

Linguistik und neue Medien GLDV-Jahrestagung vom 17. -19. März 1997 am Institut für Informatik der Universität Leipzig

*Gerhard Knorz,
Fachhochschule Darmstadt*

"Linguistik und neue Medien", ein Rahmenthema, wie geschaffen für eine Tagung in Leipzig, der - wie sie sich selbst einschätzt - Stadt der Medien, und das auch noch unmittelbar im Vorfeld der Leipziger Buchmesse. Mit ihrer ersten Tagung in den neuen Bundesländern zeigte die Gesellschaft für Linguistische Datenverarbeitung (GLDV), daß mit ihr weiterhin zu rechnen ist: Das wissenschaftliche Programm war ordentlich, die Diskussionen munter, und mit knapp 80 Teilnehmern konnte man zufrieden sein. Ein Höhepunkt besonderer Art war die Vergabe des mit 1000 DM dotierten GLDV-Preises für die beste studentische Arbeit.

1 Eröffnung der Tagung

Eine Tagung in den neuen Bundesländern bringt es mit sich, daß so mancher Teilnehmer das "Aufbauprogramm Ost" zum ersten Mal mit eigenen Augen erlebt. Den ersten Eindruck liefert der Hauptbahnhof, dessen gewaltige Stahlkonstruktion nur wie ein Wunder die Wühlarbeit um seine Fundamente herum zu überstehen scheint und das setzt sich fort an vielen Baustellen zwischen post-modernen Neubauten, restaurierten Altbäuden und solchen Häusern, deren Zustand sich über viele Jahre hinweg nur verschlechtern konnte.

Das Spannungsfeld zwischen Vergangenheit und Gegenwart, zwischen Realität und kommunizierten Eindrücken und Einschätzungen wurde auch zu Beginn der Tagung deutlich, als nach dem Vorsitzenden des Programmkomitees, Herrn Prof. Dr. Gerhard Heyer, der Rektor der Universität Leipzig, Herr Prof. Dr. Cornelius Weiss, das Wort ergriff.

Anstelle üblicher unverbindlich-freundlicher Begrüßungsfloskeln begann die Tagung mit dem interessanten und engagierten Versuch, die gastgebende Hochschule mit ihren Wurzeln (Gründung 1409!), ihrem Werdegang und den gerade abgeschlossenen Umwälzungen vorzustellen und den erreichten Stand zu bewerten. Der Rektor skizzierte die Universität Leipzig als eine wieder profilierte Hoch

schule, die mit ihren interdisziplinären Zentren eine Overlaystruktur zusätzlich zur klassischen Fakultäten-Gliederung eingerichtet hat, um das Innovationspotential an den Grenzgebieten zwischen den etablierten Fachwissenschaften aufzuspüren. Und zweifellos hatte er recht damit, daß genau dieser Aspekt zu der Tagung einer *Gesellschaft für linguistische Datenverarbeitung* besonders gut passen müsse.

2 Eingeladene Vorträge

2.1 *Organisationsformen und Mehrwertleistungen elektronischer Märkte (Kuhlen)*

Den wissenschaftlichen Teil der Tagung eröffnete *Rainer Kuhlen* (Konstanz) mit dem eingeladenen Vortrag "*Organisationsformen und Mehrwertleistungen elektronischer Märkte*". Der Vortrag war das Ergebnis einer (linearen) Wanderung durch ein reichhaltig vorgehaltenes und als Hypertext strukturiertes Material mit Übersichten, Thesen, Argumenten, Fakten und Beispielen - realisiert als interaktive Toolbook-Anwendung.

