

Angewandte Informatik

Herausgegeben

von

Dr. Peter Mertens

o. Prof. an der Universität Erlangen-Nürnberg



Sammlung Götschen Band 5013

Walter de Gruyter
Berlin · New York · 1972

Die Reihe „Informatik“ in der Sammlung Göschen umfaßt folgende Bände:

- Einführung in Teilgebiete der Informatik. Von *L. Hieber*. 2 Bände
Digitale Rechenautomaten. Von *R. Klar*.
Analog- und Hybridrechner. Von *G. Gensch*. (In Vorb.)
Datenübertragung und -fernverarbeitung. Von *K. Oetli*. (In Vorb.)
Programmierung von Datenverarbeitungsanlagen.
 Von *H. J. Schneider* u. *D. Jurksch*.
Datenstrukturen und höhere Programmieretechniken. Von *H. Noltemeier*.
Betriebssysteme I. Grundlagen. Von *E. J. Neuhold*. (In Vorb.)
Betriebssysteme II. Von *P. Caspers*. (In Vorb.)
Theorie und Praxis des Übersetzerentwurfs. Von *H. J. Hoffmann*. (In Vorb.)
Schaltwerk- und Automatentheorie. Von *C. Hackl*. 2 Bände
Graphentheorie für Informatiker. Von *W. Dörfler* u. *J. Mühlbacher*.
Einführung in die mathematische Systemtheorie. Von *F. Pichler*. (In Vorb.)
Formale Beschreibung von Programmiersprachen. Von *K. Alber*. (In Vorb.)
Angewandte Informatik. Von *P. Mertens*.
Information Retrieval. Von *O. Simmler*. (In Vorb.)
Programmiersprachen für die numerische Werkzeugmaschinensteuerung.
 Von *U. Grupe*. (In Vorb.)

©

Copyright 1972 by Walter de Gruyter & Co., vormalis G. J. Göschen'sche
Verlagshandlung – J. Guttentag, Verlagsbuchhandlung – Georg Reimer
Karl J. Trübner – Veit & Comp., Berlin 30. – Alle Rechte, einschl. der
Rechte der Herstellung von Fotokopien und Mikrofilmen, von der Verlags-
handlung vorbehalten. – Satz: IBM-Composer, Walter de Gruyter –
Druck: Mercedes-Druck, Berlin. – Printed in Germany.

ISBN 3 11 004112 X

Inhaltsverzeichnis

Vorwort und Einführung	9
Angewandte Informatik in einzelnen wissenschaftlichen Disziplinen und Bereichen	11
1. Linguistik (Herwig Krenn)	11
1.1 Einleitung	11
1.2 Document and Information Retrieval	11
1.2.1 Herstellung von Wortindices und Konkordanzen	12
1.2.2 Stilistische Analyse	12
1.2.3 Automatische Indizierung (Automatic Indexing)	13
1.2.4 Automatische Klassifizierung (Automatic Classification)	14
1.2.5 Automatisches Extracting and Abstracting	15
1.2.6 Content Analysis (Inhaltsanalyse)	15
1.3 Automatische (Maschinelle) Übersetzung	15
1.3.1 Automatische Analyse	16
1.3.2 Transferteil	17
1.3.3 Automatische Synthese	17
1.3.4 Zwischensprache	18
1.4 Informatik im Dienste der linguistischen Forschung	18
2. Physik (Hans-Jürgen Trebst)	19
2.1 Einleitung	19
2.2 Lösen von Differentialgleichungen als typische Aufgaben der Informa- tionsverarbeitung in der Physik	20
2.2.1 Einführung in die Rechenmethode	20
2.2.2 Vielteilchenprobleme	22
2.2.3 Kontinuumsprobleme	24
2.3 Sonstige Rechnungen	26
2.4 Computergestützte Auswertung von Blasenkammerbildern	27
2.5 On-line-Realzeitsysteme in physikalischen Versuchsanordnungen	29
2.6 Offene Probleme	32
3. Physiologie (Peter Finkenzeller)	33
3.1 Einleitung	33
3.2 Wechselwirkungen zwischen Physiologie und Informatik	34
3.3 Beispiele aus der Forschung des I. Physiologischen Instituts der Univer- sität Erlangen-Nürnberg	38
4. Medizin (Gustav Wagner)	45
4.1 Einleitung	45
4.2 Rationalisierung der medizinischen Dokumentation und des ärztlichen Berichtswesens	47

4.3	Computerunterstützte Diagnostik	51
4.4	Computerunterstützte Therapie	55
4.5	Computerunterstützte Krankenpflege	56
4.6	Rationalisierung der Krankenhausverwaltung	57
4.7	Steuerung des Informationsflusses im Krankenhaus	57
4.8	Elektronische Informationsverarbeitung in der prophylaktischen Medizin	59
4.9	Elektronische Informationsverarbeitung in der medizinischen For- schung	60
4.10	Programmierter Unterricht	61
4.11	Elektronische Literaturdokumentation	62
4.12	Ausblick in die Zukunft	63
5.	Ingenieurwesen (Rüdiger Hartwig und Willi Rehwald)	66
5.1	Begriffsbestimmungen	66
5.2	Entwicklung und Ziele	67
5.3	Überblick über die Anwendungsgebiete	68
5.4	Praktische Beispiele	69
6.	Betriebswirtschaft (Peter Mertens)	83
6.1	Abgrenzung	83
6.2	Aufgabenstellung – Integrierte Datenverarbeitung als Ziel	83
6.3	Modell einer integrierten Datenverarbeitung am Beispiel des Industrie- betriebes	85
6.4	Besonderheiten der Betriebsinformatik in nichtindustriellen Wirtschafts- zweigen	88
6.4.1	Handel	88
6.4.2	Kreditinstitute	89
6.4.3	Versicherungen	90
6.4.4	Verkehrsbetriebe	90
6.5	Management-Informationen-Systeme (MIS)	91
6.6	Auswirkungen auf die Betriebsführung	95
6.7	Offene Probleme – Forschungsaufgaben	95
7.	Volkswirtschaft (Hans-Jürgen Krupp)	98
7.1	Vorbemerkung zum Verhältnis von Informatik und Volkswirtschafts- lehre	98
7.2	Systematisierungsversuch der Informatikanwendungen in der Volks- wirtschaftslehre	99
7.3	Datenbanken und Indikatoren-systeme	104
7.4	Automaten als Simulatoren	106
7.5	Aufgaben für die Forschung	110
8.	Recht (Wilhelm Steinmüller)	111
8.1	Zum Aufbau der Rechtsinformatik	111
8.1.1	Einleitung	112
8.1.2	Grundbegriffe	113
8.1.3	Systematisierung	116
8.1.4	Abgrenzungen	118

