf
Ref.-Nr.: Z 8507 2425

u.a. mit dem folgenden Beitrag:

Blanding-Clark, T.:
Designing effective user interfaces and documentation. What does 'user-friendly' really mean. S. 78-88,
Mensch-Maschine-Komm.; Programmdokumentation; Manual; Benutzerfreundlich
Ref.-Nr.: Z 8507 2430

Rollinger, C. R. (Hrsg.):
Probleme des (Text-)Verstehens. Ansätze der Kuenstlichen Intelligenz.
Tuebingen, DE: Niemeyer (1984)
Kuenstl_Intelligenz; Computerlinguistik; Linguistik; Text; Struktur; Grammatik
Ref.-Nr.: Z 8507 2446

Croft, W. B.:
The role of context and adaption in user interfaces.
Int. journal of man-machine studies, London, GB 21(1984)4, S. 283-292.
Mensch-Maschine-Kommunikation; Systemforschung; Bueroorganisation; Informa-
tion_Retrieval; Recherchestrategie; Auswahl
Ref.-Nr.: Z 8507 2494

MacCracken, D. L.:
Experience with the ZOG human-computer interface system.
Int. journal of man-machine studies, London, GB 21(1984)4, S. 293-310.
Mensch-Maschine-Kommunikation; Benutzerfuehrung; Software; Maschinelle Recher-
che; Test
Ref.-Nr.: Z 8507 2495

Williams, M. D.:
What makes RABBIT run.
Int. journal of man-machine studies, London, GB 21(1984)4, S. 333-352.
Mensch-Maschine-Kommunikation; Benutzerfuehrung; Graphische_Darstellung;
Maschinelle_Recherche; Software
Ref.-Nr.: Z 8507 2497

Rissland, E. L.:
Ingredients of intelligent user interfaces.
Int. journal of man-machine studies, London, GB 21(1984)4, S. 377-388.
Mensch-Maschine-Kommunikation; Humanfaktor; Software; Kommunikationswiss.
Ref.-Nr.: Z 8507 2500

Morik, K.:
Customers' requirements for natural language systems. Results of an inquiry.
Int. journal of man-machine studies, London, GB 21(1984)5, S. 401-414.
Mensch-Maschine-Kommunikation; Frage Antwort-System; Expertensystem; Benutzer
bedarf; Benutzerfreundlich; Befragung
Ref.-Nr.: Z 8507 2501

Janda, A. (Hrsg.):
Human factors in computing systems.
Proceedings of the CHI '83 conference held in Boston, Mass., U.S.A., 12-15
December 1983. Amsterdam, NL: North Holland (1984)
Mensch-Maschine-Kommunikation; Systemforschung; Software; Hardware
Ref.-Nr.: Z 8507 2506

im gleichen Band (u.a):

Norman, D. A.:
Design principles for human-computer interfaces. S. 1-10.
Mensch-Maschine-Kommunikation; Kommandosprache; Benutzerfreundlich; Zeit-
faktor; Mathematische_Methode; Systemforschung Menutechnik
Ref.-Nr.: Z 8507 2507

Nakatani, L. H.; Rohrlich, J. A.:
Soft machines. A philosophy of user-computer interface design. S. 19-23.
Mensch-Maschine-Kommunikation; Hardware; Software; Systemforschung
Ref.-Nr.: Z 8507 2508

Jacob, R. J. K.:
Executable specifications for a human-computer interface. S. 28-34.
Mensch-Maschine-Kommunikation; Kommandosprache; Syntax; Electronic_mail;
Prototyp; Wehrwesen
Ref.-Nr.: Z 8507 2509

Roach, J. W.:
Formal specifications for modeling and developing human/computer inter-
faces. S. 35-39.
Mensch-Maschine-Kommunikation; Programmiersprache; Logik; Systemforschung;
Methode; Verkehrswesen
Ref.-Nr.: Z 8507 2510

Murrell, S.:
Computer communication system design affects group decision making. S.
63-67.
Mensch-Maschine-Kommunikation; Electronic_mail; Teleconferencing; Human-
faktor; Entscheidungshilfe; Effektivitaet; Wirkungsforschung Window system
Ref.-Nr.: Z 8507 2512

Kieras, D.:
A generalized transition network representation for interactive systems.
S. 103-106.
Mensch-Maschine-Kommunikation; Simulation; Systemforschung Generalized
Transition Networks
Ref.-Nr.: Z 8507 2513

