

Möglichkeiten interdisziplinärer Zusammenarbeit am Beispiel von Mathematik und Medizin

Von H. Werner *)

Einführung

Das in Münster im Februar 1974 abgehaltene Symposium "Computer in der Medizin" zeigt an einer Anzahl von Beispielen die Ergebnisse des erfolgreichen Einsatzes mathematischer Hilfsmittel in der Medizin. In diesem Beitrag sollen die Bedingungen untersucht werden, die für eine interdisziplinäre Zusammenarbeit meiner Meinung nach notwendig sind. Ich möchte verfolgen, wie interdisziplinäre Zusammenarbeit entsteht, wie sie abläuft und welche Voraussetzungen für ihren Erfolg eine Rolle spielen.

Dabei geht es also mehr um methodische Betrachtungen als um einzelne wissenschaftliche Resultate. Nur um meine Vorstellungen zu verdeutlichen, werde ich in einem speziellen Fall den konkreten Ablauf interdisziplinärer Zusammenarbeit beschreiben. Die Berichte über die Ergebnisse der Zusammenarbeit möchte ich den Fachvertretern der Medizin überlassen.

Interdisziplinäre Zusammenarbeit ist ein Schlagwort, das immer wieder in den Diskussionen um die Neugestaltung der wissenschaftlichen Ausbildung und Forschung auftaucht. Die progressiven Studenten rufen danach, suchen die Verbindung zwischen Praxis und Theorie, fordern Berufsbezogenheit. Deshalb sind an gewissen Universitäten Zentren gegründet worden, die der interdisziplinären Zusammenarbeit dienen sollen. Man könnte daraus den Schluß ziehen, daß es früher weniger Möglichkeiten für interdisziplinäre Zusammenarbeit gegeben hat als heute. Um dies zu analysieren, wollen wir zunächst einen Blick auf die frühere und heutige Struktur der Universität Münster werfen, die als Beispiel für andere Hochschulen stehen soll.

Früher waren die Institute bzw. Kliniken zu Fakultäten zusammengefaßt, und die Universitätsspitze wurde von einem Rektor, einem Kurator oder Kanzler und dem Senat gebildet. Heute ist die Struktur reichhaltiger. Es gibt Betriebseinheiten (früher Institute oder Kliniken), die nach fachlichen Gesichtspunkten zu Fachbereichen (beispielsweise Klinische Medizin oder Mathematik) zusammengefaßt sind; diese wiederum entsenden ihre Vertreter in die Fakultäten (Medizinische Fakultät, Naturwissenschaftliche Fakultät), in denen allerdings nur übergeordnete Verwaltungsfragen, z.B. Fragen der Prüfungsbestimmungen, behandelt werden. An der Spitze der Universitätshierar-

chie steht nach wie vor ein Rektorat und ein Senat, der allerdings im Gegensatz zu früher paritätisch durch Vertreter der verschiedenen Gruppen der Universität (Lehrstuhlinhaber, Nichtordinarien, Mittelbauvertreter, Studenten) besetzt wird. Da die Mitglieder der über die einzelnen Fächer hinausgreifenden Gremien nur einzelne Gruppenvertreter sind, werden die Wissenschaftler verschiedener aber benachbarter Fachrichtungen kaum noch zusammengeführt. So kommt es, daß heute die Fachbereichsgrenze oft gleichzeitig die Grenze der wissenschaftlichen Kontakte ist. Leider erfährt man als Mathematiker kaum, wer als neuer Kollege in der Physik oder Chemie an die Universität Münster gekommen ist, und welche wissenschaftlichen Forschungsrichtungen in diesen Fachbereichen gegenwärtig besonders gepflegt werden. Die Vielzahl der Sitzungen in den universitären Selbstverwaltungsgremien raubt die wenige für Forschungstätigkeiten verbleibende Zeit; außerdem entsteht aus diesen und anderen Gründen ein Mangel an persönlichen Kontakten. Diese aber bilden meiner Meinung nach die Voraussetzung für eine erfolgreiche wissenschaftliche Zusammenarbeit. Früher gaben die im Verhältnis zu der Zahl der heutigen Zusammenkünfte seltenen Fakultätssitzungen ebenso wie andere größere Veranstaltungen, in denen sich Professoren, Dozenten, Assistenten und Studenten trafen, Gelegenheit zu wissenschaftlichen Gesprächen. Es gab die Habilitationskolloquien, die auch dem Außenstehenden einen Eindruck von den Forschungsschwerpunkten in den Nachbardisziplinen vermittelten.

Der Verlust dieser Möglichkeiten, wissenschaftliche Kontakte anzuknüpfen und zu vertiefen, läßt sich auch nicht durch das sogenannte Projektstudium beheben, das an einer norddeutschen Universität praktiziert werden soll. Denn Zusammenarbeit kann nicht durch Organisation und Verordnung von oben erreicht werden.

Nach dieser im ganzen etwas negativen Bilanz muß man sich fragen, welche Ansatzpunkte heute für die interdisziplinäre Zusammenarbeit vorhanden sind.

Aus meiner eigenen Erfahrung sehe ich sie einmal in den zentralen Einrichtungen, in denen auch heute noch Vertreter der verschiedensten Disziplinen zusammenkommen, und wo Veranstaltungen wie Vorträge und Seminare die Wissenschaftler verschiedenster Fachrichtungen zusammenzuführen vermögen.

Eine zweite Ansatzmöglichkeit wird organisatorisch durch die Sonderforschungsbereiche gegeben, die mit Unterstüt-

*) Referat anlässlich des Journalistenseminars "COMPUTER IN DER MEDIZIN" am 18./19. Febr. 1974

zung der Deutschen Forschungsgemeinschaft gegründet werden und an denen satzungsgemäß verschiedene universitäre Einrichtungen beteiligt sein müssen.

In Münster bestand die Gelegenheit, sowohl in Sonderforschungsbereichen als auch in einer zentralen Einrichtung - dem Rechenzentrum - Erfahrungen auf dem Gebiet interdisziplinärer Zusammenarbeit zu sammeln, und ich möchte hier versuchen, die sowohl fachlichen wie psychologischen Voraussetzungen dieser Zusammenarbeit zu beschreiben.

Die fachlichen Gegebenheiten

Die Mitarbeiter des Rechenzentrums in Münster sind zum größten Teil Mathematiker, es gibt einzelne Physiker, Ingenieure und Sprachwissenschaftler. Alle verfügen neben ihrem Fachwissen über Erfahrung mit elektronischen Rechenanlagen und deren Einsatzmöglichkeiten bei der Lösung wissenschaftlicher Probleme, kurz über umfangreiches Wissen in der praktischen Informatik. Deshalb soll im folgenden der Begriff Mathematik stehen für Reine Mathematik, Angewandte Mathematik, das sind insbesondere also mathematische Methoden zur Lösung konkreter Aufgaben, und praktische Informatik.

Außerdem wurden einige Diplomanden der Mathematik, die durch ihre Arbeit im Rechenzentrum oder Institut für Numerische und Instrumentelle Mathematik ihre Neigung zu den Anwendungen entdeckt hatten, von den Kliniken als Mitarbeiter übernommen.

Dies charakterisiert die eine Gruppe der Partner, über deren Verhältnis im folgenden zu berichten ist. Es ist schwer vorauszusagen, welche Gebiete der Mathematik oder Informatik von diesen Wissenschaftlern am stärksten benötigt werden; sicherlich gehören numerische Methoden und Statistik dazu. Es hängt von der Flexibilität des einzelnen Sachbearbeiters ab, ob er bei gegebenen Problemen die geeigneten Techniken der Mathematik zur Verfügung zu stellen vermag. Zuweilen lassen sich die anfallenden Aufgaben mit verhältnismäßig einfachen mathematischen Hilfsmitteln bereits lösen. Andere Probleme wiederum stellen uns vor sehr schwierige mathematische Aufgaben, ja, sie führen uns zum Teil zu Aufgabenstellungen, die bis jetzt in dieser Form überhaupt noch nicht behandelt worden sind.

Werfen wir einen Blick auf die Medizin, so habe ich als Außenstehender den Eindruck, daß die elektronische Datenverarbeitung seit einigen Jahren in diesem Zweig der Wissenschaft sehr stark im Vordringen ist. Für mich zeichnen sich dabei drei große Bereiche ab.

- 1) Die Rationalisierung der Verwaltung in den Kliniken. Die Großkliniken unterscheiden sich hierbei wohl kaum von irgendeinem anderen Großbetrieb mit sehr vielen Kunden, und es ist natürlich, daß der heutige Ratio-

alisierungstrend in der Verwaltung auch in den Kliniken seinen Niederschlag findet und hier zu guten Erfolgen führt.

- 2) Die Aufstellung von Patientendateien, bei denen neben den für die Verwaltung benötigten kommerziellen Daten der Patienten auch die klinischen Angaben gespeichert werden. Das Endziel wäre es, sämtliche über den Patienten vorhandenen Informationen abrufbereit in einem großen Computer zur Verfügung zu haben, der seinerseits sogar diese Information zu interpretieren vermag.

Pragmatische Ansätze sind in Münster schon seit einigen Jahren gemacht worden, und zwar von einigen Kliniken mit Unterstützung von Diplom-Mathematikern. Sie haben Dateien aufgestellt, die zunächst für verschiedene Forschungsaufgaben benutzt wurden. Mit Hilfe dieser Dateien und einem Auswahlprogramm kann man sich in kürzester Zeit einen Überblick über alle für eine spezifische Fragestellung interessanten Fälle (etwa Patienten mit gleichem Krankheitsverlauf) verschaffen. Eine zweite Anwendung, die in Münster seit langem eingesetzt ist, bildet die Einladung zu Nachuntersuchungen und Betreuung nach gewissen heimtückischen Krankheiten.

- 3) Ein weiterer Bereich ist die Mathematisierung medizinischer Forschungsaufgaben einschließlich ihrer Konsequenzen für die Diagnose und Therapie. Es gehört dazu auch die Verbesserung technischer Geräte, die durch eine mathematische Fundierung physikalischer Überlegungen oder auch chemischer Untersuchungen physiologischer Vorgänge vorangetrieben wird.

Zunächst einmal werden sehr oft die Prozesse der Lebensvorgänge im biomedizinischen Bereich durch Modelle simuliert, und es ist eine der uns besonders interessierenden Aufgaben, bei der Aufstellung solcher Modelle mitzuarbeiten. Bevor wir jedoch auf diesem Gebiet die Zusammenarbeit zwischen dem Rechenzentrum und den Kunden, d.h. den Benutzern des Rechenzentrums, die aus anderen wissenschaftlichen Bereichen der Universität kommen, weiter verfolgen, will ich die Aufgaben des Rechenzentrums aus meiner Sicht etwas näher fixieren.