8.2	Anwendungsfragen	118
8.2.1	Dokumentation	118
8.2.2	Rechts- und Verwaltungsautomation	121
8.2.3	Rechnerunterstützte Informationssysteme	124
8.3	Rechtsfragen der EDV	126
9.	Öffentliche Verwaltung (Joachim Griese)	128
9.1	Einleitung und Abgrenzung zur Rechtsinformatik	128
9.2	Überblick	129
9.3	Automatische Informationsverarbeitung bei Konditionalprogrammen	130
9.3.1	Ablauf der automatischen Informationsverarbeitung	130
9.3.2	Ausgewählte Übersicht automatisierbarer Verwaltungstätigkeiten	132
9.3.3	Auswirkungen der automatischen Informationsverarbeitung auf die Form der Arbeitsteilung bei Verwaltungstätigkeiten	134
9.4	Unterstützung der Formulierung von Entscheidungsprogrammen durch automatische Informationsverarbeitung	137
9.5	Besondere Probleme automatischer Informationsverarbeitung in der Verwaltung	140
10.	Sozialforschung (Erwin Scheuch und Ekkehard Mochmann)	142
10.1	Charakterisierung	142
10.2	Forschungssituation	143
10.3	Datentypen	144
10.4	Datenverarbeitungsmodelle	146
10.4.1	Univariate Analyse	146
10.4.2	Bivariate Analyse	146
10.4.3	Multivariate Analyse	147
10.4.4	Modellbildungen	149
10.5	Entwicklung von Datenanalysesystemen	151
10.6	Sozialwissenschaftliche Textverarbeitung	153
10.7	Entwicklung integrierter Datenverarbeitungssysteme	154
11.	Raumplanung, Verkehr und Umweltschutz (Jürgen Seggelke)	158
11.1	Einleitung	158
11.2	Einige Objektbereiche der öffentlichen Planung	158
11.2.1	Stadtplanung	158
11.2.2	Schulplanung	159
11.2.3	Bevölkerungsforschung	162
11.2.4	Verkehrsplanung	163
11.2.5	Umweltschutz	164
11.3	Datenbasis	165
11.4	Datenverarbeitung	166
11.5	Datenausgabe	166
11.6	Zielforschung für die öffentliche Planung	168
12.	Verteidigung (Norbert Franßen)	169
12.1	Einleitung	169
12.2	Militärische Führung	169
12.3	Militärische Planung	173

13. Bildungswesen (Miloš Lánský)	176
13.1 Einleitung	176
13.2 Adressaten-bezogener Prozeß	177
13.3 Ziel-bezogener Prozeß	177
13.4 Lehrstoff-bezogener Prozeß	178
13.5 Algorithmus-bezogener Prozeß	178
13.6 Medium-bezogener Prozeß	179
13.7 Soziostruktur-bezogener Prozeß	179
13.8 Ein Beispiel für die Anwendung der Methode VERBAL	180
13.8.1 Basaltext und Begriffsskelette	180
13.8.2 Herstellung der natürlichen Reihenfolge mit SEQUAL	182
13.8.3 Verteilung von Explanationen nach VERBAL	183
13.8.4 Die Ausarbeitung des Lehrprogrammes	187
Zum langfristigen Wissenschafts- und Forschungsprogramm einer Angewandten Informatik (Adolf Adam)	190
Stichwortverzeichnis	197

Autoren dieses Bandes

1. *Linguistik*
Priv.-Doz. Dr. *Herwig Krenn*, Universität Bochum
2. *Physik*
Dipl.-Phys. Dr. *Hans-Jürgen Trebst*, Universität Erlangen-Nürnberg
3. *Physiologie*
Dipl.-Phys. Dr. *Peter Finkenzeller*, Universität Erlangen-Nürnberg
4. *Medizin*
Prof. Dr. *Gustav Wagner*, Deutsches Krebsforschungszentrum, Heidelberg
5. *Ingenieurwesen*
Dr. Ing. *Rüder Hartwig* und Dr. Ing. *Willi Rehwald*, IBM Deutschland, Sindelfingen
6. *Betriebswirtschaft*
Prof. Dr. *Peter Mertens*, Universität Erlangen-Nürnberg
7. *Volkswirtschaft*
Prof. Dr. *Hans-Jürgen Krupp*, Universität Frankfurt
8. *Recht*
Prof. Dr. *Wilhelm Steinmüller*, Universität Regensburg
9. *Öffentliche Verwaltung*
Dipl.-Ing. Dipl.-Wirtsch.-Ing. Dr. *Joachim Griese*, Universität Erlangen-Nürnberg
10. *Sozialforschung*
Prof. Dr. *Erwin Scheuch* und Dipl.-Kfm. *Ekkehard Mochmann*, Universität zu Köln
11. *Raumplanung, Verkehr und Umweltschutz*
Dipl.-Ing. Dipl.-Wirtsch.-Ing. *Jürgen Seggelke*, Bundesministerium für Bildung und Wissenschaft, Bonn
12. *Verteidigung*
Dipl.-Phys. Dr. *Norbert Franßen*, IABG, Ottobrunn
13. *Bildungswesen*
Prof. Dr. *Milos Lánský*, Forschungs- und Entwicklungszentrum für objektivierte Lehr- und Lernverfahren, Paderborn
14. *Zum langfristigen Wissenschafts- und Forschungsprogramm einer Angewandten Informatik*
Prof. Dr. *Adolf Adam*, Hochschule Linz