Revesman, M. E.; Greenstein, J. S.:
Application of a model of human decision making for human/computer communi-
cation. S. 107-111.
Mensch-Maschine-Kommunikation; Mathematisches_Modell; Effektivitaet
Ref.-Nr.: Z 8507 2514

Dumais, S. T.; Landauer, T. K.:
Using examples to describe categories. S. 112-115.
Information_Retrieval_System; Datenfeld; Datenstruktur; Mensch-Maschine-
Kommunikation Menu
Ref.-Nr.: Z 8507 2515

Tennant, H. R.; Ross, K. M.; Thompson, C. W.:
Usable natural language interface through menu-based natural language
understanding. S. 154-160.
Mensch-Maschine-Kommunikation; Sprache; Benutzerfreundlich Menu
Ref.-Nr.: Z 8507 2519

Kelly, J. F.:
An empirical methodology for writing user-friendly natural language compu-
ter applications. S. 193-196.
Mensch-Maschine-Kommunikation; Frage-Antwort-System; Benutzerfreundlich;
Computerlinguistik; Effektivitaet
Ref.-Nr.: Z 8507 2521

Brenner, E. H.; Lucey, J. H.; Martinez, C. L.; Meleka, A.:
American Petroleum Institute's machine-aided indexing and searching project.
.. Science and technology libraries, New York, NY, US 5(1984)1, S. 49-62.
Indexierungsverfahren; Rechnerunterstuetzt; Referat; Schlagwort; Auswahl;
Methode; Recherche; Expertensystem
Ref.-Nr.: Z 8507 2549

Elithorn, A. (Hrsg.); Banerji, R. (Hrsg.):
Artificial and human intelligence.
Amsterdam, NL: North-Holland (1984)

Kuenstliche_Intelligenz; Systemforschung; Psychologie; Humanfaktor; Mensch-Maschine-Kommunikation; Modell
Ref.-Nr.: Z 8507 2590

(u.a.) mit folgenden Beiträgen:

Wilks, Y.; Bien, J.:
Beliefs, points of view and multiple environments. S. 147-171,.
Kuenstliche_Intelligenz; Frage-Antwort-System; Mensch-Maschine-Kommunikation; Humanfaktor; Systemumgebung
Ref.-Nr.: Z 8507 2594

Nagao, M.:
A framework of a mechanical translation between Japanese and English by analogy principle. S. 173-180.
Kuenstliche_Intelligenz; Maschinelle_Uebersetzung; Modell; Humanfaktor; Japanisch; Englisch.
Ref.-Nr.: Z 8507 2595

Taylor, M. M.:
The bilateral cooperative model of reading. A human paradigm for artificial intelligence. S. 239-250.
Kuenstliche_Intelligenz; Lesen; Modell; Humanfaktor; Mustererkennung
Ref.-Nr.: Z 8507 2597

Sternberg, R. J.:
Common and uncommon issues in artificial intelligence and psychology. S. 281-288.
Kuenstliche_Intelligenz; Psychologie; Methode; Vergleich
Ref.-Nr.: Z 8507 2598

Colombetti, M.; Guida, G.; Pernici, B.; Somalvico, M.:
Reasoning in natural language for designing a data base. S. 297-303,.
Systemforschung; Datenbank; Entwurf; Sprache; Inferenz; Kuenstliche-Intelligenz; Wissensrepraesentation
Ref.-Nr.: Z 8507 2599

Ballmer, T. T.:
Die Allgemeinsprache als Grundlage des menschlichen Wissens.
In: Henzler, R. G. (Hrsg.): Anwendungen in der Klassifikation Band 1. Proceedings, 8. Jahrestagung der Gesellschaft fuer Klassifikation e.V., Hofgeismar, 10.-13. April 1984. Frankfurt, DE: INDEKS (1985) S. 3-18.
Klassifikationsmerkmal; Sprache; Semantik; Syntax; Pragmatik
Ref.-Nr.: Z 8507 2617

Wilss, W.:
Menschliche Uebersetzung und maschinelle Uebersetzung.
Lebende Sprachen, Berlin, DE 30(1985)1, S. 10-15.
Uebersetzungswesen; Masch._Uebersetzung; Vergleich; Software; Entwickl.tendenz
Ref.-Nr.: Z 8507 2692

Boehm, B.:
Probleme der Textverarbeitung an wissenschaftlichen Einrichtungen. Zum Stand der Pilotstudie Textverarbeitung an der Universitaet Regensburg.
LDV-Forum, Frankfurt, DE 2(1984)2, S. 4-10.
Textverarbeitung; Implementierung; Anforderung; Bildschirm; Arbeitsplatzgestaltung; Benutzerfreundlich; Software; Wissenschaft
Ref.-Nr.: Z 8507 2693

Bauer, G.:
Wissensdarstellung fuer eine effektive Mensch-Rechner-Interaktion.
Rechentechnik Datenverarbeitung, Berlin, DD 22(1985)2, S. 33-36.