Aufgaben eines Universitätsrechenzentrums

Die Aufgaben des Rechenzentrums sind Dienstleistungen im weitesten Sinne des Wortes. Leider wird dieser Ausdruck heute sehr oft abwertend gebraucht, im Gegensatz zum Begriff der Forschung, die mit einem gewissen Snobismus als eine vornehmere Tätigkeit angesehen wird. Ich versuche immer wieder deutlich zu machen, wie gerade der Dienst für andere und die Zusammenarbeit mit anderen ein großes Maß an Befriedigung zu bringen vermag. Auch die Lehre und die angewandte Forschung sind Teil der universitären Dienstleistung.

In einem Rechenzentrum kann man die folgenden Gruppen von Aufgaben unterscheiden.

- 1) Die Bereitstellung von Rechenkapazität auf geeigneter Hardware, d.h. einmal auf der Rechanlage selbst, sei sie digital oder analog, weiterhin auf den notwendigen peripheren Geräten, also den Ausgabegeräten, Zeichengeräten, Druckern und auf den Eingabegeräten, den Lochkartengeräten, die den Kunden in unserem Hause zur Verfügung stehen.
- 2) Die Hilfestellung bei der Vorbereitung auf den Gebrauch einer Rechanlage. Sie erfolgt durch Programmierkurse, weiter durch eine Programmberatung, die Hilfeleistung anbietet, wenn ein Anfänger mit seinem Programm nicht zurecht kommt, beispielsweise, wenn er Schwierigkeiten bei der Interpretation einer Fehlermeldung der Rechanlage hat, und schließlich in der Anleitung zum geschickten, d.h. insbesondere zum sparsamen Gebrauch der zur Verfügung gestellten Betriebsmittel. Daß dies Aufgaben des Rechenzentrums sind, ist eine überall in den Universitäten akzeptierte Auffassung. Nach meiner Meinung und auch der meiner Mitarbeiter geht aber eine gute Dienstleistung weit über dies hinaus, denn sie soll zum möglichst effektiven Einsatz der Rechanlage führen.
- 3) Es findet deshalb auch eine Problembberatung statt, in der auf Fragen des Wissenschaftlers hin, der das Rechenzentrum in Anspruch nimmt, größere wissenschaftliche Beiträge zur Lösung seines Problems geliefert werden. Ich halte diese Hilfestellung besonders dann für wichtig, wenn ein wissenschaftlicher Bereich für die Datenverarbeitung erst erschlossen werden muß. Üblicherweise sind dies heute Gebiete, in denen statt der formal logischen Arbeitsweisen andere Verfahren vorherrschend sind. Während wir im Bereich der Physik oder auch der Chemie heute in dieser Beziehung kaum noch zu helfen brauchen, es sei denn, daß es sich um große neue Fragestellungen handelt, so sind es neben der Medizin vor allem auch die Fachbereiche der Geisteswissenschaften, die bei der computergerechten Formulierung ihrer Probleme unsere Unterstützung in Anspruch nehmen. In diesem sogenannten nichtnumerischen Bereich liegt in Münster ebenfalls ein Schwerpunkt der EDV-Arbeit.

Es kommt bei der Problembberatung oft zu einer echten Teamarbeit, d.h. daß Wissenschaftler sehr verschiedener Fachgebiete ihr Fachwissen zusammenbringen, um etwas entstehen zu lassen, das mehr ist als nur die Summe der einzelnen wissenschaftlichen Beiträge.
- 4) Das Rechenzentrum führt schließlich eigene Projekte allein oder gemeinsam mit anderen Universitäts-Einrichtungen durch, die unter Umständen Pilotcharakter für andere Universitäten besitzen können. Eines dieser Projekte unseres Rechenzentrums ist die automatische

Übersetzung von Texten in Blindenschrift, die sowohl mathematischen wie formal-linguistischen Sachverstand der beteiligten Mitarbeiter verlangt. Ein anderes, von der DFG gefördertes Projekt wurde in Zusammenarbeit mit der Universitätsbibliothek zur Rationalisierung der Buchausleihe durchgeführt.

Die verschiedenen Bereiche der Dienstleistung setzen natürlich ein mehr oder weniger hohes fachliches Können voraus. Meine Erfahrung ist, daß dieses Können allein für eine erfolgreiche Teamarbeit nicht ausreicht, sondern es vielmehr noch auf andere Qualitäten ankommt.

Psychologische Bedingungen für erfolgreiche Teamarbeit

Eine erfolgreiche Team- bzw. Projektarbeit setzt bei den Beteiligten die folgende Haltung voraus:

- 1) Aufgeschlossenheit gegenüber Fragen, die außerhalb des eigenen Fachgebietes liegen und Bereitschaft, sich in ein fremdes Problem zunächst einmal einzuhören und anschließend so hinein zu versenken, daß man die Fragestellung für die EDV aufbereiten, d.h. formalisieren kann. Diese letzte Tätigkeit erfordert eine konstruktive Einstellung zum Problem, nicht etwa eine Abwehrhaltung, wie man sie bei manchem Fachwissenschaftler beobachtet, der froh ist, wenn er aufzeigen kann, daß die Mittel seines Faches nicht zur Lösung des an ihn herangetragenen Problems geeignet sind.
- 2) Eine entsprechende Bereitschaft wird beim Partner des Anwendungsgebietes vorausgesetzt. Er muß nämlich willens sein, den Gedankengängen des EDV-Beraters zu folgen und zu prüfen, ob mit der Formalisierung wirklich das von ihm gestellte Problem beschrieben wird. Daß der ins Rechenzentrum kommende Kunde bereit ist, sich dieser Aufgabe zu stellen, kann man in der Regel annehmen.
- 3) Auf beiden Seiten ist Geduld und das Bemühen nötig, die Fachsprache des anderen wenigstens so weit zu erlernen, wie es zu einer Kommunikation über das zur Diskussion stehende Problem notwendig ist, und umgekehrt dem Gesprächspartner die eigene ihm fremde Begriffswelt möglichst ohne viele Fachausdrücke zu verdeutlichen. Zusammenarbeit setzt, das halte ich für entscheidend, gegenseitige Achtung und die Anerkennung des geleisteten wissenschaftlichen Beitrages des Partners voraus, obwohl oder gerade weil man selbst darüber kein Urteil abzugeben in der Lage ist.

Ich möchte mir erlauben, diesen Gesichtspunkt auch negativ zu formulieren. Wenn man einen Mathematiker nur als Programmierer behandelt, der eine Rechnung durchzuführen hat, ohne daß er die zugrundeliegende Aufgabenstellung kennt, dann kann man keinen eigenständigen Beitrag von seiner Seite erwarten, dann entwickelt sich keine echte Zusammenarbeit. Die in Münster erzielten Erfolge beruhen wohl zumeist darauf, daß es zu einer intensiven Zusammenarbeit gekommen ist.

Entstehen und Verlauf einer Teamarbeit möchte ich hier an einem methodischen Beispiel erläutern, das typisch für eine solche Arbeit ist, wobei natürlich im einzelnen Abweichungen auftreten werden.

- 1) Das Entstehen einer Zusammenarbeit ist oft von Zufällen abhängig. Es kann sein, daß ein Kunde mit einer gezielten Frage zu uns kommt, Daten vorlegt und Unterstützung bei der Auswertung haben möchte. Es kann aber auch sein, daß das aus der Sicht der EDV eigentlich zentrale Problem mehr durch einen Nebensatz angesprochen wird, als daß es den ursprünglichen Grund der Frage darstellt. Einmal legte mir ein Kollege Meßdaten vor, für deren Auswertung er sich bereits ein statistisches Verfahren hatte empfehlen lassen. Dieses führte aber dazu, daß die Mittelwerte nicht einmal im Vertrauensbereich, der berechnet wurde, lagen. Dies schien ihm unmöglich, und so kam es zu einer Rücksprache bei mir. Inhaltlich ließen sich die Meßdaten als Kurven darstellen, wobei 4 Kurven die Veränderung gewisser Teilchenanzahlen als Funktion der Zeit darstellen, und es fand zwischen diesen Gruppen durch das Altern der einzelnen Teilchen ein Übergang statt. Wenn man von diesem Sachverhalt hört, so denkt der Mathematiker zunächst an ein Kompartimentmodell oder an einen Wachstumsprozeß und erinnert sich daran, daß dieser Vorgang üblicherweise durch ein System von gekoppelten gewöhnlichen Differentialgleichungen dargestellt wird. Dies war auch mein erster Gedanke.
- 2) Nachdem wir die empirischen Fakten eingehend besprochen hatten, versuchte ich, einen mathematischen Formalismus aufzustellen und die Kurven mit Hilfe mathematischer Methoden zu approximieren. In meiner Begeisterung für diese sehr glatt verlaufenden Kurven ließ ich mich zu der Vermutung hinreißen, daß man aus diesem zeitlichen Kurvenverlauf auch auf das dynamische Verhalten des Entwicklungsprozesses schließen könne. Mit Hilfe von Approximationsprozessen versuchte ich einen Exponentialausgleich der Kurven und hoffte, auf diese Weise die Verweildauer der Teilchen in den einzelnen Gruppen ermitteln zu können.
- 3) Die erhaltenen Ergebnisse wurden dann mit dem Kliniker diskutiert, und es ergaben sich durchaus beträchtliche Abweichungen gegenüber dem, was er erwartete. Gleichzeitig stellte ich fest, daß die gemessenen Daten nicht ausreichten, um den Einschwingvorgang in den stationären Zustand hinreichend deutlich erkennen zu lassen.
- 4) Es wurden deshalb zum einen in der Klinik neue Versuchsreihen durchgeführt und die gewünschten zusätzlichen Werte gemessen. Gleichzeitig versuchte ich, den Mechanismus genauer zu simulieren, der bei den Übergangsprozessen abzulaufen schien. Die zunächst gemachte Annahme eines Kompartimentmodells erwies sich als nicht angemessen.

Nachdem das Modell verbessert war und einen plausiblen Mechanismus für den Ablauf des Reifungsprozesses ergab, ließ sich mathematisch doch wieder, aber nur näherungsweise eine Relation durch Differentialgleichungen ermitteln, wie ich sie ursprünglich vermutet hatte. Mit den neuen Meßdaten konnten dann die Auswertungen wiederholt und aus dem neuen Modell rechnerisch Ergebnisse ermittelt werden, die in guter Näherung auch mit dem übereinstimmten, was man aus anderen medizinischen Ergebnissen erwartete.