Vorwort und Einführung

Friedrich L. Bauer, der sich um die Einführung der Informatik in der Bundesrepublik Deutschland wesentliche Verdienste erworben hat, schrieb zusammen mit seinem Schüler *P. Deussen* zum Studium der Informatik:

„Im Studium der Informatik steht die Beschäftigung mit Rechenanlagen . . . im Vordergrund. Aus Gründen der geistigen Ökonomie und nach allgemeinen Prinzipien der Wissenschaft wird sich das Studium nicht mit einer Aufzählung der Rechneranwendungen begnügen dürfen und können. Statt dessen muß es versuchen, Gemeinsamkeiten in den verschiedenen Anwendungsgebieten aufzudecken, sozusagen abstrahierte Anwendungen finden. . . . Eine Hauptaufgabe der Informatik ist es also, Gemeinsamkeiten verschiedener Aufgaben herauszuarbeiten und abstrahiert vom konkreten Einzelfall zu beschreiben, damit sie für weitere Anwendungen erschlossen werden können“¹⁾).

Diese Aussage von *Bauer* und *Deussen* könnte man über das Grobkonzept eines Studienplanes hinaus als ein Wissenschaftsprogramm der Informatik deuten.

Die vorliegende kleine Schrift beschäftigt sich mit dem Gebiet der Angewandten Informatik, die – geht man einmal von den Datenverarbeitungs-Förderungsprogrammen der Bundesregierung aus – die jüngere Teildisziplin der Informatik ist. Zwar unterscheidet sie sich bis zu einem gewissen Grade von der Informatik (Kerninformatik), wie sie im ersten Satz des obigen Zitates von *Bauer* und *Deussen* gesehen wird, nämlich darin, daß bei der Angewandten Informatik die Beschäftigung mit Rechenanlagen nicht so sehr im Vordergrund der Betrachtung steht. Vielmehr wird – zumindest auf längere Sicht – der Computer eher nur einen, wenn auch ungeheuer wichtigen, Bestandteil größerer Informationssysteme darstellen, die ihrem Charakter nach Mensch-Maschinen-Systeme sind.

Die von *Bauer* und *Deussen* genannte Aufgabenstellung, nämlich Gemeinsamkeiten in den verschiedenen Anwendungsgebieten aufzudecken, gilt jedoch in vollem Ausmaß auch für die Angewandte Informatik.

Um dem so gesteckten Ziel näherzukommen, erscheint zunächst eine Materialsammlung des in den verschiedensten Teildisziplinen Erreichten und mehr noch der offenen Probleme unumgänglich. Eine solche Materialsammlung einzuleiten, ist eines der Anliegen dieses Bändchens. In Anbetracht der (etwa verglichen mit der Kerninformatik) noch viel weniger tiefgehen-

¹⁾ *Bauer, F. L.* und *P. Deussen*, Ein junges Fachgebiet mit Zukunft: Die Informatik. In: *Süddeutsche Zeitung* vom 16. 12. 1970, Nr. 300, Beilage.

den systematisch-theoretischen Durchdringung der Angewandten Informatik und deren geringerem wissenschaftlichen Reifegrad lassen sich allerdings dabei Aufzählungen vorerst kaum vermeiden.

Ein zweites Anliegen ist es, eine Brücke zwischen Fachleuten der verschiedensten Disziplinen zu schlagen. Dabei geht es nicht nur um die Verständigung zwischen Computeranwendern in verschiedenen Teilbereichen, sondern auch um die Kommunikation zwischen dem Informatiker und dem Fachmann der Disziplin, auf die Forschungsergebnisse und Arbeitsergebnisse der Informatik angewendet werden sollen. In diesem Zusammenhang wollen wir vor allem auch darauf hinweisen, daß es oft genug nicht nur um „Anwendungen“ der Informatik im engeren Sinne geht, sondern daß diese junge Disziplin geeignet ist, das Wesen traditioneller Wissenschaftsbereiche tiefgreifend zu beeinflussen, wie es etwa in dem Beitrag von *Krupp* über Beziehungen zwischen Informatik und Volkswirtschaftslehre, in dem von *Scheuch* und *Mochmann* über Informatik und Sozialforschung oder in dem von *Steinmüller* über Rechtsinformatik zum Ausdruck kommt. Andererseits mögen die Ansprüche der Anwendungsdisziplinen auch in vielfältiger Weise auf das Forschungs- und auch das Lehrprogramm der Kerninformatik rückwirken.

Ein drittes Anliegen der vorliegenden kleinen Schrift ist es, dem interessierten Laien und auch solchen Persönlichkeiten, die die Ausbildungs- und Forschungspolitik in der Bundesrepublik mit zu verantworten haben, einen Überblick über den Stand der Computeranwendungen in einer größeren Zahl von Disziplinen zu vermitteln.

Dies ist in Verbindung mit den langfristigen Programmen zur Förderung der Datenverarbeitung und im Zusammenhang mit den Plänen zur Einrichtung neuer, moderner Studiengänge wichtig, die eines Tages dazu beitragen sollen, talentierte Nachwuchskräfte für die schwierige Aufgabe der automatischen Informationsverarbeitung in den verschiedensten Bereichen des wissenschaftlichen und des gesellschaftlichen Lebens heranzubilden.

Die vorliegende Schrift konnte nur als Sammelveröffentlichung entstehen, denn es ist schlechthin undenkbar, daß ein einzelner Autor die zahlreichen hier vertretenen Disziplinen hinreichend übersieht.