Mensch-Maschine-Kommunikation; Frage-Antwort-System; Semantisches_Netz; Wissensrepräsentation
Ref.-Nr.: Z 8507 2699

Borgman, C. L.:
Psychological research in human-computer interaction.
Annual review of information science and technology, White Plains, NY, US
19(1984) S. 33-64.
Mensch-Maschine-Kommunikation; Humanfaktor; Informationsverhalten; Textverarbeitung; Information_Retrieval
Ref.-Nr.: Z 8507 2734

Lesk, M.:
Programming languages for text and knowledge processing.
Annual review of information science and technology, White Plains, NY, US
19(1984) S. 97-128.
Programmiersprache; Geschichte; Textverarbeitung; Kuenstliche_Intelligenz; Expertensystem; Bewertung; Wissensrepräsentation
Ref.-Nr.: Z 8507 2736

Tucker, A. B.; Nirenburg, S.:
Machine translation. A contemporary view.
Annual review of information science and technology, White Plains, NY, US
19(1984) S. 129-160.
Maschinelle_Uebersetzung; Projekt; Bewertung; Grammatik; Woerterbuecherstellung
Implementierung; Kuenstliche_Intelligenz; Parsing
Ref.-Nr.: Z 8507 2737

Larkin, J. H.:
Information processing in scientific thought.
In: Warren, K. S. (Hrsg.): Selectivity in information systems. Survival of the fittest. New York, NY, US: Praeger (1985) S. 10-26.
Informationsflut; Kuenstliche_Intelligenz; Humanfaktor; Expertensystem; Intelligenz; Menschliche Informationsverarbeitung
Ref.-Nr.: Z 8507 2744

Cherry, J. M.:
Human-computer dialogue.
In: Kent, A. (Hrsg.): Encyclopedia of library and information science Volume 38, Supplement 3. New York, NY, US: Dekker (1985) S. 205-225
Mensch-Maschine-Kommunikation; Systemforschung; Kommandosprache; Informationsverhalten; Humanfaktor
Ref.-Nr.: Z 8507 2851

Kotov, R. G.:
Language in real communication as an object of applied linguistics.
International forum on information and documentation, Moskva, SU 9(1984)4, S. 17-21.
Mensch-Maschine-Kommunikation; Sprache; Kuenstliche_Sprache; Entwicklungstendenz
Ref.-Nr.: Z 8508 2990

Piotrovsky, R. G.:
The semiotical interpretation of machine translation.
Int. forum on information and documentation, Moskva, SU 9(1984)4, S. 22-26.
Maschinelle_Uebersetzung; Computerlinguistik; Semiotik; Mensch-Maschine-Kommunikation
Ref.-Nr.: Z 8508 2991

Santaella, E. M.; Slamecka, V.:

Toward native language software for information management.

Information processing and management, Oxford, GB 20(1984)4, S. 527-534.

Mensch-Maschine-Kommunikation; Sprache; Kommandosprache; Software; Mehrsprachig; Benutzerführung; Benutzer; Richtlinie

Ref.-Nr.: Z 8508 3114

Jonak, Z.:

Automatic indexing of full texts.

Information processing and management, Oxford, GB 20(1984)5/6, S. 619-627.

Maschinelles_Indexierungsverfahren; Volltext; Indexierungsverfahren; Manuell; Vergleich; Suchfrage; Semantik; Methode; CSSR

Ref.-Nr.: Z 8508 3122

Cross, G. R.; De Bessonnet, C. G.:

Representation of legal knowledge for conceptual retrieval

Information processing and management, Oxford, GB 21(1985)1, S. 35-44.

Jura; Information_Retrieval; Expertensystem Wissensrepräsentation

Ref.-Nr.: Z 8508 3128

Taylor, A.:

But I have promises to keep - PRECIS, an alternative for subject access.

Technical services quarterly, New York, NY, US 2(1984)1/2, S. 75-90.

Maschinelles_Indexierungsverfahren; Sachkatalogisierung;

Belkin, N. J.; Henrmaton provision mechanism.

Inf, GB 3(1984)3, S. 122-141.

System V.; Thomas, R. C.:

Experimental adaptive interface.