Dieser Verlauf ist durchaus typisch für eine solche Zusammenarbeit. Es werden erste Ansätze gemacht und aufgrund von Vergleichen mit der Erfahrung verbessert. Zuweilen wird man diesen Vorgang auch mehrfach zu iterieren haben, bis Ergebnisse herauskommen, die mathematisch konsistent und zum anderen auch dem Mediziner plausibel erscheinen.

Es ist ebenfalls typisch für die geschilderte Entwicklung, daß die Untersuchungen nicht zügig zum Abschluß kamen, denn sowohl der Mediziner wie auch der Leiter eines Rechenzentrums werden zeitlich so stark in Anspruch genommen, daß es Perioden gibt, etwa während des Semesterbetriebes, in denen die Arbeit wochenlang ruht. Es folgen dann wieder Zeiten, in denen die Entwicklung sehr rasch voranschreitet.

- 5) Es bleiben noch die Ergebnisse zu formulieren. Dies geschieht entweder durch eine gemeinsame Arbeit oder, und das ist vielleicht sogar noch befriedigender, durch Entwicklungen eines neuen Hilfsmittels für die Diagnose oder Therapie.

Der Vorteil der Teamarbeit gegenüber der Einzelarbeit kann darin liegen, daß fachliche Ergebnisse erzielt werden, für die dem einen das Rüstzeug gefehlt hätten und dem anderen die Anregungen und die wissenschaftlichen Grundlagen. Es entsteht also mehr als jeder für sich im stillen Kämmerlein oder Labor hätte erzielen können. Die Ergebnisse werden weitreichender und oft auch genauer sein, als ohne das quantitative Hilfsmittel der Datenverarbeitung.

Für wichtig halte ich auch, daß der Mathematiker nicht von seiner Gruppe getrennt wird, sondern teilhat am Pool der Erfahrung, die eine große Gruppe besitzt und die in Gesprächen laufend erweitert wird. Langfristig wird ein Wissenschaftler in einem Team nur dann von Nutzen sein, wenn er beständig versucht, sich in seiner eigenen Disziplin auf dem laufenden zu halten. Für die Mitarbeiter des Rechenzentrums sehe ich als Ergebnis dieser gemeinsamen Arbeiten auch die Befriedigung an, einer interessanten Tätigkeit nachzugehen, deren Wert als Beitrag zum zivilisatorischen Fortschritt man erkennen kann.

Ich verkenne durchaus nicht die Freude, die ein Gelehrter an einem Ergebnis haben kann, das er durch formales Nachdenken erzielt hat. Ich selber kann mich an diesem Spiel, das häufig in der Mathematik getrieben wird, erfreuen.

Es wäre der Entwicklung der Wissenschaft sicherlich abträglich, wenn dieses freie Spiel der Gedanken eingeengt würde. Als Vertreter des Faches Angewandte Mathematik sehe ich jedoch unsere Aufgaben in folgendem:

- a) Es ist zu erproben, ob die heute gelehrtten Methoden auch anwendbar sind, d.h. ob sie sich in der Praxis bewähren und im wahrsten Sinne des Wortes mit Fleisch und Blut versehen werden können.
- b) Die behandelten Probleme sollen aus der Praxis stammen. Die spezifische Aufgabe der Angewandten Mathematik ist es dann, durch die Behandlung einer Vielzahl von Aufgaben zu erkennen, welche Problemstellungen unabhängig von der einzelnen Aufgabe im mathematischen Sinne gleichartig sind, und dafür weittragende Lösungsmethoden zu entwickeln.

Diese Ausführungen waren mehr methodischen Gesichtspunkten gewidmet als der Darstellung von Einzelergebnissen. Es dürfte dem Fachvertreter der Medizin leichter fallen als dem Mathematiker, darzustellen, wie mathematische Methoden und Datenverarbeitung die Medizin in der Forschung zu unterstützen vermögen. Man kann diese Unterstützung wohl in drei große Gruppen einteilen:

- 1) Die Anwendungen statistischer Methoden zur Auswertung medizinischer Messungen. Die Methoden sind im wesent-

lichen bekannt, interessant sind die neuen und oft recht scharfen Ergebnisse.

- 2) Das Bemühen um die quantitative Fassung biologischer und medizinischer Gesetzmäßigkeit, wie es in dem obigen Beispiel anklang.
- 3) Die Anwendung der Ergebnisse in der Diagnose und bei der Behandlung.

Ich darf abschließend zusammenfassen, daß es mir darauf ankam, herauszuarbeiten, daß für das Gelingen einer guten Zusammenarbeit nicht allein fachliche Qualifikationen ausreichen, sondern auch ein gegenseitiges Verständnis der Partner, viel Geduld und das intensive Bemühen, gemeinsam zum Erfolg zu kommen, notwendig sind.

Anschrift des Verfassers:

Prof. Dr. H. Werner
Rechenzentrum
der Universität Münster
44 Münster
Roxeler Straße 60

Prerequisites and programs for on-line data acquisition in clinical laboratories, connected with the medical information – system WAMIS

Von A. Marksteiner

Zusammenfassung

Das Rechenzentrum der Medizinischen Fakultät der Universität Wien unterstützt seit 1968 die Stationen und Stationslabors einiger Kliniken, sowie ein Zentrallaboratorium bei der Durchführung ihrer täglichen Routine Aufgaben. Diese Unterstützung wird vor allem deshalb gegeben, um die Labordaten zum Zwecke späterer wissenschaftlicher Auswertungen vollständig erfassen zu können. Die Verwaltung der im voraus geplanten Laboranforderungen und Therapien stationärer Patienten durch die EDV konnte dem medizinischen Personal der Stationen einen Teil der täglichen Schreibarbeiten abnehmen und darüber hinaus sicherstellen, daß diese Daten computergerecht erfaßt und gespeichert wurden. Ein großzügiger Ausbau des Medizinischen Rechenzentrums ermöglichte die Konzeption des Wiener Allgemeinen Medizinischen Information Systems (WAMIS). Dieses erfaßt sowohl stationäre als auch ambulante Patienten. Das Erfassen der Laborergebnisse dieses nunmehr stark ausgeweiteten Patientenkreises, kann nicht mehr auf konventionelle Weise durch Ablesen der Ergebnislisten erreicht werden.

On-line Anschluß der Laborgeräte an das Informations-System kann Abhilfe bringen, wenn eine Reihe von Voraussetzungen erfüllt werden. Die wichtigste Forderung ist die nach verwechslungsfreier Zuordnung der Daten am Ausgang der Laborgeräte zu den betreffenden Patienten. Dies kann nur erfüllt werden, wenn die Proben eine automatisch lesbare Identifikation tragen. Die kommerziell angebotenen Probenidentifizierungs-Systeme erfüllen nicht alle Forderungen nach leichtem Aufbringen, Lesen und Vervielfachen der Probenidentifikation.

Es wurde ein beliebig oft wiederverwendbarer Proben-Identifikations-Ring entwickelt, der eine 18 Bit Information trägt und die obigen Forderungen erfüllt.

Die entwickelten Programme tragen sowohl den Bedürfnissen der Stationen, das Informations-System als elektronisches Notizbuch für im voraus geplante Befunde und Therapien einzusetzen, Rechnung als auch jenen der Laboratorien, da durch den on-line Anschluß der meisten Laborgeräte ein Großteil der Befunde, ohne Schreibarbeiten des Laborpersonals, erfaßt und an die befundanfordernden Stellen ausgegeben wird. Durch den neu entwickelten Proben-Identifikations-Ring, der für eine Reihe von Probengefäßen paßt, ist die Zuordnung der Meßwerte zu den betreffenden Patienten problemlos.

In 1967 the 2nd Clinic of Internal Medicine, University of Vienna, was equipped with an IBM 360/16K Computer. Statistical evaluation of the patient's medical history data was the main goal at that time. This goal could be reached only by a complete data collection. It proved impossible to enter the data of a patient's medical history after their discharge, since the data are contained there in a great variety of forms. To distill the "real data" from that lot of paper was beyond our capability (1).

A complete data collection must start at the origin of the data.

In a Clinic of Internal Medicine nearly all of the data have their origin at the wards and in the laboratories.

In the first stages of completion an organization was established that helped to reduce clerical work at the wards and provided the laboratories with computerized "working lists" for data collection. Key punching of the "working lists" was the data entry to the computer.

Reduction of clerical work at the wards was accomplished by changing the request-procedure of laboratory tests. In a Clinic of Internal Medicine laboratory tests are requested often and repeatedly. It was the duty of the nurse to fill up all the request-forms for the different laboratories, to prepare all the labels for the test tubes etc., that resulted from a simple word of the physician during the round: "for Mr. X we perform during the next three days all laboratory tests of our program A".

In order to establish a very modest system, which served as a basis from which we gained a lot of experience for the further developments, several prerequisites had to be realized.

1. A requestform, which contains most of the tests that are performed at the clinic, was designed.

On this form all the tests requested for one patient are marked. With an easy mnemonic code the tests can be requested for certain days (1=Monday, 2=Tuesday etc.) daily (T, German "täglich") or every certain day (1F=every Monday, F=German "fortlaufend"). Before the installation of the teleprocessing the requestforms have been sent to the computing center, and after keypunching the data were stored at the request-file.

2. A system of three DA-files was established.

- a) PATI: The first record of this file contained the numbers of all indoor patients, the second record those of indoor patients, discharged within the last 21 days.

- b) NAME

- c) REQUEST

The position of the patient's number in file PATI was the key to the files NAME and REQUEST.

3. The coded testnumbers of the requestform must be completed with additional information that is necessary for further processing.

- a) Information about the laboratory, where the test is performed

- b) Information about the container, necessary for specimen collection

- c) Information about the day(s), when the test can be performed.

4. Every day, after the round, the file REQUEST was searched for tests that were requested for tomorrow. For each of these tests a record was built up and written to a SD-file PERFORM.

5. From the sorted file PERFORM two kinds of lists and selfadhesive labels were printed

- a) The "blood collection lists" and self adhesive labels for the wards.

- b) "Working-lists" for the laboratories.

According to the "blood collection lists" blood was collected by the physicians of the wards and sent to the labs. The results of the tests were recorded on the "working lists" by the technicians. After keypunching the data were entered to the computer to a SD-file RESULT.

6. From the merged files PERFORM and RESULT the reports for the wards were written. At the same time the data were stored on magnetic tapes for statistical evaluation.