Der Herausgeber ist allen beteiligten Autoren für ihre Mitwirkung zu großem Dank verpflichtet, unter anderem auch dafür, daß sie sich dem problematischen Zwang unterzogen, außerordentlich umfangreiche Themen auf wenigen Seiten abzuhandeln.

Schließlich gilt mein Dank auch meinem Mitarbeiter *Horst Zwirner* für die Unterstützung der redaktionellen Arbeiten.

Angewandte Informatik in einzelnen wissenschaftlichen Disziplinen und Bereichen

1. Linguistik

Von Herwig Krenn

1.1 Einleitung

Die linguistische Informatik umfaßt alle Anwendungen der elektronischen Informationsverarbeitung, die unter der Kontrolle eines Programmes stehen, das Prinzipien der linguistischen Theorie beinhaltet. Andere Termini dafür sind Computational Linguistics, Automatic Language Processing und Mechanolinguistics.

Die Verbindung von Informatik und natürlicher Sprache geht hauptsächlich auf die Bemühungen um das Problem der automatischen Übersetzung zurück. Heute ergibt diese Verbindung ein weitverzweigtes Gebiet, von dem hier nur einige wichtige Ausschnitte behandelt werden sollen.

1.2 Document and Information Retrieval

Unter diesem englischen Terminus versteht man die Wiedergewinnung einer mittels natürlicher Sprache ausgedrückten Information. Wenn es sich um einen bibliographischen Hinweis (Angabe) handelt, spricht man von Document Retrieval¹), in allen anderen Fällen von Information Retrieval. Die sogenannte Informationsexplosion unserer Zeit verlangt nach Möglichkeiten, die unüberschaubare Flut an Information, die gleichzeitig an allen möglichen Orten der Welt in sprachlicher Formulierung angeboten wird, nachträglich und schnell verfügbar zu machen. Besonders Bibliotheken werden täglich immer mehr mit dem Problem konfrontiert, die in den gesammelten Schriftstücken enthaltene Information zu jedem beliebigen Zeitpunkt und unter beliebigen Gesichtspunkten „wiederauffinden“ zu lassen. Zur Lösung dieses Problems bietet sich der Einsatz der EDV an. Man benötigt ein System, das in natürlicher Sprache kodierte Information behandeln kann. Die Anforderungen, die an ein im Language Processing versiertes elektronisches System gestellt werden, weisen ganz verschiedene Schwierigkeitsgrade auf. Die Schwierigkeiten sind noch relativ einfach, solange rein sprachliche Ausdrücke (linguistic expressions) zu bearbeiten sind; sie werden aber immer größer und sind größtenteils noch ungelöst, wenn

¹) Verschiedene Document-Retrieval-Verfahren werden zur Zeit auf einigen Gebieten bereits erfolgreich eingesetzt und erprobt. Gewöhnlich werden die „Anfragen“ in Form von Schlüsselwörtern gestellt, von denen eine bestimmte Menge vorgegeben ist. Die Maschine sucht sodann in ihrem Katalog, in dem jedes Schriftstück des betreffenden Gebietes vermerkt ist, anhand der Schlüsselwörter die relevanten bibliographischen Angaben heraus.

vom Computer verlangt wird, den durch sprachliche Formung bezeichneten Inhalt (content) zu behandeln.

Im folgenden werden verschiedene der EDV bislang abverlangte Leistungen vorgeführt und in ihrer Problematik kurz diskutiert. Beginnen wir mit den Arbeiten aus dem Bereich des Language Processing, die der Computer noch relativ leicht bewältigen kann:

1.2.1 Herstellung von Wortindices und Konkordanzen

Unter Wortindex versteht man die alphabetisch geordnete Auflistung aller Wörter eines Textes, wobei jedes Wort eine Referenzliste zu allen seinen Vorkommen im Text erhält. Als Referenzsysteme benutzt man sehr häufig Paragraph-, Seiten- und Zeilenangaben. Natürlich ist der Aufbau der Referenzsysteme von der Art des Textes abhängig.

Bevor das erste Wort eines Textes in den Computer eingelesen wird, steht der Paragraph-, Satz-, Seiten- etc. und Wortzähler auf Null. Der Wortzähler fügt nach jedem eingelesenen Wort eine 1 hinzu, bis das Satzende erreicht ist. Bei diesem Punkt wird die Satzählung auf 1 gestellt, die Wortählung beginnt wieder bei Null. Analog wird bei der Zählung von Paragraphen, Kapiteln etc. verfahren.

Von den bereits vorliegenden Indexing Systems seien besonders drei erwähnt: Die „depth-indexing“-Technik von *Simmons & McConlogue* [21], die Programme von *Clayton* [4] und *Smith* [22].

Eine Konkordanz ist im wesentlichen dasselbe wie ein Wortindex, nur mit dem Unterschied, daß zusätzlich zur Referenzliste mit jedem Wort ein Kontext seines Vorkommens mitgeliefert wird. Dabei wird als unmittelbarer Kontext sehr oft die Zeile verstanden, in der das Wort vorkommt.

Es versteht sich von selbst, daß Konkordanzen für den interessierten Benutzer weit größeren Wert haben als Wortindices. Der Benutzer einer Konkordanz braucht nicht mehr alle Stellen des Textvorkommens aufsuchen, um ein bestimmtes Phänomen studieren zu können. (Weiterführende Literaturangaben findet der Leser bei *Borko* [2/S. 211 ff.].)

1.2.2 Stilistische Analyse

Bei der stilistischen Analyse von Texten wird vom Computer ebenso wie bei der Herstellung von Wortindices und Konkordanzen lediglich eine Bearbeitung eines sprachlichen Ausdrucks verlangt. Die Bedeutung der sprachlichen Zeichen spielt keine Rolle.