Information techn-Nr.: Z 8508 3157

Hartmann, C.:

Dialog- und Kommandofunktionen. Synopse bei grossen Online-Informationssystemen.

Rechentechnik Datenverarbeitung, Berlin, DD 22(1985)3, S. 30-32

Mensch-Maschine-Kommunikation; Dialogsystem; Kommandosprache; Vergleich

Ref.-Nr.: Z 8508 3180

Turner, J. A.:

Computer mediated work. The interplay between technology and structured jobs.

Conference 5: Information systems, Tucson, AZ, US: 1984. Communications of the ACM, Baltimore, MD, US 27(1984)12, S. 1210-1217.

Mensch-Maschine-Kommunikation; Systemumgebung; Arbeitsablauf; Humanfaktor; Arbeitszufriedenheit; Bueroorganisation

Ref.-Nr.: Z 8508 3217

liebe meine GLDV-Mitglieder !

Bonn, im November 1985

Ursache dieses Schreibens ist, daß ich in letzter Zeit mehrmals darauf angesprochen wurde, doch mal zu sagen, was es denn eigentlich mit diesem Verein, der GLDV, auf sich habe. Und da die Antwort immer recht kurz ausgefallen ist, möchte ich diese Gelegenheit benutzen, mich an mehrere von Ihnen zugleich zu wenden.

Das tue ich in keiner offiziellen Funktion, sondern als ein 'altes' GLDV-Mitglied, das mittlerweile gelernt hat, daß dieser Verein - wie sicher auch andere - auch dadurch charakterisiert ist, daß vom persönlichen Einsatz einzelner Mitglieder eine Menge abhängt und daß man oft einen langen Atem braucht, um etwas auf die Beine zu stellen.

Einen Pressesprecher, Public-Relations-Menschen oder ähnliches gibt es nämlich nicht; der Vorstand, der die Vereinsaufgaben neben den üblichen vielfältigen dienstlichen Aufgaben in Lehre und Forschung ehrenamtlich übernimmt, hat oft anderes zu tun; zu einem Vereinssekretariat haben wir es noch nicht gebracht - der Verein hat ja nicht einmal einen festen Sitz an einem Ort, nur einen Ort für den Gerichtsstand, wie das ja nun mal sein muß.

Wenn Sie also nicht gleich nach Ihrer Beitrittserklärung einen persönlichen Brief erhalten, der die offizielle Freude über Ihren Beitritt ausdrückt, so wundern Sie sich bitte nicht. Der Vorstand muß ja Ihrem Beitritt erst zustimmen, und der Vorstand kommt nur von Fall zu Fall wenige Male im Jahr zusammen. (Im Augenblick sind dazu Reisen von Hannover, Heidelberg, Essen und Regensburg nötig.)

Besser ist, Sie verstehen sich als neues Mitglied vor allem als jemanden, der nicht nur Interesse an der GLDV bekundet, sondern unverbrauchte Tätigkeitsbereitschaft mitbringt.

Wenn das folgende Ihr Selbstverständnis in dieser Richtung beeinflussen könnte, wird das sicher zu Ihrer Zufriedenheit mit dem Verein beitragen.

Trotzdem - die Frage " Was bietet mir denn die GLDV?" läßt sich glücklicherweise mehrfach beantworten:

- Sie bietet den Mitgliedern - den kostenlosen Bezug des in ehrenamtlicher Redaktion erscheinenden "LDV-Forums", bisher auch der Tagungsbände und zusätzlich den von Arbeitspapieren der Arbeitskreise;
- Einladungen zur Teilnahme an den Jahrestagungen mit verbilligter Tagungsgebühr, (studentische Mitglieder sind von der Tagungsgebühr befreit)
 - den verbilligten Bezug der Zeitschrift "Sprache und Datenverarbeitung" (SDV),
 - neben vielerlei Möglichkeit zur Information auf den Jahrestagungen die unbezahlbare Gelegenheit im Rahmen des Vereins eigene Arbeiten zur LDV und in Nachbarbereichen vorzustellen, zu diskutieren und weiterzutreiben.

Damit Sie Atmosphärisches im Verein besser einschätzen können, lassen sie uns kurz auf seinen Werdegang zurückblicken:

Der Verein wurde 1975 gegründet - zunächst als eine lockere Vereinigung von wissenschaftlichen Mitarbeitern und wenigen Professoren.