From this first stage of completion we learned that it is necessary to differentiate between the ward's laboratories and the central-laboratory. The technicians from the wards' laboratories liked the "working lists" for recording the results because these labs got the specimen only from their ward.

For the central-lab this solution was not a most favourable one. Only 40% of the samples processed there are from indoor patients. For the samples that came from outdoor ward's patients no "working list" could be printed. These specimens arrived at the central laboratory together with the request-forms.

So the "aid" of the computer, given to the central laboratory proved to be doubtful, since an additional differentiation between indoor and outdoor patients had to be introduced.

In order to standardize the sample processing in the central lab we stopped the printing of the "working list" for the central lab. Instead of this, request-forms, similar to those of the outdoor ward's patients, were printed by the computer.

The request-forms were sent, together with the "blood-collection-lists" and the self adhesive labels, to the wards.

By this way of organization no more difference between indoor and outdoor patient's samples resulted.

For both kinds there arrived both samples and request-forms in the central-laboratory. The request-forms served for data recording.

As long as there was no possibility for a real-time data processing this organization was superior because it met the requirements of a medical reality much better. If the "working lists" are prepared in advance according to the requests there is always a difference between the number of requested tests and the samples that arrive at the laboratory. This difference can be easily explained. A working list prepared in advance can not take into account all events that will prevent a blood collection tomorrow (patient has already eaten, he does not feel well etc.). If the request-forms are prepared by the computer the clerical work of the nurse will not increase remarkably but the lab will receive only as many requests as it receives samples, since the nurse controls the blood collection. Request-forms not used are sent back to the computing center. From this again we learned that every program that deals with patients must not be designed like an invoice-program.

After an internal reorganization - the computing center now has to serve all clinics of the Medical Faculty of Vienna - the center was equipped on a large scale in 1972.

The new hardware equipment allowed us to start with the design and programming of a Medical Information System (WAMIS) (2).

Part of WAMIS is a package of programs that achieves a relief in clerical work both for the staff of the wards and

the laboratories. In order to reach that goal we had to take into consideration that our batch-programs could not decrease the workload of the laboratory. On the contrary the chief of the central laboratory suspected that part of the increase of the requests was caused by the comfort of our request-system.

The real bottleneck in laboratory work is shown by an example (3):

A lab that is capable of performing 100 different tests (this figure is a realistic one for the central laboratory of a University Clinic (4)) and that receives specimens of 500 patients daily has to distribute these samples within a matrix that contains 50.000 positions.

If there are done 3 to 6 tests from each specimen, after the distribution 1.500 to 3.000 positions of the 50.000 possible ones will be occupied in a very random manner. During the distribution it always must be assured that the sample identification and the information about the kind of test is transferred in an accurate way. All samples within one row of the matrix are for the same test. The results of the tests have to be connected with the requests. Then the matrix can be cut into columns. Each column now contains the results of one patient.

From this point of view it is obvious that a computer can reduce the clerical workload of a central laboratory only by a new approach to the problems of

SAMPLE IDENTIFICATION
SPECIMEN DISTRIBUTION and
DATA ENTRY to the COMPUTER.

A machine-readable sample identification is a final prerequisite to on-line data acquisition from laboratory instruments. Only if each test-tube bears an identification that is fed automatically to the computer, together with the result of the test, a relief of clerical work is achieved for the staff of a central laboratory.

Specimen distribution is not the main problem in a laboratory equipped with modern multi-channel analyzers. By a

Fig. 1: Test tubes with Sample Identification Ring



skilled organization of specimen collection in accordance with the (modern) laboratory equipment this problem can be minimized.

But in our case we had to draw our attention to the sample distribution, since our programs have to fit for all departments of the Vienna University Hospital. And many of the laboratories are not equipped with multi-channel analyzers but with a series of single-channel instruments and automatic samplers. So sample distribution is an essential part of their workload.

If each sample is attached with a machine-readable identification that bears information about the patient and about the tests to be performed, automated or semi-automated devices for sample distribution can be constructed.

For this reason we had to look for a method of automated sample identification. Our demands for this identification:

- a) machine-readable
- b) easy to apply to different kinds of tubes
- c) easy to reproduce
- d) security
- e) no auxiliary organization for the maintenance of the system.

The commercially available systems at that time (1972) did not satisfy all our demands.

The identification systems used were the following:

PUNCHED CARDS
PUNCHED TUBES or STOPPERS
OPTICAL MARKS ON TUBES
OPTICAL MARKS ON SELFADHESIVE LABELS
(to be fixed around the tube)
OPTICAL MARKS ON SELFADHESIVE LABELS
(to be fixed on special containers)

All systems that work with labels do not allow the reproduction of the information without further manipulation and without an auxiliary label maintenance system.

Systems that work with punched tubes or optical marks on tubes can be reproduced but reading and writing is done by mechanical devices.

By these mechanical devices the speed of reproduction is limited and the method is restricted to special types of tubes.

The development of a reusable "Identification Ring" (5) satisfied all our above mentioned demands for sample identification.

The information is applied to and read by the "Identification-Ring" electro-magnetically without any mechanical device.

The ring can be applied to different kind of tubes, so that nearly all samples that arrive in the laboratory can be identified in an unequivocal manner (Figure 1).

The Identification Ring bears a 18 bit information. A 11 bit laboratory series number, 6 bit method number and a parity bit.

The problem of data entry was solved by an on-line linkage between the laboratory-instruments and the central computer. The main problems of on-line linkage, to install kilometers of cables (6) and the difficulties in communication between technicians and computer (7) were solved by a new equipment, recently developed by an Austrian company. This equipment allows the data transmission to and from 32 laboratory instruments by a single bipolar cable. "Lab-terminals" serve as interfaces between the laboratory instruments and the process control computer. Communication between the technicians and the system is also done by these "lab terminals".

Thus, the hardware problems solved, we started designing a package of programs that is capable of relieving the work of the ward staff as well as of the laboratory personal.

Depending on the circumstances there are four different types of requests for tests that have to be settled by our programs:

The four different types arise from a combination of the following conditions:

1. The request may be planned in advance (A)
or immediate (I)

Requests of type A are mainly for indoor patients.

Requests of type I are mainly for outdoor patients that come to an outpatient's ward without advance notice or in case of indoor patient's "emergency requests".

2. The requested tests are performed either with instruments linked directly to the computer (+) or not (-).

	Type of request	Instrument connected on-line
Type 1	A	+
Type 2	A	-
Type 3	I	+
Type 4	I	-

This four different types of request must be taken into consideration for the "visible" organization as well as for the organization of the databank.

The "visible" organization takes place in two different areas: WARDS and LABORATORIES.

The actions performed there, are:

REQUEST-
SAMPLE- handling
RESULT-

From the point of view of the MIS, an organization has to be established that is capable of acquiring the results of all tests without detriment to the type of request under the conditions to minimize the actions at the wards and laboratories.

REQUEST handling:

At the wards the requestforms are identified with the patient's identification number by means of an addressing plate. The addressing plate was prepared by the computer at the first admission of the patient.

If the tests are to be performed immediately (emergency requests for indoor patients or requests for outdoor patients) the nurse also prepares some self-adhesive labels with that addressing plate. It can be easily seen from the request-form how many tubes have to be sent to the central-laboratory, viz. how many self-adhesive labels have to be prepared.

The data of the requestforms are entered to the Medical Information System by visible display units either at the wards requesting the tests or at a central office. Requests for tests that are planned in advance are stored in a file named REQUEST.

Once a day a search is made in this file for tests that are ordered for tomorrow and can be performed tomorrow.

Tests done from the same specimen and in the same functional unit of the laboratory (e.g. liver tests, kidney tests) are grouped together to form a record in the file LABSER if the specimen collection can be done in the same container. For each record in the file LABSER a self-adhesive label is printed.

From these labels and from "WORKING LISTS", also printed for each ward, the nurses know how many tubes have to be sent to the laboratory.

Requests for tests that have to be performed immediately are not stored in the file REQUEST. For these requests the records of the file LABSER are generated directly from the VDU input. Whether a requested test is stored at first in the file REQUEST from which once a day the records of the file LABSER are generated, or directly in the file LABSER depends on the type of requestform that is used. But there is only one program to enter the data from the request forms.

SAMPLE handling:

At the wards: The samples are collected in containers identified with self-adhesive labels. The patient's identification number and his name is printed on these labels which have been prepared either by the computer or by means of an addressing-plate.

In the central laboratory the patient's identification number of the self-adhesive label at the test tube is read by the technician and entered to the computer by a VDU. This causes all records of the file LABSER of this patient to be displayed at the screen of the VDU. Now the technician points with a lightpen to the correct record, attaches a Sample Identification Ring to the test-tube and puts it into the ID-writer. By this selection the sample number is transferred to the ID-writer placed beside the VDU.

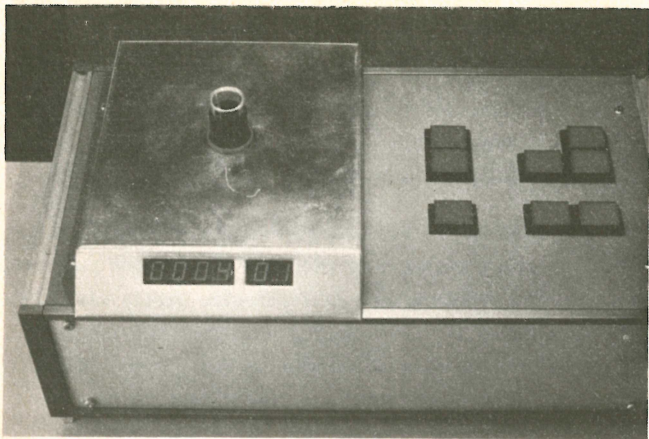
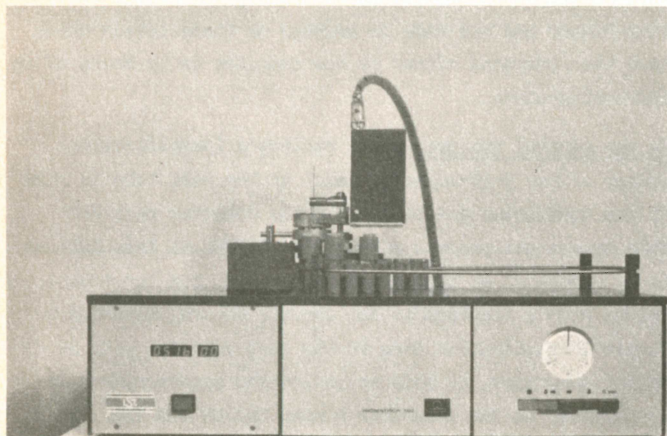


Fig. 2: ID-reader, part of the distribution unit

The sample number consists of a four digit laboratory series number (access to the file LABSER) and a two digit method number. This method number also indicates whether the sample has to be distributed furthermore. Beside the VDU there are also as many racks as there are functional units in this laboratory. The identified sample is placed in its corresponding rack. After the procedure of identification, "working lists" for each functional unit of the laboratory are printed out by a laboratory matrix printer. The "working lists" show the technicians of each functional unit how the samples are to be distributed. Since sample distribution is done semi-automated the "working lists" are optional for most functional units. Only those functional units that also use instruments not connected on-line to the computer do need the "working lists" in any case for recording the results of the off-line tests.