Stil wird hier verstanden als das Auftreten von Mustern in frei gestaltbaren Vorgängen. Auf sprachliche Gestaltung übertragen heißt dies, Stil ist eine des öfteren oder sehr oft wiederkehrende, ja schablonenartige bestimmte

Formung des sprachlichen Ausdrucks. Die stilistische Analyse soll diese jeweilige bestimmte Formung eruieren.

Bevor ein stilistisches Phänomen gesichtet und als solches erklärt werden kann, bedarf es der Überprüfung von sehr viel Sprachmaterial. Es versteht sich von selbst, daß die Elektronik hier wertvolle Hilfe leisten kann. Da sich Stilistika auf der ganzen Breite sprachlicher Gestaltung etablieren können, stehen viele der linguistischen Computertätigkeiten im Dienste der stilistischen Analyse.

Stilistische Phänomene müssen nicht notwendig statistisch ermittelt werden, aber es läßt sich beobachten, daß sie sehr oft von implizit statistischen Prozeduren abhängig sind. Damit umschließt die Computational Stylistics die Frequenzzählung eines jeden Wortes im Text ebenso wie die Angabe seiner genauen Position und natürlich die Zählung aller möglichen Positionen.

Stilistische Analysen haben sehr oft die Ermittlung des Autors für einen bestimmten Text zum Ziele.

Beispielsweise haben *Mosteller* und *Wallace* [17] ein statistisches Meßverfahren, das sogenannte *Bayes*-Theorem angewandt, um die „Federalist Papers“, die zwischen 1787 und 1788 von Alexander Hamilton, James Madison und John Jay verfaßt worden sein sollen, auf diese drei Autoren zu verteilen. Auf statistische Eigenheiten bestimmter Wortvorkommen stützt sich auch *Morton* [16] bei der Entscheidung der Frage, welche der Paulus-Briefe wirklich vom Apostel Paulus verfaßt wurden.

Wegen der riesigen Materialmasse, deren Untersuchung einer jeden stilistischen Analyse vorausgehen muß, bietet sich die computerunterstützte Speicherung und Wiederauffindung von sprachlichen Daten an. Man denkt heute auch bereits an bedeutungsvolle Stilanalysen politischer Reden oder an den Einsatz von Computern für die Analyse von defektem Sprachmaterial, die Psychologen und Psychiatern wertvolle Einsichten vermitteln kann. Auch für die Kriminalistik kann die computerunterstützte Stilanalyse Bedeutung erlangen, etwa bei der Bestimmung der Urheberschaft von Erpresserbriefen usw. (Weiterführende Literaturangaben enthält die Veröffentlichung von *Borko* [2/S. 211 ff.].)

1.2.3 Automatische Indizierung (*Automatic Indexing*)

Unter „Indexing“ versteht man die Zuweisung einer Menge von Schlüsselwörtern oder Deskriptoren an einen bestimmten Text. Es geht darum, den Inhalt eines Textes durch Schlüsselwörter zu speichern und damit ein Mittel zur Wiederauffindung des Inhaltes zur Verfügung zu haben. Computer sollen beim *Automatic Indexing* genau jene zeitraubende Arbeit verrichten, die heute noch Bibliothekare und Spezialisten tun. Da es sich um die Bearbeitung von Textinhalten handelt, ist die Programmierung die-

ses Language Processing ungleich schwieriger als die Bearbeitung sprachlich-formaler Ausdrücke.

Zur Lösung des Problems wurden bislang besonders folgende vier Techniken erforscht:

1. das Statistical Indexing
2. das Permutation Indexing
3. das Citation Indexing
4. das Association Indexing

Zu 1: Dem statistisch vorgehenden Indexing liegt die Annahme zugrunde, daß mit dem häufigeren Gebrauch eines Wortes in einem Text die Wahrscheinlichkeit wächst, daß das betreffende Wort ein relevanter Indikator für den Inhalt des Textes bzw. für den behandelten Gegenstand ist.

Zu 2: Das Permutation Indexing geht von der Auffassung aus, daß der Autor eines wissenschaftlichen Schriftstückes im allgemeinen versuchen wird, einen informativen Titel zu gebrauchen. Somit müßte der Titel des Schriftstückes mehr Hinweise auf den behandelten Gegenstand enthalten als irgendein Satz des Schriftstückes. Beim Permutation Indexing wird die Selektion der Schlüsselwörter auf die Wörter des Titels beschränkt. Durch die Permutation wird der Titel derart umgestellt, daß jedes signifikante Wort der Reihe nach in eine Index-Wort-Position in einer Index-Zeile gebracht wird. Ein signifikantes Wort kann jedes Wort sein, es sei denn es findet sich auf einer im Computer gespeicherten Verbotliste, weil es für sich allein nicht aussagekräftig ist. Auf dieser Verbotliste stehen beispielsweise die Artikel, Präpositionen, Konjunktionen.

Zu 3: Zur Technik des Citation Indexing führte die Entdeckung, daß die Ähnlichkeit der bibliographischen Referenzen (Zitate) zweier Schriftstücke auch Ähnlichkeit hinsichtlich des behandelten Gegenstandes impliziert. Kessler [11] hat diese Beziehung „bibliographic coupling“ genannt. Er stellte die Hypothese auf, daß wissenschaftliche Texte eine bedeutungsvolle Beziehung zueinander haben, wenn sie eine oder mehr bibliographische Referenzen gemeinsam haben.

Zu 4: Der menschliche Geist speichert Ideen nicht alphabetisch geordnet, sondern nach bestimmten assoziativen Verknüpfungsgesetzen. Beim Association Indexing wird eine analoge Organisation angewandt. (Auch hierzu empfehlen wir die Literaturangaben bei Borko [2/S. 123 ff.].)