- Sie sehen, einige Mitglieder kennen sich also schon lange und sind daher zu Vereins-Insidern geworden. Lassen Sie sich aber dadurch nicht hindern, den ersten Schritt zum Kennenlernen zu tun, auch wenn es so scheint, als seien die Insider sich selbst genug, mit Wichtigerem befaßt und nicht gerade aufgeschlossen gegenüber Neuen.

Übrigens ist das "Linguistische Kolloquium" in ähnlicher Weise entstanden

- als Arbeitskreis für eine neue Linguistik und gegen die herrschende Lehre.

Dem war vorausgegangen - ich meine, dem Gründen - daß sich im Februar 1973 in München am neugegründeten Institut für Phonetik und Sprachliche Kommunikation (IPK), Gastgeber: Prof. Tillmann, ca. 50 Personen trafen, die in irgendeiner Weise, meist in der Forschung, auf LDV-Gebiet tätig waren. Mitarbeiter folgender Institutionen waren vertreten:

von Universitätsinstituten: Germanistisches Institut der Universität Saarbrücken,
Seminar für Deutsche Philologie der Universität München,
Institut für Phonetik und Sprachliche Kommunikation,
Institut für Rechtsphilosophie, beide Uni München,
Rechenzentrum der Universität Köln,
Fachbereich Sprach- und Literaturwissenschaft sowie
Rechenzentrum der Universität Regensburg,
Institut für Kommunikationsforschung und Phonetik (IKP) der Universität Bonn,
Institut für Datenverarbeitung und Mathematisches Institut der TU München,
Fachbereich Sprachwissenschaft der Universität Konstanz,
Germanistisches Institut der TH Aachen,
Institut für Informatik der Universität Hamburg,
Institut für Informatik der Universität Stuttgart;

von Instituten und Institutionen in anderer Trägerschaft:

Institut für Deutsche Sprache (IDS), Mannheim, mit seinen Außenstellen in Bonn und Freiburg,
Institut für Linguistische Maschinelle Sprachverarbeitung (LIMAS), Bonn,
Institut für mathematisch-empirische Systemforschung (MESY), Aachen;

von Behörden, bzw. Behörden nachgeordneten Instituten:

Presse- und Informationsamt der Bundesregierung (BPA),
Bundessprachenamt,
Forschungsinstitut Funk und Mathematik (FFM);

aus der Industrie: Siemens, München.

In dieser Zusammenstellung bin ich der Teilnehmerliste gefolgt, denn notiert habe ich das damals nicht.

- Übrigens existieren heute nicht mehr: LIMAS, MESY, FFM und die genannten Außenstellen des IDS.

Die Teilnehmer kamen zusammen, um Informationen aus ihrer Arbeit auszutauschen und um Forschungsarbeiten zu koordinieren.

Sie bildeten zu diesem Zweck fünf Arbeitskreise mit den Themen:

1. Gesprochene Sprache
2. Linguistische Theorie, Methoden, Analysesysteme
3. Hardware, Software
4. Juristische Fragen, Strukturprobleme
5. Projektthemen, Projektnehmer, Projektgeber.

Die Arbeitkreisthemen waren nicht scharf voneinander abgegrenzt, was zu einer Zusammenarbeit zwischen den Arbeitskreisen, aber auch zu Änderungen bei den jeweiligen Zielsetzungen führte.

Immerhin waren die Ergebnisse der ersten beiden Arbeitstagungen 1973 und 1974 geeignet, die Formulierung gemeinsamer Ziele zu erreichen:

- i) Um die Forschungsaktivitäten im LDV-Bereich zu koordinieren, sollte eine Informationszentrale über laufende Projekte eingerichtet werden.
- ii) Es sollte ein "projektübergreifender Katalog" von "anwendungsseitig und forschungsstrategisch interessanten Zielsetzungen" aufgestellt werden, um im Zusammenhang damit die "Nahtstellen" zu definieren, an denen Kooperation möglich wäre.
- iii) Die Existenz einer Informationszentrale, die auch gemeinsame Zielsetzungen herausstellt, wird forschungspolitisch bedeutsam werden, und die Informationszentrale kann als solche Institution ihren Einfluß beratend auf Projektgeber und Projektnehmer geltend machen.

Da die Zusammenarbeit der Mitarbeiter aus unterschiedlich orientierten Projekten unter einer gemeinsamen Zielsetzung im Mittelpunkt der Zusammenkünfte stand, wurde der Vereinigung der Arbeitskreise der Name "LDV-Fittings" gegeben, - was den Wunsch ausdrückte, an Nahtstellen Bindeglieder zu schaffen.