Sample distribution is done by means of a distribution unit. This consists of one ID-reader (Figure 2) and 6 ID-writers. The sample is placed in the ID-reader and the sample-number is transferred to the computer. There the correspondent record in file LABSER is read and the information how this sample has to be distributed, again is transmitted to the ID-writers.

Fig. 3: Commercially available automated sampler, provided with ID-reader



These ID-writers identify the necessary amount of tubes.

For the last stage, the measurement, two types of ID-readers have been developed. One type fits to an automated sampler and can be used for series analysis (Figure 3) another type is for single analysis.

In any case the data at the output of the laboratory instruments and the sample identification are transferred to the computer. By this each data-record can be adjoined in an unequivocal manner to a patient.

Results from tests performed off-line, are recorded by means of "working lists". Again the data entry to the computer is done by VDU. Thus we reached our big goal, firstly to relieve the nurses at the wards from a lot of their clerical work, since they can use the computer as an electronic notebook, secondly to diminish the workload of the central laboratory by on-line data acquisition of the majority of tests, and thirdly to improve and accelerate the data flow within the hospital.

Far away from being exhaustive this brief report only could deal with some of the problems that have arisen.

A detailed description and a report of our experience will be subject of another paper.

References:

1. BRUNNER, H., PAUMGARTNER, G., GRABNER, G. u.a. (1968): "Computer in der Medizin" ed. by K. Fellingner, Verlag Brüder Hollinek, Wien
2. GRABNER, H. and GRABNER, G. (1974): Aims and structure of the Vienna General Medical Information System WAMIS. Paper submitted to MEDINFO 74
3. BÜTTNER, H., EGGSTEIN, M., KELLER, H. u.a. (1972): Dt.Ges.f.Klin.Chemie e.V. Mitteilungen 2/72, 29
4. RICHTERICH, R. (1971): Klinische Chemie. 3. erw.Aufl., Verlag S.Karger, Basel
5. BANCSEICH, J., HADRIAN, W. and MARKSTEINER, A. in press
6. EGGSTEIN, M., GRÄSER, W. and LIEBICH, H.M. (1973): Paper presented at IBM Customer Executive Seminar on Laboratory Information Systems, Paris
7. FLYNN, F.v. (1970): Proceedings of the IFIP-TC4 Working Conference on Information. Processing of Medical Records

Anschrift des Verfassers:

Dr. Alois Marksteiner
 Department of Medical Computer Sciences,
 University of Vienna
 A-1097 Wien
 Garnisongasse 13

Automatische Übersetzung in Blindenkurzschrift

Von W. A. Slaby *)

Zusammenfassung

In diesem Bericht soll das Konzept des MARKOV-Systems formaler Übersetzungsregeln als eine Möglichkeit vorgestellt werden, wie sich Übersetzungsprozesse - in diesem Fall der Prozeß der Übersetzung von Schwarzschrift in Blindenkurzschrift - geeignet formalisieren lassen und damit einer Automatisierung zugänglich gemacht werden können.

Nach der Definition des Begriffs des MARKOV-Systems formaler Übersetzungsregeln und einigen Erläuterungen zu seiner Plausibilität wird an einem konkreten Beispiel aus dem Übersetzungsprozeß in Blindenkurzschrift demonstriert, welches Verfahren wir in Zusammenarbeit mit J. SPLETT vom FB Germanistik der Universität Münster entwickelt haben, um für diesen Übersetzungsprozeß ein vollständiges MARKOV-System formaler Übersetzungsregeln aufzustellen.

Summary

In this report we introduce the concept of a MARKOV-system of formal translation-rules as a means for appropriately formalizing any computable translation, thus especially the translation of an inkprint text of any language into the corresponding grade 2 braille.

After the definition of the notion of MARKOV-system of formal translation-rules and some explanations about its plausibility we demonstrate the method which we have developed together with J. SPLETT of Münster University in order to generate a complete MARKOV-system for the translation process of inkprint into German grade 2 braille.

Bei der Blindenschrift handelt es sich um ein von dem Franzosen Louis Braille (1809-52) entwickeltes Zeichensystem, bei dem jeder Buchstabe, jede Ziffer und jedes Satzzeichen durch ein sog. Braille-Zeichen, d.h. eine Kombination von bis zu 6 erhobenen Punkten dargestellt wird, die als Matrix von drei Zeilen und zwei Spalten angeordnet sind (Beispiel: $\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}$ = A, $\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}$ = Z). Diese Braille-Zeichen kann der Blinde durch Abtasten mit den Fingern lesen.

Ein Text in Blindenvollschrift (engl.: grade 1 braille) wird nun dadurch erzeugt, daß man jedes Zeichen eines Schwarzschrift-Textes durch das entsprechende Braille-Zeichen ersetzt. [Hierbei sind geringfügige Ausnahmen zugelassen, die es gestatten, häufig vorkommende Buchstabenkombinationen wie st, sch, ie u.ä. durch ein einziges Braille-Zeichen zu ersetzen, wenn diese Buchstabenkombinationen etymologisch zu einem Bestandteil eines zusammengesetzten Wortes gehören und nicht die Wortgrenze zwischen den Komponenten eines zusammengesetzten Wortes überschreiten (so darf z.B. die Buchstabenkombination "st" in "Haustür" nicht durch das Braille-Zeichen für "st" übersetzt werden).] Eine solche 1-1-Übersetzung führt wegen der Größe der Braille-Zeichen zu äußerst umfangreichen und platzaufwendigen Braille-Büchern. Daher hat man in den Blindenschriftsystemen der verschiedenen Sprachen jeweils ein Regelsystem entwickelt, das es gestat-

tet, einige Wörter oder Ketten von Buchstaben durch ein einziges oder höchstens zwei Braille-Zeichen zu übersetzen. In diesen Fällen spricht man von ein- bzw. zweiförmigen Wortkürzungen (engl.: word contractions) und Teilwortkürzungen (engl.: part-word contractions). Das Resultat einer Übersetzung in Blindenschrift, die derartige Wort- und Teilwortkürzungen verwendet, nennt man Blindenkurzschrift (engl.: grade 2 braille).

Da der Vorrat von 63 Braille-Zeichen für die Vielzahl von Buchstabengruppen und ganzen Wörtern, die man kontrahieren möchte, nicht ausreicht und daher fast alle Braille-Zeichen eine mehrfache Bedeutung haben (Beispiel: $\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}$ bedeutet "sich", wenn dieses Zeichen allein steht, und es bedeutet "en", wenn es als Teilwortkürzung innerhalb eines Wortes verwendet wird), muß es neben der Aufstellung der Wort- und Teilwortkürzungen und einer Beschreibung ihrer normalen Verwendung ein umfangreiches System von Ausnahmeregeln geben, das die korrekte Anwendung von Kontraktionen sicherstellt. Da es außerdem Wort- und Teilwortkürzungen gibt, die überlappt in einem zu übersetzenden Wort vorkommen können (Beispiel: "bellen", in dem die kürzbaren Buchstabengruppen "be", "el" und "ll" vorkommen), muß unter den verschiedenen Wort- und Teilwortkürzungen eine Rangfolge definiert sein, mit welcher Priorität die einzelnen Kontraktionen anzuwenden sind. Diese Ausnahme- und Prioritätsregeln berücksichtigen in entscheidendem Maße die morphologische Struktur der einzelnen Wörter (wie z.B. Wortbestandteile zusammengesetzter Wörter, Präfix und Suffix

*) Überarbeitetes Manuskript eines Referats anlässlich des Journalistenseminars "COMPUTER IN DER MEDIZIN" am 18./19. Febr. 1974

eines Wortes, präpositionale Verwendung von Buchstabengruppen u.ä.), wodurch bei der Automatisierung dieses Übersetzungsvorgangs linguistische Probleme von enormer Komplexität auftreten.

Bevor wir nun darangehen können, den Prozeß der Übersetzung in Blindenkurzschrift zu formalisieren, müssen zunächst Begriffe wie "Blindenschriftwort" und "Übersetzung" u.ä. eine formale Klärung erfahren.

Definition 1:

Jede endliche, nichtleere Menge Σ werde Alphabet genannt. [Bei unserem Übersetzungsproblem haben wir es in der Hauptsache mit zwei Alphabeten zu tun:

- dem Alphabet der großen lateinischen Buchstaben, arabischen Ziffern, Satzzeichen
 $\Sigma_1 = \{A, B, C, \dots, X, Y, Z, 0, 1, \dots, 8, 9, \dots, !, ?, ;, :, \dots, -, ", \dots\}$
- dem Alphabet der Braille-Zeichen
 $\Sigma_2 = \{ \begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}, \dots, \begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix} \}$,
 welches aus $2^6 - 1 = 63$ Zeichen besteht.]

Jede endliche (evtl. leere) Folge

$$w = x_1 x_2 x_3 \dots x_{n-1} x_n$$

von Elementen $x_1, x_2, \dots, x_{n-1}, x_n$ eines Alphabets Σ heie Wort über Σ .

Das leere Wort, d.h. das Wort, welches aus keinem Zeichen besteht, werde mit ϵ bezeichnet.

Weiter bezeichne

$$\Sigma^* = \{w \mid \exists n \in \mathbb{N}_0 : \exists x_1, \dots, x_n \in \Sigma : w = x_1 \dots x_n\}$$

die Menge aller Wörter über dem Alphabet Σ .

Induktiv definiert man die Länge eines Wortes $w \in \Sigma^*$ $l(w)$ durch

$$l(\epsilon) := 0$$

$$l(ua) := l(u) + 1 \quad \forall u \in \Sigma^*, \quad a \in \Sigma$$

Jede Teilmenge L von Σ^* heie formale Sprache über Σ .