1.2.4 Automatische Klassifizierung (*Automatic Classification*)

Es ist die Aufgabe eines Klassifikationssystems, eine bestimmte Materialmasse systematisch zu gruppieren. Da der Computer nur Wörter „inhaltlich“ behandeln kann, muß ein automatisches Klassifikationssystem auf statistischen Prinzipien beruhen. Schriftstücke werden vom EDV-System nach den Wörtern, die sie enthalten, klassifiziert. Man kann bei wissen-

schaftlichen Dokumenten davon ausgehen, daß in Schriftstücken, die über verschiedene Themen handeln, verschiedene Wortmengen gebraucht werden. (Diese Annahme gilt allerdings nicht generell für sprachliche Aussagen. Man kann zum Beispiel mit verschiedenen Wörtern auch dasselbe ausdrücken.) Somit können wissenschaftliche Schriftstücke nach Ähnlichkeiten oder Differenzen im gebrauchten Wortschatz klassifiziert werden. (Weiterführende Literaturangaben finden Sie wiederum bei *Borko* [2/S. 123 ff.])

1.2.5 Automatisches Extracting and Abstracting

Besonders Bibliotheken sind daran interessiert, von den gesammelten Schriftstücken wesentliche Auszüge, sogenannte Resümees, zur Verfügung zu haben. Bislang haben von Rechenanlagen produzierte Abstracts den Nachteil, daß sie nur aus einzelnen Sätzen des Originaltextes bestehen, die nach bestimmten Kriterien extrahiert werden. Bei den extrahierten Sätzen handelt es sich zwar um repräsentative Sätze des Originals, aber man ist dennoch bestrebt, ein Computerprogramm zu entwickeln, das befähigt ist, ein Schriftstück zu lesen und darüber ein Resümee in konventionellem Prosa-Stil zu schreiben. Dieses Ziel setzt aber voraus, daß das Problem der Spracherzeugung durch den Computer gelöst ist. Davon ist man aber noch weit entfernt.

Zum Problem des Abstracting gibt es bereits viele Forschungsarbeiten (*Luhn-Studie*, *Baxendale-Studie*, *Oswald-Studie*, *Acsi-Matic-Studie*, *Ramo-Wooldridge-Studie* u. a.). (Hierzu enthält die Schrift von *Borko* ebenfalls weiterführende Literaturangaben [2/S. 177 ff.])

1.2.6 Content Analysis (Inhaltsanalyse)

Um die Semantik eines Textes behandeln zu können, muß der Computer vorläufig noch auf einzelne Schlüsselwörter oder Schlüsselsätze des Textes zurückgreifen. Auf diesem schwierigen Gebiet gibt es noch keine zufriedenstellenden Ergebnisse (vgl. neuere Lösungsversuche von *Laffal* [14], *Ströbl* [23], *Feitscher* [5]).

1.3 Automatische (Maschinelle) Übersetzung

Unter automatischer Übersetzung versteht man die „Übersetzung einer gegebenen natürlichen (Quellen-)Sprache in eine andere (Ziel-)Sprache mit Hilfe elektronischer Datenverarbeitungsmaschinen“ [12].

Für keinen Bereich der Computerlinguistik wurde soviel Energie aufgewendet wie für die Bewältigung der automatischen Übersetzung. Die automatische Übersetzung gliedert sich in folgende Etappen:

1.3.1 Automatische Analyse

Input der automatischen Analyse ist der Text der Quellsprache. Dieser Text kann gesprochen oder geschrieben vorliegen. Die Überführung gesprochener Sprache in geschriebene Sprache, die für eine automatische Simultanübersetzung von großer Wichtigkeit wäre, kann noch nicht durchgeführt werden. Noch sind die relevanten Lauteigenschaften und ihre Entsprechungen auf der Ebene der Schriftzeichen nicht genügend erforscht. Schriftlich fixierte Texte hingegen werden in absehbarer Zeit durch sogenannte optische Leser direkt in die Elektronik eingelesen werden können, und zwar trotz typographischer Unterschiede zwischen einzelnen Texten.

Der eingegebene Text ist nach einem bestimmten Analyseprogramm zu bearbeiten. Diese Operation wird als „parsing“ bezeichnet. Die Instruktionsvorschriften an das Parsing-Programm müssen selbstverständlich formal sein. Zur Durchführung eines dem Computer angepaßten Parsing-Programmes wurden bereits zahlreiche Strategien entwickelt (vgl. *Hays* [9/S. 106 ff.] und *Varga* [24]). Ziel des Parsing ist es, dem eingegebenen Text eine Reihe von grammatischen Informationen abzugewinnen, die für das Gelingen der Übersetzung benötigt werden. Solche grammatischen Informationen sind morphologische, syntaktische und semantische Aussagen über die Struktur des eingegebenen Textes. Demnach vollzieht sich das Parsing in drei Schritten:

1. morphologische Analyse
2. syntaktische Analyse
3. semantische Analyse

Zu 1: Morphologische Analyse:

Die Wörter des eingegebenen Textes werden zuerst einer morphologischen Analyse unterworfen. Aufgabe der morphologischen Analyse ist es, die Wörter des Eingabetextes mit den Wörtern des im Computer gespeicherten Wörterbuchs der Quellsprache zu vergleichen beziehungsweise bei diesem Vergleich bestimmte Differenzen zu ermitteln. Bei den im Wörterbuch gespeicherten Wörtern handelt es sich um Wortstämme ohne jeden morphologischen Zusatz. Es liegen bereits einige für morphologische Prozesse funktionierende Analyse-Algorithmen vor (vgl. *Ramo-Wooldridge* [19] und *Chapin* [3]).

Zu 2: Syntaktische Analyse:

Unter Syntax versteht man das System von Regeln, das die Bildung von Strukturen sogenannter wohlgeformter Sätze festlegt. Die syntaktische Struktur ist die Menge der Beziehungen zwischen den Wörtern eines Satzes bzw. Wörtern und größeren Einheiten eines Satzes. Ein Algorithmus, der einen Text bzw. Sätze syntaktisch analysieren soll, muß diese Beziehungen

identifizieren und syntaktisch klassifizieren, das heißt ihre hierarchischen Verhältnisse bestimmen.