(- so wie Rohre durch ein übergreifendes Bindeglied miteinander "gefittet" werden)

In der 1975 beschlossenen Satzung des Vereins findet sich dieser Gedanke in § 2,1:

" Zweck der Vereinigung ist die Förderung der wissenschaftlichen linguistischen Datenverarbeitung mit dem Ziel der Kooperation und Koordination von Forschungsvorhaben."

Unter den besonderen Aufgaben des Vereins sind in § 3 genannt:

- Forschungsdokumentation
- Daten- und Programmaustausch
- Beratung (bei Forschungsaufgaben)
- Forschungsplanung.

In § 10 wird als besondere Einrichtung eine "Clearingstelle LDV" (s.o. i)) geplant, deren Dienste die Mitglieder des Vereins in Anspruch nehmen dürfen, der gegenüber sie aber auch zur Information und Zusammenarbeit verpflichtet sind. (s. § 8 Rechte und Pflichten der Mitglieder).

Daß die Formulierung des § 8 von verdeckter Brisanz war, stellte sich unter anderem 1976 heraus, als auf einer Mitgliedervollversammlung der Antrag gestellt wurde, den einschränkenden Zusatz

"soweit nicht Rechte Dritter beeinträchtigt werden"

der "Informationspflicht" hinzuzufügen. Diese "Rechte Dritter" sind z.B. Rechte der Arbeitgeber oder Rechte der Projektfinanzierer und Auftraggeber gegenüber den Mitarbeitern des Projektes.

Der Änderungsantrag wurde übrigens mit knapper Mehrheit abgelehnt. Soweit ich mich erinnere, sind wegen des Pflichtenkonflikts einige Mitglieder darauf ausgetreten.

- Übrigens sind auch nicht alle Initiatoren dem Verein beigetreten; 1976 wurden 30 Mitglieder des Vereins gezählt.

Daß ein Gutteil des ersten Elans reduziert wurde, als die Arbeitskreise in Detail-Arbeiten einstiegen, ist leicht verständlich, wenn man bedenkt, daß recht verschiedene Individuen mit unterschiedlichen Interessen und in divers strukturierten Forschungsprojekten tätig sich zusammengefunden hatten. (Wenn Sie daraufhin noch einmal die Liste der im Hintergrund stehenden Institute und Institutionen rekapitulieren, wird Ihnen deutlich, warum ich sie aufgeführt habe.) Das ist u.a. daran abzulesen, daß der Arbeitskreis "Linguistische Theorie, Methoden, Analysesysteme" unter einer mehr praktischen Zielsetzung "Terminologiefragen" 1976 wiederbelebt werden sollte.

Aber auch das war ebensowenig erfolgreich, - nicht nur weil Vereinsmitglieder sich nicht auf bestimmte Begriffsbedeutungen einigen konnten, sie waren teil-

weise auch nicht bereit, ihren höchst eigenen Gebrauch von Fachtermini festzulegen oder festlegen zu lassen.

Damit stehen wir letztlich vor Problemen der Mensch-Mensch-Kommunikation.

- Diese Sensitivität gegenüber Begriffen ist einmal als typisches Verhalten von Geisteswissenschaftlern apostrophiert worden. Die Beurteilung stammt aber wohl aus einer gewissen Resignation, sie erklärt nicht, und das Unbehagen bleibt, da doch ein Teil dieser Geisteswissenschaftler zugleich hellhörig Forschende auf dem Gebiet der verbalen menschlichen Kommunikation sind und mittlerweile auch Lehrende, die ihre wissenschaftlichen Vorstellungen in Wörtern weitergeben und ein Interesse an möglichst wenig mißverständlicher Kommunikation haben.

Immerhin, die Arbeiten des Arbeitskreises "Clearingstelle" führten zur Übernahme der LDV-Literaturdokumentation durch die Zentralstelle für Dokumentation (ZDOK) in Frankfurt, die zukünftig die Abonnenten mit dem sogenannten Standardprofil LDV versorgte (d.i.: Sammlungen bibliographischer Angaben neu erschienener LDV-Literatur).

Das Institut für Deutsche Sprache gab eine Kurzdarstellung von LDV-Projekten im Rahmen seiner Clearings-Aktivitäten für Neueres Deutsch. In Zusammenarbeit des IKP Bonn mit der Gesellschaft für Mathematik und Datenverarbeitung (GMD), dort dem Institut für Software-Technologie (IST), wurde der Stand der Anwendersoftware LDV mit Einzelangaben in Zusammenstellungen des Informationszentrums Sozialwissenschaften, Bonn, aufgenommen und von letzterem in Zusammenarbeit mit dem IdS 1982 veröffentlicht.