Sind $v = x_1 x_2 \dots x_r$ und $w = z_1 z_2 \dots z_s$ beliebige Wörter über Σ , so definiert man die Verkettung von v und w durch

$$vw := x_1 x_2 \dots x_r z_1 z_2 \dots z_s$$

[Man nennt Σ^* zusammen mit der Operation der Verkettung auch das von Σ erzeugte freie Monoid.]

Seien $x, y \in \Sigma^*$ beliebige Wörter über dem Alphabet Σ .

Man definiert:

$$x \text{ Teilwort von } y : \exists u, v \in \Sigma^* : y = uxv$$

Jedes $x \in \Sigma_2^*$ (Σ_2 Alphabet aller Braille-Zeichen) heie Blindenschriftwort (Braille-Wort).

Unter diesem formalen Aspekt lät sich jede Übersetzung einer Sprache L_1 über einem Alphabet Σ in eine Sprache L_2 über einem Alphabet Δ als eine Vorschrift auffassen, die jedem Wort $w_1 \in L_1$ (mindestens) ein Wort $w_2 \in L_2$ zuordnet.

Definition 2:

Seien Σ und Δ Alphabete, L_1 formale Sprache über Σ , L_2 formale Sprache über Δ sowie $w_1 \in L_1, w_2 \in L_2$ beliebig. Man definiert:

- T Übersetzungszusammenhang zwischen L_1 und L_2
 $: \mathfrak{X} T \subseteq L_1 \times L_2$ [T ist Relation auf $L_1 \times L_2$]
 und $\forall w_1 \in L_1 : \exists w_2 \in L_2 : (w_1, w_2) \in T$
- w_2 Übersetzung von w_1 bzgl. T : $\mathfrak{X} (w_1, w_2) \in T$

Betrachtet man bei geeignet gewählten formalen Sprachen L_1 über Σ_1 (etwa L_1 formale Sprache aller orthographisch korrekten Wörter einer natürlichen Sprache s) und L_2 über Σ_2 den Übersetzungszusammenhang T_s , der jedem Wort $w \in L_1$ die nach dem Blindenkurzschrift-Regelsystem der natürlichen Sprache s korrekte Braille-Übersetzung zuordnet, so erscheint es vor allem auch unter dem Aspekt der Automatisierung des Übersetzungsprozesses plausibel, daß man an einen derartigen Übersetzungszusammenhang T_s folgende weitere Forderungen stellt:

- T_s ist eine Funktion
 d.h. jedem $w_1 \in L_1$ wird durch T_s genau ein $w_2 \in L_2$ zugeordnet.
 Man schreibt dann für das durch $w_1 \in L_1$ eindeutig bestimmte $w_2 \in L_2$ mit $(w_1, w_2) \in T_s$ auch: $T_s(w_1)$
- T_s ist injektiv
 d.h. verschiedenen Wörtern aus L_1 werden durch T_s auch verschiedene Übersetzungen zugeordnet.
- T_s ist rekursive Wortfunktion auf Σ_1^*
 d.h. T_s ist effektiv berechenbar

Einen solchen Übersetzungszusammenhang T_s , der zusätzlich diese drei Forderungen erfüllt, möchte ich im folgenden als Braille-Übersetzungsabbildung der natürlichen Sprache s bezeichnen. Die Übersetzung des Wortes ERDENKEN $\in L_1$ bzgl. T_d , der Braille-Übersetzungsabbildung der deutschen Sprache, lät sich dann schreiben als

$$T_d(\text{ERDENKEN}) = \begin{matrix} \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \end{matrix}$$

Um nun einen Algorithmus zu entwickeln, der eine solche Braille-Übersetzungsabbildung effektiv berechnet, müssen wir die interne Struktur und Definition dieser Abbildung näher untersuchen. Sie ist durch ein kompliziertes, in der "Systematik der deutschen Blindenkurzschrift" [2] festgeleg-

tes System von Regeln gegeben, das die korrekte Anwendung der einzelnen Wort- und Teilwortkontraktionen sicherstellt. Unter anderem finden sich dort Übersetzungsvorschriften wie

ER ist zu übersetzen durch das Braille-Zeichen $\begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$
 EN ist zu übersetzen durch das Braille-Zeichen $\begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$
 EL ist zu übersetzen durch das Braille-Zeichen $\begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$
 LL ist zu übersetzen durch das Braille-Zeichen $\begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$
 H ist zu übersetzen durch das Braille-Zeichen $\begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$
 E ist zu übersetzen durch das Braille-Zeichen $\begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$
 L ist zu übersetzen durch das Braille-Zeichen $\begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$
 N ist zu übersetzen durch das Braille-Zeichen $\begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$

Diese Art von Übersetzungsvorschriften läßt sich leicht in der folgenden Weise formalisieren:

Definition 3: [System formaler Übersetzungsregeln (Semi-THUE-System)]

Ein Paar (Γ, \mathfrak{R}) heie System formaler Übersetzungsregeln (Semi-THUE-System) genau dann, wenn die folgenden Bedingungen erfüllt sind:

1. Γ ist Alphabet
2. $\mathfrak{R} \subseteq \Gamma^* \times \Gamma^*$, endliche, nichtleere Menge

Jedes Paar $(u, v) \in \mathfrak{R}$ heie formale Übersetzungsregel. Für $(u, v) \in \mathfrak{R}$ schreibt man auch $u \rightarrow v$.

Um die Wirkung eines Systems formaler Übersetzungsregeln (Γ, \mathfrak{R}) auf ein Wort w über Γ zu erklären, muß definiert werden, was es heißen soll, daß eine formale Übersetzungsregel $u \rightarrow v$ auf w anwendbar ist.

Sei (Γ, \mathfrak{R}) ein System formaler Übersetzungsregeln, $w \in \Gamma^*$ ein Wort über Γ und $u \rightarrow v \in \mathfrak{R}$ eine formale Übersetzungsregel.

Man definiert:

$u \rightarrow v$ anwendbar auf w
 : $\exists u$ Teilwort von w
 [$\exists x, y \in \Gamma^* : w = xuy$]
 Def.1

Ist $u \rightarrow v \in \mathfrak{R}$ anwendbar auf w , so nennt man $z := xvy$ das Resultat der Anwendung von $u \rightarrow v$ auf w .

Betrachten wir dazu ein Beispiel.

Wenn wir die oben erwähnten Übersetzungsvorschriften für die Übersetzung in Blindenkurzschrift in dieser Weise formalisie-

ren, so erhalten wir folgendes spezielle System formaler Übersetzungsregeln (Γ, \mathfrak{R})

mit $\Gamma = \Sigma_1 \cup \Sigma_2$, Σ_1, Σ_2 wie in Definition 1
 und $\mathfrak{R} = \{ ER \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}, EN \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}, EL \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}, LL \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}, \\ N \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}, H \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}, E \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}, L \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix} \}$

Betrachtet man nun z.B. das Wort $w = \text{LERNEN}$, so sieht man leicht, daß die formale Übersetzungsregel $ER \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ auf w anwendbar ist mit dem Resultat $z_1 = L \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix} \text{NEN}$. Als nächstes ist etwa die formale Übersetzungsregel $EN \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ auf z_1 anwendbar, wobei sich als Resultat der Anwendung das Wort $z_2 = L \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix} N \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ ergibt. Zuletzt wenden wir die formalen Übersetzungsregeln $L \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ und $N \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ an und erhalten als Endresultat die korrekte Blindenkurzschrift-Übersetzung des Wortes LERNEN, nämlich $\begin{smallmatrix} \bullet\bullet & \bullet\bullet & \bullet\bullet & \bullet\bullet \\ \bullet\bullet & \bullet\bullet & \bullet\bullet & \bullet\bullet \\ \bullet\bullet & \bullet\bullet & \bullet\bullet & \bullet\bullet \end{smallmatrix}$.

An dieser Stelle sollte man bemerken, daß in der Auswahl der anzuwendenden formalen Übersetzungsregel eine gewisse Willkür liegt, wenn mehr als eine formale Übersetzungsregel auf ein Wort w anwendbar ist. Bei dem obigen Beispiel etwa sind zu Beginn die formalen Übersetzungsregeln $ER \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$, $EN \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$, $L \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$, $E \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ und $N \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ auf das Wort $w = \text{LERNEN}$ anwendbar.

Wenn wir nun zuerst die Regel $E \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ anwenden, um etwa das erste "E" in "LERNEN" zu übersetzen, so erhalten wir sicher nicht die korrekte Blindenkurzschrift-Übersetzung von $w = \text{LERNEN}$.

Ähnliche Komplikationen treten dadurch auf, daß es kontrahierbare Buchstabengruppen gibt, die überlappt in einem Wort auftreten können. Betrachten wir etwa das Wort $w = \text{HELL}$ und versuchen, darauf das obige System formaler Übersetzungsregeln anzuwenden. Dabei macht man zunächst sicherlich nichts falsch, wenn man als erste die Regel $H \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ anwendet. Aber im nächsten Schritt stehen wir vor dem Problem, ob wir zuerst die Regel $EL \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ und dann die Regel $L \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ oder zuerst die Regel $E \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ und dann die Regel $LL \rightarrow \begin{smallmatrix} \bullet\bullet \\ \bullet\bullet \\ \bullet\bullet \end{smallmatrix}$ anwenden müssen, was offensichtlich zu unterschiedlichen Resultaten führt.

Da die "Systematik der deutschen Blindenkurzschrift" für diesen Fall eine eindeutige Übersetzung in Blindenkurzschrift vorsehen muß, ist dort festgelegt, daß die Teilwortkontraktion LL mit höherer Priorität anzuwenden ist als die Teilwortkontraktion EL.

Wenn wir also die interne Definition einer Braille-Übersetzungsabbildung adäquat formalisieren wollen, müssen wir einen Formalismus entwickeln, der uns in die Lage versetzt, diese Metaregel der Priorität mit zu berücksichtigen. Eine Möglichkeit, dieses Ziel zu erreichen, stellt das Konzept des MARKOV-Systems formaler Übersetzungsregeln dar.

Definition 4: [MARKOV-System formaler Übersetzungsregeln]

Ein Quadrupel $\mathfrak{M} = (\Sigma, \Delta, \Gamma, \mathfrak{R})$ heie MARKOV-System formaler Übersetzungsregeln, wenn die folgenden Bedingungen erfüllt sind:

1. Σ, Δ, Γ sind Alphabete mit $\Sigma \subseteq \Gamma$ und $\Delta \subseteq \Gamma$
 $[\Sigma$ heie Inputalphabet, Δ Outputalphabet und Γ Arbeitsalphabet]
2. (Γ, \mathfrak{R}) ist ein System formaler Übersetzungsregeln und \mathfrak{R} ist vollständig und linear geordnete Menge.