Für die syntaktische Analyse existieren bereits zahlreiche Computerprogramme mit verschiedener theoretischer Basis, z. B.

der *Cocke*-Algorithmus in mehreren Versionen,
der *Kuno-Oettinger* Syntactic Analyzer,
die *Mitre* Syntactic Analysis Procedure,
der Parser von *Martin Kay*,
die „Fulcrum“-Methode von *Garvin*,
das Saarbrücker Verfahren und die Analyseprogramme
der Universität Freiburg i. Br.

Die syntaktische Strukturanalyse hängt vom zugrunde gelegten Grammatiktyp ab. Heute werden vielfach generative Grammatiken für die Darstellung der Quellsprache verwendet. Damit kann der syntaktische Analysealgorithmus aus den Regeln dieser Grammatik abgeleitet werden. Aufgabe der syntaktischen Analyse ist es dann, zu einem gegebenen Satz die Ableitungen zu liefern.

(Literatur hierzu: *Rhodes* [20], *Oomen* [18] und bei *Borko* [2/S. 322f.])

Zu 3: Semantische Analyse:

Die Ermittlung semantischer Information aus dem Input-Text gestaltet sich äußerst schwierig. Die semantische Analyse, die Informationen über die lexikalischen Einheiten (Wörter) durch eine Menge von semantischen Merkmalen repräsentiert, ist noch nicht genügend entwickelt. Die Speicherung semantischer Merkmale zusammen mit Regeln, die anzeigen, welche Kombinationen der semantischen Merkmale zulässig sind, könnte zum Beispiel dazu beitragen, daß der Computer bei der Übersetzung mehrdeutiger Wörter die richtige Auswahl trifft. (Vgl. zu diesem Problem *Hays* [9/S. 188ff.] und *Wilks* [25].)

1.3.2 Transferteil

Unter „Transferteil“ verstehen wir die eigentliche Etappe der Übersetzung, das heißt die Umsetzung der Analyseergebnisse des Parsing in die Zielsprache. Das Resultat des Transferteiles ist eine äquivalente Umsetzung in Ausdrucksmittel der Zielsprache, allerdings handelt es sich hierbei noch um eine unvollkommen geformte, rohe Sprache, die noch weiter geändert werden muß.

1.3.3 Automatische Synthese

Die automatische Synthese bildet die abschließende Operation des Übersetzungsprozesses. Ein Synthesealgorithmus erzeugt den Text in der Zielsprache, dabei wird beispielsweise die richtige Wortfolge hergestellt. Als

Synthesealgorithmen eignen sich besonders generative Grammatikmodelle. (Vgl. dazu *Woolley* [26] und *Maas* [15].)

1.3.4 Zwischensprache

Die Analysealgorithmen werden immer in Hinblick auf bestimmte Zielsprachen formuliert. Da für die wechselseitige Übersetzung in mehrere Sprachen die erforderlichen Algorithmen immer zahlreicher werden, versucht man heute, verschiedene Typen von Zwischensprachen zu entwickeln. Zwischensprachen machen die Analyse der Quellsprache von der Zielsprache unabhängig. (Vgl. *Andreev* [1].)

1.4 Informatik im Dienste der linguistischen Forschung

Der Einsatz der Informatik bietet sich heute auch für die linguistische Forschung selbst an. Eine Grammatik ist mathematisch gesehen ein so komplizierter Kalkül, daß es kaum mehr möglich ist, die aus ihm ableitbaren Ausdrücke zu beherrschen. Die EDV eignet sich für die Überprüfung linguistischer Aussagen, zum Beispiel der Implikationen grammatischer Regeln. Die Grammatik wird als Rechenprogramm in einem Computer realisiert. Auf diese Weise lassen sich Grammatiken mechanisch testen, und der Linguist wird beim Schreiben von Grammatiken auf bedeutungsvolle Weise unterstützt. (Vgl. dazu *Friedman* [6] und *Gross* [8].)

Literatur

- [1] *Andreev, N.*: The Intermediary Language as the Focal Point of Machine Translation. In: Booth, A. D. (Hrsg.): Machine Translation, Amsterdam 1967, S. 3 ff.
- [2] *Borko, H.*: Automated Language Processing, New York u. a. 1967.
- [3] *Chapin, G.*: On the Syntax of Word-derivation in English, Bedford 1967 (Bericht der MITRE Corp.).
- [4] *Clayton, T.*: The Preparation of Literary Text for multiple automated Studies: Comprehensive Identification and the Provision of Discriminants. In: IBM (Hrsg.), Proceedings of IBM Literary Data Processing Conference 1964, S. 171 ff.
- [5] *Feitscher, W.*: Zur Bewertung von Texten der deutschen Sprache, Dokumentation/Information, Heft 13. VI. Kolloquium über Information und Dokumentation, Themenkreis 1, TH Ilmenau 1969.
- [6] *Friedmann, J., T. H. Bredt, R. W. Doran, B. W. Pollack und T. S. Martner*: A Computer Model of Transformational Grammar, New York 1971.
- [7] *Garvin, P. L.*: Computer Participation in Linguistic Research, In: Language 38 (1962), S. 385 ff.
- [8] *Gross, L. N.*: A Computer Program for Testing Grammars on-line, Bedford 1968 (Bericht der MITRE Corp.).
- [9] *Hays, G.*: Introduction to Computational Linguistics, New York 1967.
- [10] *Kay, M. und T. W. Zieve*: Natural Language in Computer Form, RAND-Bericht RM-4390 (1965).