An diesen Verflechtungen ist zu sehen, mit welchen Schwierigkeiten die Clearingstätigkeiten verbunden waren.

Ich vermag jedoch nicht zu sagen, ob es allein finanzielle Probleme waren, die zu dieser anfangs vom Verein vehement abgelehnten Dezentralisierung geführt haben, oder ob die offizielle Forschungspolitik (z.B. IuD-Programm des BMFT) die Gründung einer Zentralstelle unmöglich gemacht hat oder ob mangelnde Einigkeit oder Durchsetzungsvermögen und zögerliches Handeln des Vereins dafür ursächlich waren. (Dafür wird es wohl mehrere Lesarten geben, wenn sie auf den Tisch kommen.)

Während zunächst halbjährliche Mitgliederversammlungen vorgesehen waren, die auch bis 1977 fast regelmäßig in München, Bonn und Mannheim stattfanden, wurden die Mitgliederversammlungen ab 1978 im Zusammenhang mit wissenschaftlichen Symposien als Jahrestagungen durchgeführt, zunächst in Essen, dann

1979 in Bonn - zusammen mit der Jahrestagung der Association for Literary and Linguistic Computing (ALLC),

1980 in Saarbrücken

1982 in Koblenz

1983 in Trier

1984 in Heidelberg

1984 in Hannover.

Seit der Jahrestagung in Heidelberg ist der Name des Vereins in "Gesellschaft für Linguistische Datenverarbeitung" geändert, der "Fittings"-Gedanke blieb aber grundlegend.

An offiziellen Aktivitäten des Vereins ist der Vorstand beteiligt, der Beirat, Organisationskomitees der Jahrestagungen und die Arbeitskreise. Darüber hinaus sind gewiß die Aktivitäten wesentlich, die einzelne Mitglieder von Fall zu Fall ihrer Interessenlage entsprechend verbinden, z.B. im Forschungsbereich.

Ich vermute, daß diese bi- oder multilaterale Zusammenarbeit das, was in der LDV geschieht, erheblich bestimmt; es ist aber letztlich nicht Sache des Vereins, da die übrigen Mitglieder allenfalls von Ergebnissen, selten von Verhandlungen Kenntnis erlangen.

Die Gesamtlage der LDV in ihrer finanziellen Ausstattung, in ihrer Stellensituation wie in apparativen Möglichkeiten macht es dem Einzelnen zunehmend schwerer, Vereinsaufgaben zu übernehmen - man denke beispielsweise daran, wie viele LDV-Lehrstühle oder Teil-LDV-Lehrstühle nur mit halbtags besetzten Sekretariaten oder auch gar keinen festen Schreibkräften ausgestattet sind, - so daß oft schon die bloße Verbreitung von Mitteilungen an technischen Dingen scheitert (und das in unser informationsmitteilungsfreudigen Zeit!). Wenn man etwas im Verein erfragen will oder etwas anregen, stellt sich als erstes die Frage, mit wem man Kontakt aufnehmen soll, wer wohl am sichersten zu erreichen ist und dann - ob man nicht bis zur nächsten Jahrestagung warten soll. So versendet sicher auch viel Initiative und Energie.

Eine relativ lockere Vereinigung hat allerdings auch ihr Gutes: Es gibt keine kontinuierliche Vereinspolitik betreibenden, maßgeblichen Repräsentanten des Vereins. Das läßt Einzelaktivitäten Freiraum, der jedoch als solcher erkannt werden muß, denn er wird nicht ausdrücklich freigegeben. Von wem auch?

Immerhin gibt es auch heute Arbeitskreise, deren Arbeitsweise und Zusammensetzung übrigens so verschieden ist, daß Sie Anfragen am besten an eine der Kontaktpersonen richten. (Auf der Jahrestagung in Göttingen wird erstmals konkurrenzlos Zeit für Arbeitskreise eingeplant sein; so gibt es möglicherweise auch Gelegenheit zur Vorstellung.)