Die informationswissenschaftliche Orientierung des Beitrags war vom Veranstalter bewußt als ein "Blick über den Tellerrand" akquiriert worden. Daß im Kontext netzbasierter multimedial gestalteter virtueller Märkte die "alten" Fragestellungen von Computerlinguistik und Information Retrieval in neuem Gewand aktueller werden als sie jemals waren, konnte die Tagungsteilnehmer nur freuen. Die Beispiele und der Erfahrungshintergrund für die behandelten Aspekte wie Regionalisierung, Akzeptanz, Vertrauensbildung, Qualitätsstandards entstammten - wie könnte es anders sein - der *electronic mall Bodensee*, an deren Aufbau die Universität Konstanz maßgeblich mit beteiligt ist. Schade eigentlich, daß Kuhlen sich zu sehr von der Furcht drängeln ließ, die Tagungsteilnehmer könnten mit dem behandelten Gegenstandsbereich schon zu sehr vertraut sein: Den voll durchformulierten projezierten Statements gleichermaßen wie dem gesprochenen Wort zu folgen, war im vorgelegten Schnelldurchgang durchaus eine Herausforderung.

2.2 *Wiederverwendbarkeit lexikalischer Ressourcen (Thunnair)*

Der zweite eingeladene Vortrag der Tagung hatte die Aufgabe, die Tagungsteilnehmer auf den zweiten Tag einzustimmen. Und wer *Gregor Thurmair* kennt, brauchte sich über den interessanten, handfesten und überaus munteren Beitrag aus München nicht zu wundern. "*Wiederverwendbarkeit lexikalischer Ressourcen*" war der Vortragstitel und das Projekt A VENTINUS war das Thema. Das Projekt A VENTINUS setzt einen bunten Strauß an Techniken für eine pragmati-

sche automatische Übersetzung, für Information Extraction, für Indexierung und Retrieval auf Text- und Faktendatenbanken ein, um die polizeiliche Arbeit im Bereich des organisierten Verbrechens über Länder- und vor allen Dingen Sprachgrenzen hinweg zu unterstützen. Die Vortragsperspektive bestand in der Frage, welche sich überschneidende und divergierende Anforderungen die verschiedenen Systemkomponenten an die lexikalischen Ressourcen stellen. Und wie die jeweils benötigten lexikalischen Materialien in einer zentralen Lexikonarchitektur verwaltet und zur Verfügung gestellt werden können.

Das Projekt AVENTINUS steht am Anfang seiner Realisierung. Ab November 1998 wird man abschließend urteilen können, inwieweit die Projektziele als erreicht gelten können.

Die freimütige Offenlegung der Schwierigkeiten, die ein Unternehmen hat, das eine zentrale Komponente für ein aus Geheimhaltungsgründen völlig unbekanntes Hintergrundsystem entwickelt, die erfrischend pragmatische Art, mit der die Problemstellungen auf den Punkt gebracht wurden und die technisch zweifellos spannende Aufgabenstellung machten für mich diesen Vortrag zu einem Highlight der Tagung. Schade, daß mit dem Vortragenden anschließend nicht mehr zu reden war: Auf wichtige Leute wartet schon das Flugzeug, um sie zum nächsten wichtigen Termin zu bringen.

3 Tagungsprogramm

Das Tagungsprogramm war weitgehend durchgängig als Parallelprogramm mit Fachvorträgen und Systemvorführungen organisiert.

3.1 *Electronic Publishing und Anwendungen*

Die erste Sektion, konkurrierend zu Vorführungen zum Thema *Lexikon und Terminologie*, war mehr oder weniger passend "*Electronic Publishing und Anwendungen*" überschrieben.

Katja Krüger (Universität Stuttgart) beschrieb in ihrem Beitrag "*Mehrsprachige computergestützte Texterschließung für Übersetzer und Terminologen*" einen Ansatz, der abgeschlossene Übersetzungen, die in großer Anzahl zweisprachig in elektronischer Form vorliegen, als Ressource für die aktuelle und langfristige Arbeit von Übersetzern und Terminologen erschließen soll: Durch Extraktion von Termen, Bereitstellung von Äquivalenz-Kandidaten und Kontextbelegungen zur intellektuellen Kandidatenauswahl sowie Unterstützung beim Aufbau eines Glossars. In der Diskussion wurde insbesondere kritisch diskutiert, ob die eingesetzten Verfahren zur linguistischen Termextraktion unter Verwendung

von allgemeinen Affixlisten und anwendungsspezifischen Wortstämmen tatsächlich problemadäquat sein können.