- [11] *Kessler, M. M.*: Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing. In: *American Documentation* 16 (1965), S. 223 ff.
- [12] *Klaus, G.*: Wörterbuch der Kybernetik, Frankfurt – Hamburg 1969.
- [13] *Klein, W.*: Bibliographie zur Computerlinguistik. In: *Linguistische Berichte* 11 (1971), S. 101 ff.
- [14] *Laffal, J.*: Total or Selected Content Analysis, Preprint Nr. 24, International Conference on Computational Linguistics, Sänga-Säby 1969.
- [15] *Maas, H.-D.*: Die Synthese deutscher Sätze im Zusammenhang mit maschineller Sprachübersetzung. In: *Muttersprache* 9/10 (1969), S. 274 ff.
- [16] *Morton, A. Q.*: A Computer challenges the Church. In: *The Observer* vom 3. Nov. 1963, S. 21.
- [17] *Mosteller, F.* und *D. L. Wallace*: Inference in an Authorship Problem. In: *Journal of the American Statistical Association* 58 (1963), S. 275 ff.
- [18] *Oomen, U.*: Automatische syntaktische Analyse, Mouton 1968.
- [19] *Ramo-Wooldridge* (Hrsg.): *Machine Translation Studies of Semantic Techniques*. AF 30(602)-2036, Technical Report Nr. 1 to Rome Air Development Center, Griffis AFB, New York – Los Angeles 1960.
- [20] *Rhodes, I.*: A new Approach to the mechanical Syntactic Analysis of Russian. In: *Mechanical Translation* 6 (1961), S. 33 ff.
- [21] *Simmons, R. F.* und *K. L. McConlogue*: Maximum Depth Indexing for Computer Retrieval of English Language Data. In: *American Documentation* 14 (1963) 1, S. 68 ff.
- [22] *Smith, P. H.*: A Computer Program to generate a Text Concordance. In: *IBM* (Hrsg.): *Proceedings of IBM Literary Data Processing Conference* 1964, S. 113 ff.
- [23] *Ströbl, A.*: Ein Vorschlag zur automatischen Bestimmung der Bedeutung eines Textes. In: *Muttersprache* 9/10 (1969), S. 266 ff.
- [24] *Varga, D.*: Problems of Parsing Systems, Preprint Nr. 61, International Conference on Computational Linguistics, Sänga Säby 1969.
- [25] *Wilks, Y.*: Interactive Semantic Analysis of English Paragraphs, Preprint No. 8, International Conference on Computational Linguistics, Sänga-Säby 1969.
- [26] *Woolley, G. H.*: Automatic Text Generation, Preprint No. 37, International Conference on Computational Linguistics, Sänga-Säby 1969.

2. Physik

Von Hans-Jürgen Trebst

2.1 Einleitung

Es liegt im Wesen der Physik, mit mathematisch formulierten Theorien umzugehen, Meßwerte zu verarbeiten und sie mit den zugehörigen Theorien in Verbindung zu bringen. Demnach gehören die Durchführung von Rechnungen und damit auch die Anwendung von Rechenhilfsmitteln seit jeher zur Ausübung der Physik. Der Einsatz von Datenverarbeitungsanlagen

hierfür stellt also gar keine Besonderheit dar, sondern nur die ganz normale Weiterentwicklung ihrer Arbeitsmethoden. Es gibt aber eine Reihe von Anwendungen, bei denen der Einsatz der elektronischen Datenverarbeitung in quantitativer oder in qualitativer Hinsicht wesentlich über die früheren Arbeitsmethoden hinausführt und mit denen wir uns in dieser Übersicht befassen wollen. Vorher werden wir uns aber auch einen Überblick über die wichtigsten vorkommenden Rechenmethoden verschaffen.

2.2 Lösen von Differentialgleichungen als typische Aufgaben der Informationsverarbeitung in der Physik

2.2.1 Einführung in die Rechenmethoden

Die Berechnung von Wurfbahnen ist eine der Grundaufgaben der Mechanik, und die Beschäftigung mit diesem Problem gab der Mathematik und mathematischen Physik schon seit jeher fruchtbare Anregungen. Das Verhalten des bewegten Körpers wird durch eine Differentialgleichung beschrieben. Löst man die Differentialgleichung, so erhält man die Bahn des Körpers. Für einen vereinfachten Fall lernt man in der Schule als Lösung die Wurfparabel kennen. Wenn die vereinfachenden Annahmen nicht mehr zulässig sind, steigt der Rechenaufwand stark an, und die Lösung läßt sich nicht mehr in geschlossener Form angeben, sondern nur noch numerisch. Es ist daher verständlich, daß die Berechnung von Geschosßbahnen mit zu den Gebieten zählte, die gegen Ende des zweiten Weltkrieges in den USA die Entwicklung elektronischer Rechenmaschinen vorwärts drängte (die anderen waren Berechnungen von Kernreaktoren und „Atom“-bomben und Probleme der Wettervorhersage). Differentialgleichungen treten in der Physik und in der Technik noch an vielen anderen Stellen als „Bewegungsgleichungen“ auf, die Veränderungen irgendeines Zustandes beschreiben. Ihre Lösung ist also eine der typischen Aufgaben in der ganzen Physik. Die Bahn eines Geschosses wird nach den Newton'schen Gesetzen beschrieben durch die Differentialgleichung für seine Geschwindigkeit v

$$\frac{dv}{dt} = \frac{F}{m} \quad v = v(t)$$

und die Beziehung

$$\frac{dx}{dt} = v \quad x = x(t)$$

für seinen Ort x . (Tatsächlich sind es nicht nur 2, sondern $3 \cdot 2$ Gleichungen für die 3 Raumrichtungen, d. h. v , x und F sind Vektoren. Dazu kommen weitere $3 \cdot 2$ Gleichungen für die Drehbewegung.) Dabei ist F die Kraft, die jeweils auf das Geschosß der Masse m wirkt. Sie setzt sich aus der Erdranzie-