Arbeitskreis MIKROCOMPUTER: Dr.U.Klenk, Ernst-Barlach-Weg 3, 3400 Göttingen

LDV UND NACHBARN: Dr.G.Knorz, TH Darmstadt, FB Informatik, DVS II
Karolinenplatz 5a, 6100 Darmstadt

LEXIKOLOGIE: Dipl.inf.G.Frackenpohl, Lügenhausenerstr. 102,
5064 Rösrath 1

MASCHINELLE ÜBERSETZUNG: Dr.H.J.Weber, Zum Eschberg 2, 6601 Saarbrücken-Schafbrücke

AUSBILDUNG UND BERUFSPERSPEKTIVEN: Dr.M.Lutz-Hensel, p.a.Institut
für Kommunikationsforschung und Phonetik der
Univ.Bonn, Poppelsdorfer Allee 47, 5300 Bonn 1

In Anbetracht des Programms, unter dem der Verein einmal angetreten ist, möchte ich sagen, daß er sich heute noch im Widerstreit verschiedener, größtenteils verdeckter Interessen befindet - oder wieder in einer neuen Selbstfindungsphase. Diese Situation können die Mitglieder nutzen.

Hat bei der Gründung offiziell die Forschung im Mittelpunkt der Interessen gestanden, so ist mittlerweile auch die LDV-Lehre ins Blickfeld geraten, das umso mehr, als der Verein jetzt ja auch studentische Mitglieder hat, die nicht in LDV-Forschungsprojekten ihr Interesse an LDV entdeckt haben. Und auch die beruflichen Werdegänge von LDV-Absolventen außerhalb der Universitäten werden uns zukünftig mehr beschäftigen.

Haben Sie also keine Scheu!

Begrüßen Sie sich als Mitglied der GLDV und laden Sie sich zur Mitarbeit im Verein ein!

*Unter völliger Mißachtung von Sprechaktsregeln
- aber mit sehr freundlichen Grüßen
Ihre M. Lutz-Hensel*

SITZUNG DES BEIRATES DER GLDV

Protokoll

Auf seiner ersten Sitzung am 25.10.85 in Frankfurt wählte der GLDV-Beirat I. Batori einstimmig zu seinem Sprecher.

Dann befaßte er sich mit einer Reihe von Themen, die für die Arbeit der GLDV wichtig sind:

Fortbildungsveranstaltungen der GLDV sollten dezentral initiiert werden, sich am konkreten Bedarf orientieren und sich wegen der Verringerung des Reiseaufwandes möglichst an die Jahrestagung anschließen. Die Teilnahmegebühren sollen gestaffelt werden, damit Studenten wenig belastet werden.

Der Beirat sieht eine inhaltliche Lücke bei den deutschsprachigen Fachzeitschriften, die das LDV-Forum ausfüllen könnte, wenn es sich langsam auf eine Fachzeitschrift hin entwickelt. Das Thema wird bei der nächsten Sitzung erneut behandelt.

Bei der Jahrestagung in Göttingen soll eine Veranstaltung zum Thema "LDV-Ausbildung" ohne Parallelveranstaltungen stattfinden. Eine Diskussion zur Klärung des Verständnisses von Computerlinguistik im Bezug zu LDV und philologischer Datenverarbeitung wird für möglichst den ersten Abend angeregt. Mittelfristig soll die Basis für eine notwendiger erscheinende kritische Bewertung von Studiengängen in LDV/CL gelegt werden.

In München wird ein neues Studienfach Informationswissenschaftliche Sprach- und Literaturforschung geplant. Am Fachbereich 14 soll ein neues Institut mit einer H4-Stelle und zwei H3-Stellen gegründet werden. Die Ausschreibungen sollen noch dieses Jahre beginnen. Weitere Mitarbeiter sollen über Projekte finanziert werden.

Das Institut bietet in Ergänzung des schon begonnenen Aufbaustudiums einen grundständigen Studiengang an, dessen Absolventen ihre Tätigkeitsfelder in der informationsverarbeitenden Industrie, in Verlagen, Redaktionen, Bibliotheken und Museen finden sollen. Vorbild für das Institut ist das Center for the Study of Language and Information (CSLI) an der Universität Stanford. Arbeitsschwerpunkte sollen z.B. informationswissenschaftliche Methoden zur Überprüfung geisteswissenschaftlicher Theorien und sprach- und textbezogene KI sein. Das Institut soll mit der Industrie in einem für die Geisteswissenschaften bisher unüblichen Ausmaß zusammenarbeiten.

Der Beirat unterstützt einen Vorstoß der GLDV bei der DFG, der zu einer besseren Vertretung der Computerlinguistik in den Fachausschüssen der DFG führen soll und einen Schwerpunkt "Informationsverarbeitung" anregt.

Brigitte Endres-Niggemeyer

Weitere "Mitteilungen aus der GLDV", die erst kurz vor Drucklegung eingingen, finden sich auf Seite 136