Sei $w \in \Gamma^*$ ein beliebiges Wort über Γ .

Man definiert:

\mathfrak{M} anwendbar auf w : $\exists u + v \in \mathfrak{R} : u + v$ anwendbar auf w

Sei \mathfrak{M} anwendbar auf w und sei $u_0 + v_0$ die erste formale Übersetzungsregel - erste bzgl. der Ordnung auf \mathfrak{R} -, die anwendbar auf w ist. Da u_0 also Teilwort von w ist, existieren $x, y \in \Gamma^*$ derart, daß

$w = xu_0y$ und $l(x)$ minimal mit dieser Bedingung.

Man definiert dann

$$\mathfrak{M}(w) := xv_0y$$

und erweitert diese Definition auf $\mathfrak{M}^n(w)$ durch

$$\mathfrak{M}^0(w) := w$$

$$\mathfrak{M}^n(w) := \mathfrak{M}(\mathfrak{M}^{n-1}(w)) \quad \text{für } n \in \mathbb{N},$$

vorausgesetzt, daß $\forall i (0 < i < n-1) :$
 \mathfrak{M} anwendbar auf $\mathfrak{M}^i(w)$

Man sieht nun leicht, daß für jedes $w \in \Sigma^*$ genau einer der beiden folgenden Fälle eintritt:

1. Fall: es gibt ein $r_w \in \mathbb{N}_0$, so daß
 $\forall i (0 < i < r_w - 1) : \mathfrak{M}$ anwendbar auf $\mathfrak{M}^i(w)$
 und \mathfrak{M} nicht anwendbar auf $\mathfrak{M}^{r_w}(w)$

2. Fall: $\forall n \in \mathbb{N}_0 : \mathfrak{M}$ anwendbar auf $\mathfrak{M}^n(w)$

Ist im 1. Fall $\mathfrak{M}^{r_w}(w) \in \Delta^*$, so heißt $\mathfrak{M}^{r_w}(w)$ das Resultat der Anwendung von \mathfrak{M} auf w .

Da wir hier hauptsächlich an der Formalisierung einer Braille-Übersetzungsabbildung T_S interessiert sind, die einen Übersetzungszusammenhang zwischen einer natürlichen Sprache (s) und der korrespondierenden Blindenkurzschrift darstellt, hat das MARKOV-System formaler Übersetzungsregeln im allgemeinen die Form:

$$\mathfrak{M}_S = (\Sigma_1, \Sigma_2, \Gamma, \mathfrak{R}_S).$$

Da darüberhinaus wohl jede vernünftig definierte Braille-Übersetzungsabbildung T_S eine rekursive Wortfunktion von Σ_1^* in Σ_2^* und damit effektiv berechenbar ist, lät sich \mathfrak{R}_S immer so wählen, daß für alle $w \in \Sigma_1^*$ der 1. Fall von Definition 4 erfüllt ist.

Betrachten wir als ein Beispiel das folgende spezielle MARKOV-System formaler Übersetzungsregeln

$$\mathfrak{M} = (\Sigma_1, \Sigma_2, \Gamma, \mathfrak{R})$$

mit $\Gamma = \Sigma_1 \cup \Sigma_2$

und $\mathfrak{R} = \{1. LL \rightarrow \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}, 2. EL \rightarrow \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}, 3. ER \rightarrow \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix},$
 $4. EN \rightarrow \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}, 5. E \rightarrow \begin{smallmatrix} \bullet \\ \bullet \end{smallmatrix}, 6. H \rightarrow \begin{smallmatrix} \bullet \\ \bullet \end{smallmatrix},$
 $7. L \rightarrow \begin{smallmatrix} \bullet \\ \bullet \end{smallmatrix}, 8. N \rightarrow \begin{smallmatrix} \bullet \\ \bullet \end{smallmatrix} \}$

zusammen mit dem Wort $w = ERHELLEN$.

\mathfrak{M} ist anwendbar auf w und wir erhalten die Ableitungskette

$$\mathfrak{M}(w) = ERHE \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} EN = : z_1$$

$$\mathfrak{M}(z_1) = \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} HE \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} EN = : z_2$$

$$\mathfrak{M}(z_2) = \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} HE \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} = : z_3$$

$$\mathfrak{M}(z_3) = \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} H \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix} = : z_4$$

und schließlich

$$\mathfrak{M}(z_4) = \begin{smallmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$$

welches die korrekte Blindenkurzschrift-Übersetzung des Wortes ERHELLEN ist. Außerdem sieht man sofort, daß \mathfrak{M} nicht mehr anwendbar ist.

Setzt man zunächst einmal voraus, man hätte ein MARKOV-System formaler Übersetzungsregeln \mathfrak{M}_S aufgestellt, welches vollständig eine Braille-Übersetzungsabbildung T_S beschreibt. Dann gilt offensichtlich für jedes $w \in \Sigma_1^*$:

$$T_S(w) = \mathfrak{M}_S^{r_w}(w)$$

Also ist der Algorithmus, der die Braille-Übersetzungsabbildung T_S berechnet, nichts anderes als ein effektives Verfahren, welches die Anwendung eines speziellen, als Parameter übergebenen MARKOV-Systems \mathfrak{M}_S auf irgendein eingegebenes Wort $w \in \Sigma_1^*$ durchführt. Dieser Algorithmus, deswegen universeller MARKOV-Algorithmus genannt, werde hier durch ein PL/I-ähnliches Computerprogramm beschrieben.

Dazu sei WORD die Variable, auf der zu Beginn das zu übersetzende Wort $w \in \Sigma_1^*$ gespeichert wird. LSIDE(N) und RSIDE(N) seien zwei eindimensionale arrays der Länge N, die die linken bzw. rechten Seiten der formalen Übersetzungsregeln des MARKOV-Systems aufnehmen. [Dabei wird N als Parameter an das Programm übergeben und danach erst Speicherplatz für LSIDE(N) und RSIDE(N) reserviert.]

Der wesentliche Teil des Algorithmus ist dann durch die folgenden Instruktionen gegeben:

```
LOOP:
  do I = 1 to N;
  M = index (WORD, LSIDE(I));
  if M > 0
  then do;
    ersetze das Vorkommen von LSIDE(I) in WORD ab
    Position M durch RSIDE(I);
    go to LOOP;
  end;
end;
```

Um für den speziellen Übersetzungszusammenhang der Braille-Übersetzungsabbildung T_d (für die deutsche Blindenkurzschrift) ein vollständiges MARKOV-System formaler Übersetzungsregeln M_d aufzustellen, müssen die linguistischen Probleme, die vor allem bei der korrekten Übertragung der Ausnahme- und Prioritätsregeln der "Systematik der deutschen Blindenkurzschrift" auftreten, auf die Ebene formaler Zeichenketten transponiert werden.

Dazu wird in die formalen Übersetzungsregeln des MARKOV-Systems soviel an Informationen darüber aufgenommen, welche Buchstabenkombinationen als Umgebung der einzelnen Wort- und Teilwort-Kontraktionen in deutschen Wörtern auftreten können, wie zur Sicherstellung einer korrekten Übersetzung erforderlich ist. Welche Methode wir in Zusammenarbeit mit Dr. J. SPLETT [4] vom Germanistischen Seminar der Universität Münster entwickelt haben, um diese Informationen über die Wortkonstituenz aus den Wörtern der deutschen Sprache herauszukristallisieren, möchte ich am Beispiel der Teilwortkürzung "sam" $\hat{=} \begin{matrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{matrix} \hat{=} "s"$ demonstrieren.

Nach den Regeln der deutschen Blindenkurzschrift darf die Teilwortkürzung "sam" $\hat{=} \begin{matrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{matrix}$ nur dann angewandt werden, wenn es sich bei der zu übersetzenden Buchstabengruppe um die Nachsilbe "sam" wie z.B. in "grausam", "einsam" und "verlangsam" handelt. Dabei ist die Einschränkung zu beachten, daß die Anwendung nach einfachen Vokalen (wie in "lobesam") nicht gestattet ist.

Um nun eine möglichst vollständige Liste aller deutschen Wörter zu erstellen, in denen das Morphem "sam" als Wortbestandteil vorkommt, sind wir nach der folgenden Methode vorgegangen:

1. Automatische Erstellung einer Liste aller Wörter des Deutschen Wörterbuchs von MACKENSEN, die die Buchstabenkombination "sam" als Wortbestandteil enthalten. Hierzu wurde ein Computerprogramm entwickelt, welches diesen Test auf vorgegebene Wortbestandteile für jedes Wort des auf Magnetband gespeicherten Wörterbuchs durchführt und die Liste in einer geeigneten Sortierfolge ausdrückt.

2. Manuelle Ergänzung der Liste, dadurch daß alle Wörter anderer Wörterbücher in die Liste aufgenommen werden, die mit "sam" beginnen oder auf "sam" enden und noch nicht in der Liste verzeichnet sind.