Das aus dem Ansatz resultierende integrierte System Termland stellt eine HTML-Schnittstelle zur Verfügung, so daß die Nutzer eine einfache und vertraute Browseroberfläche verwenden können.

Der zweite Beitrag der Sektion widmete sich einer Fragestellung, die sich für viele der Anwesenden bereits praktisch gestellt haben dürfte: " *Vom Text zur Hypermedia-Anwendung: Methodisches Vorgehen bei der Hypertextualisierung nichtstandardisierter Textsorten* ". Angelika Storrer (IDS Mannheim) versuchte in diesem Beitrag, die eigene Vorgehensweise bei der Umsetzung einer umfangreichen Print-Publikation (wissenschaftliche Grammatik des Deutschen) in eine ergänzende Hypermedia-Anwendung (GRAMMIS, Grundlagen eines grammatischen Informationssystems) zu reflektieren und zu systematisieren. Für die Beurteilung, inwieweit die vorgestellten Prinzipien tatsächlich in allgemeiner Weise gelten und gleichzeitig über das hinausgehen, was von einem "umsichtigen HypermediaDesigner" ohne weitere Hintergrundtheorie erwartet werden kann, habe ich mir das Studium der schriftlichen Fassung vorgenommen.

Außerordentlich lebhaft wurde zum letzten Vortrag der Sektion diskutiert: Christian Wolff stellte den Beitrag " *Zur Relevanz der linguistischen Pragmatik bei der Entwicklung von Multimedia-Anwendungen* " vor, für den er zusammen mit Gerhard Heyer verantwortlich zeichnet. Ausgangspunkt war das allgemein empfundene Unbehagen darüber, daß die Entwicklung von Multimedia-Anwendungen im Informatikbereich weitgehend technikzentriert untersucht und betrieben wird. Klassische Theorien der linguistischen Pragmatik, angewandt auf den neuen Kontext, könnten einen substantiellen Beitrag zur Behebung dieses Defizites leisten, so der Ansatz aus Leipzig, der eingeständenerweise noch ziemlich am Anfang seiner Entwicklung steht. Der Überlegung, daß man die Sprechakttheorie sinnvoll auf den neuen Gegenstandsbereich anwenden könne, wurde auch nicht widersprochen. Aber wieso kann's nicht auch die Rhetorik sein, so wurde gefragt? Und, hier äußerten sich gleich mehrere Teilnehmer engagiert, sollte man sich nicht eher am Ansatz des Kommunikationsdesign orientieren, der sehr viel stärker vom Einzelfall ausgeht? Alles in allem eine Thematik, mit der man ComputerlinguistInnen (und nicht nur solche) hinter dem Ofen vorlocken kann...

3.2 *Neue Medien und Ausbildung*

Die Sektion IV mit Systemvorführungen zum Thema " *Neue Medien und Ausbildung* " konkurrierte mit Fachvorträgen zu "Computerlexikographie" . Ich habe mich für ersteres entschieden und dies nicht bereut. Alle drei vorgestellten Systeme waren einer sorgfältigen Betrachtung wert.

RECALL, Projekt und Demonstrator für den computerunterstützten Fremdsprachenunterricht, wendet CL- und AI-Methoden an, um den Lernenden realitätsnah und motiviert beim Erwerb einer Sprache zu führen und zu fördern. Das System wurde von *Peter Gerstl* (IBM-WZ Heidelberg) vorgestellt, und zeigte an vielen Details, von welchem Wert eine didaktische Konzeption ist, die auf einer empirischen Grundlage aufbauen kann.

Der Titel "Visualisierung russischer fachsprachlicher Grammatik in einem interaktiven System" schreckt den des Russischen Unkundigen zu Unrecht ab: Die Vorführung durch Friedrich Wenzel (Universität Hannover) und eine Studierendengruppe vermittelte den Eindruck, daß hier auf ganz pragmatische Weise ein sehr brauchbarer Ansatz entwickelt wurde, Naturwissenschaftlern und Ingenieuren den Erwerb der russischen Sprache zu erleichtern. Zwei Schlüsseleigenschaften machen das System so interessant: Zum einen die Verwendung animierter graphischer (Sprach-)Strukturen, mit denen interaktiv gearbeitet werden kann. Und zum anderen die Technik der Interaktion selbst: Lehrende und Lernende verwenden einen roten Laserpunkt auf der Projektionsfläche in gleicher Weise und mit gleicher Wirkung wie eine Maus. Möglich macht dies eine Erkennungsoptik, die auf dem LDC-Display angebracht wird.

Die Vorführung zum Vortrag am Vortag lieferte *Angelika Starrer* (IDS Mannheim): Sie stellte den Demonstrationsprototyp *GRAMMIS- I* mit seiner Oberfläche und seinen Navigationsmöglichkeiten vor und demonstrierte, mit welchen Konzepten und Visualisierungen versucht wurde, dem elektronischen Medium einen Mehrwert gegenüber einer gedruckten Buchversion mitzugeben. Die anspruchsvolle Grundlage macht den dargebotenen Stoff für Hochschulen, nicht so sehr jedoch für allgemeinbildende Schulen interessant. Gegen die Zusicherung, einen Evaluierungsfragebogen ausgefüllt zurückzuliefern, kann man den Prototypen anfordern und verwenden.

3.3 *Panel" Sprachgenerierung H*

Der Dienstag Nachmittag begann mit einem von Roland Hausser (Universität Erlangen) moderierten Panel zum Thema *Sprachgenerierung*. Den Teilnehmern wurde ein Fragenkatalog vorgelegt, dessen Beantwortung den überwiegenden Teil der Veranstaltungszeit erforderte. Also wenig Diskussion und viel Vortrag, aber immerhin eine gute Gelegenheit, das Thema facettenreich aus unterschiedlichen Perspektiven vorgestellt zu bekommen. Die Teilnehmer waren Karin Harbusch (Universität Koblenz-Landau), Thomas Pechmann (Universität Leipzig), Dietmar Rösner (Universität Magdeburg), Anita Steube (Universität Leipzig).

3.4 *Vorträge zum GLDV-Preis '97 für die beste studentische Arbeit*

Man mußte sich entscheiden, ob man am Dienstag Nachmittag die Sektion *Syntax und Parsing* besuchen wollte, oder aber aus erster Hand erfahren wollte, ob und wie der - nach einem Vorversuch i 995 in Regensburg - erste offizielle Wettbewerb für den GLDV-Preis '97 für die beste studentische Arbeit in Schwung kommt. Mit nur drei Einreichungen (Universität Stuttgart und Universität Koblenz) konnte man zunächst durchaus skeptisch sein, aber die Veranstaltung selbst belehrte etwaige Zweifler eines Besseren.

Im einzelnen wurden folgende Beiträge auf durchweg hohem Niveau vorgestellt:

- *Anke Kölzer (Universität Koblenz): "Lexana - Ein System zur Lexikon- und Grammatikanalyse für Kategoriale Unifikationsgrammatiken"*
- *Sergej Melnik (Universität Leipzig): "EPOS: Electronic Publishing mit OODB und SGML"*
- *Karsten Böhm; Timo Böhme (Universität Leipzig): "SQUIRREL: DBS-based Information Retrieval System for the WWW"*

Die Jury hatte keine leichte Aufgabe, sich zwischen den 3 Einreichungen zu entscheiden. Thematisch waren Computerlinguistik, Information Retrieval und Electronic Publishing abgedeckt, und sowohl die Präsentation wie auch die schriftlichen Fassungen waren als eine überzeugende Werbung anzusehen, sowohl für die Idee eines solchen Wettbewerbs und auch für die Institution, aus der die Arbeit kam. Da dem GLDV-Wettbewerb ein eigener Bericht gewidmet ist, soll hier auf eine inhaltliche Beschreibung verzichtet werden.

Bis zur Abendveranstaltung im historischen Rahmen der Moritzbastei mußten sich die Teilnehmer von Wettbewerb und Tagung gedulden. Doch nach dem Buffet war es soweit: Im Namen der Jury würdigte Gerhard Knorz (FH Darmstadt) die einzelnen Arbeiten und überreichte unter großem Beifall den mit 1.000 DM dotierten Preis an die Preisträgerin, Frau Anke Kölzer. Alle Beteiligten schienen sich einig darin, daß das Ziel sein muß, den so begonnenen Wettbewerb als eine feste Institution der Disziplin zu etablieren.

3.5 *Und sonst noch...*

Am Mittwoch, den dritten Tagungstag standen noch 2 Sektionen über *"Information Retrieval und Statistik"* und *"Corpora und Lexikon"* auf dem Programm sowie ein abschließendes Panel zur Frage *"Wie praxisrelevant ist die computerlinguistische Ausbildung?"* Und wer sich die Zeit für die am Donnerstag beginnende Leipziger Buchmesse noch erlauben konnte, der wird sich den GLDV-Workshop im Messehof zum Thema *"Neue Medien - eine Herausforderung für die Sprachverarbeitung"* vorgemerkt haben.

Alles in allem eine gut organisierte und anregende Tagung, überschaubar, aber dennoch - so das allgegenwärtige Zitat - mit mehr Teilnehmern als die letztjährige, von 5 CL-Fachgesellschaften veranstaltete KONVENS. Wie immer man einen solchen Vergleich bewerten mag, Leipzig war eine Reise wert!

"Neue Medien - Eine Herausforderung an die Sprachverarbeitung" Bericht über den letzten Tag der GLDV Jahrestagung 1997 und den GLDV- Workshop auf der Leipziger Buchmesse"

*Dr. Christian Wolff,
Universität Leipzig*

Am letzten Tag der GLDV '97 standen noch einmal eine Fach- und eine Demonstrationssektion auf dem Programm.

Die abschließende Demo-Sektion *Corpora & Lexikon* zeigte anhand zweier Vorträge die große Spannweite derzeitiger Arbeiten corpus- und lexikonbasierter Ansätze. Uwe Quasthoff, Leipzig, stellte das Projekt *Der Deutsche Wortschatz* vor, dessen Ziel, die Sammlung einer umfassenden Vollformendatenbank des Deutschen durch verteiltes Arbeiten erreicht werden soll. Eine CD-ROM mit 2,5 Mio. Vollformen, die frei verfügbar ist und auch allen Tagungsteilnehmern mit auf den Weg gegeben wurde, ermöglicht die dezentrale Ergänzung der Wortsammlung. An zahlreichen Beispielen konnten die unterschiedlichen Einsatzgebiete einer solchen Sammlung demonstriert werden. Naheliegend, daß sich eine lebhaftige Diskussion über den Wortbegriff und die Randbedingungen dieses Projektes anschloß.

Markus Schulze, Uni Erlangen stellte anschließend das System *Malaga* vor - basierend auf dem von Roland Hausser entwickelten Ansatz der links-assoziativen Grammatik stellt es eine Programmiersprache für Anwendungen der Sprachanalyse dar, die in unterschiedlichsten Szenarien benutzt werden kann. Ausgangspunkt der Demonstration war die Morphologiekomponente von *Malaga*, DMM,