Da aber die größten Probleme in der korrekten Anwendung von Kontraktionen bei zusammengesetzten Wörtern auftreten, wird die Liste nach folgendem manuellen Verfahren weiter ergänzt:

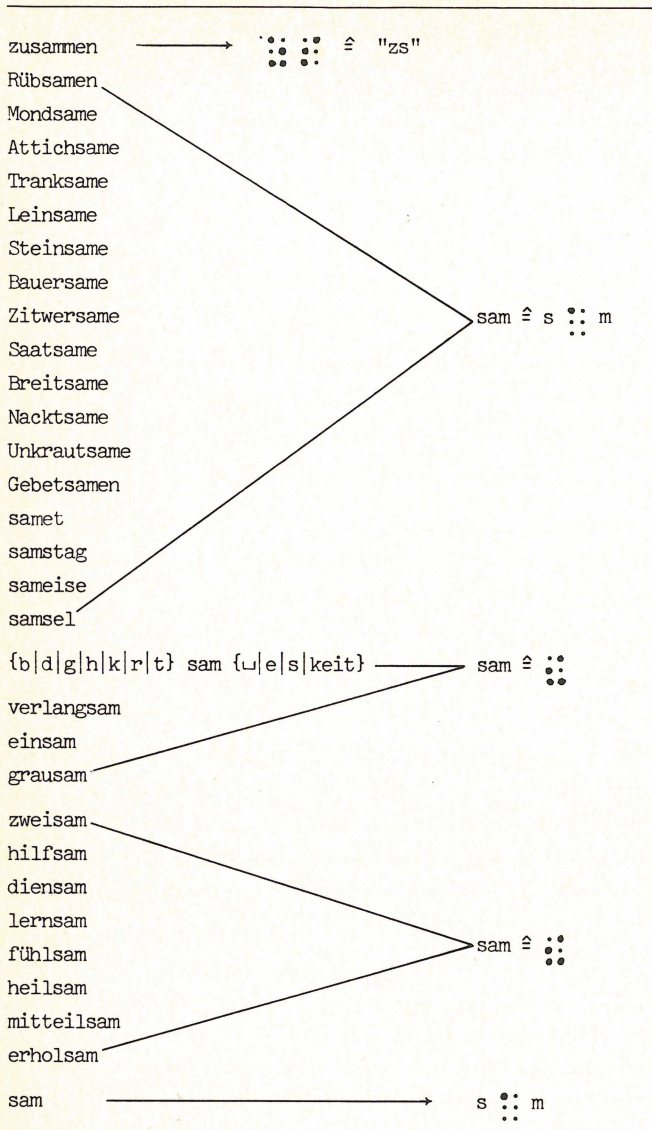
- Produktion einer Liste A1 aller Wörter, die auf "s" enden bzw. die "es" oder "s" als Fugenmorphem bei Wortzusammensetzungen verwenden (z.B. Stand(es), Arbeit(s))
- Produktion einer Liste B1 aller Wörter, die mit "am" beginnen (z.B. Amt, Ampulle, Amulett, Ampel)
- Produktion einer Liste A2 aller Wörter, die auf "sa" enden (z.B. Mensa, Prosa, Hansa)
- Produktion einer Liste B2 aller Wörter, die mit "m" beginnen.

Durch systematisches Testen aller Möglichkeiten, welche Wörter aus Liste A1 bzw. A2 sich mit welchen Wörtern aus Liste B1 bzw. B2 zu sinnvollen zusammengesetzten Wörtern kombinieren lassen, können weitere deutsche Wörter mit Wortbestandteil "sam" (wie z.B. Halsamulett, Verkehrsampel, Standesamt, Mensamahl u.ä.) aufgespürt und als Ergänzung in die Hauptliste aufgenommen werden.

Nach dieser Erstellung einer möglichst vollständigen Liste aller Wörter, die als Wortbestandteil das Morphem "sam" enthalten, kann man diese Wörter daraufhin analysieren, welche Buchstaben bzw. Buchstabenkombinationen als Umgebung des Morphems "sam" auftreten können. Dabei stellt sich heraus, daß - abgesehen von wenigen Ausnahmen - das Morphem "sam" als Nachsilbe höchstens nach den Buchstaben b, d, g, h, k, r, t auftritt. Außerdem kann "sam" als Nachsilbe höchstens die Endungen "e", "es", "er", "en", "em", "st..." sowie bei Substantivierung die Endung "keit" annehmen.

Hinzu kommen noch Verbformen von "einsam" und "langsam", bei denen die Endung "t" (bei "verlangsam" und "vereinsamt") einige Schwierigkeiten bereitet, sowie Substantive wie "Gehorsam" und "Gewahrsam", bei denen als Endung zusätzlich ein "s" auftreten kann. Betrachtet man die Endungen, die "sam" als Nachsilbe annehmen kann, so stellt man leicht fest, daß die Endungen "e" und "en" dadurch Probleme verursachen, daß es eine Reihe von zusammengesetzten Wörtern mit dem Wortbestandteil "Same" oder "Samen" (wie in "Leinsamen" oder "Hanfsame") gibt.

Wie diese Probleme bei der Aufstellung eines MARKOV-Teilsystems formaler Übersetzungsregeln für das Morphem "sam" gelöst werden, zeigt die folgende Aufstellung, bei der die formalen Übersetzungsregeln nur angedeutet werden:



Die Tatsache, daß nach dieser Analysierungsmethode insgesamt mehr als 400 Wort- und Teilwort-Kontraktionen zu untersuchen sind, die zum Teil (wie bei "st") weit komplizierterer Natur sind, läßt uns vermuten, daß diese Untersuchungen trotz der Beteiligung einiger Mitarbeiter des Rechenzentrums und einer Gruppe von Studenten der Germanistik erst im Laufe des Jahres 1975 abgeschlossen werden können.

Unter diesem Aspekt des enormen Arbeitsaufwands stelle ich zur Zeit Untersuchungen darüber an, in welcher Richtung das allgemeine Konzept des MARKOV-Systems spezialisiert werden muß, um eine Automatisierung des Verfahrens der Generierung der formalen Übersetzungsregeln des MARKOV-Systems zu ermöglichen. Über die Ergebnisse dieser Untersuchung und über das Programmsystem, welches interaktiv die Generierung der formalen Übersetzungsregeln durchführt, wird berichtet werden.

Literaturverzeichnis

- [1] DOST, W. (1965): Übersetzung natürlicher Sprachen und Blindenschriftübersetzung. Diplomarbeit, Hamburg
- [2] FREUND/MITTELSTEN, SCHEID/BRITZ (1972): Systematik der deutschen Blindenkurzschrift (Entwurf). Deutsche Blindenstudienanstalt, Marburg
- [3] SLABY, W.A. (1973): A universal braille translator. In: Proceedings of the Intern. Conference on Computational Linguistics, Pisa
- [4] SPLETT, J. (--): Linguistische Probleme bei der automatischen Produktion der deutschen Blindenkurzschrift. Voraussichtl. in: Z.f.Dialektologie und Linguistik
- [5] WERNER, H. (1974): The historical development of automatic braille production in Germany. In: SIGCAPH News, Ass. for Computing Machinery, N.Y.

Anschrift des Verfassers:

Dipl.-Math. Wolfgang A. Slaby
Rechenzentrum der
Universität Münster
44 Münster
Roxeler Straße 60

Aufbau und Organisation eines integrierten Real-Time Datenverarbeitungssystems für die Strahlentherapie und Nuclearmedizin

Von H. Gremmel, U. Günther, J. Pomplun, K. Stamer und H. Wendhausen

Zusammenfassung

Es wird ein Prozessdatenverarbeitungssystem für die Nuclearmedizin und Strahlentherapie einer Radiologischen Klinik beschrieben. Neben den generellen Anforderungen an ein System dieser Art wird die hardware- und softwaremäßige Realisierung geschildert.

Summary

A real time data processing system for a radiological clinic concerning nuclear medicine and high voltage therapy is described. Besides the general demands for a system of this kind the configuration of hardware and software is explained.

Einleitung

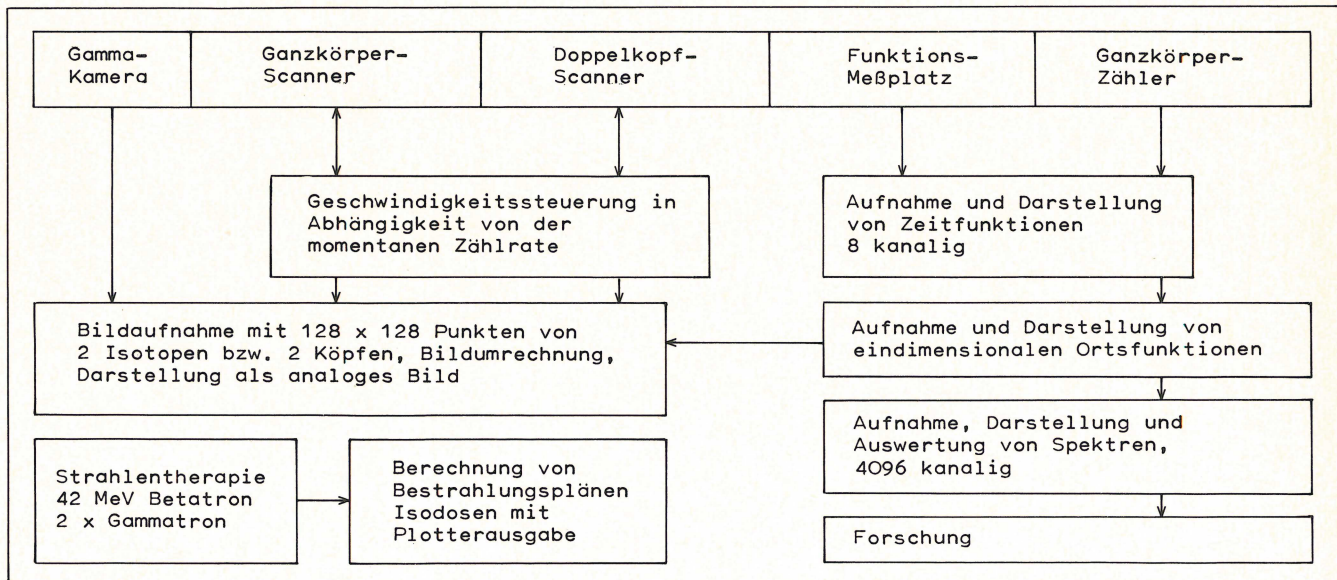
Im Neubau der Radiologischen Universitätsklinik in Kiel war für die folgenden Aufgabenbereiche eine Datenverarbeitung geplant (siehe Abbildung 1):

1. Die Signale einer Gammakamera mit Bildern aus 128 x 128 Punkten von maximal zwei Isotopen sollen aufgenommen und analog dargestellt werden. Die Möglichkeit einer Bildumrechnung wie Glättung, Refokussierung und Kontrastanpassung soll gegeben sein.
2. Ähnliche Anforderungen werden an die Verarbeitung der Meßwerte eines Ganzkörper-Doppelkopfscanners mit zwei 8-Zoll Detektoren gestellt. Zusätzlich ist eine Geschwindigkeitssteuerung des Scanners durch den Rechner verlangt, um Bereiche ohne nennenswerte Aktivitätsbe-

legung mit Höchstgeschwindigkeit zu überfahren. Interessierende Bezirke sollen in einigen Fällen mit gleicher Statistik abgetastet werden, d.h. der Scanner verweilt hier so lange über jedem Punkt, bis eine vorgewählte Impulszahl erreicht ist.

3. Die Anforderungen beim Ganzkörperscanner und einem konventionellen Doppelkopfscanner sind völlig identisch.
4. Ein Funktionsmeßplatz mit vier Detektoren und acht Kanälen liefert Zeitfunktionen der Aktivität über vier Körperpunkten, die gespeichert und mathematisch ausgewertet werden sollen.
5. Eine Hauptaufgabe des Ganzkörperzählers ist die Aufnahme und Auswertung von Spektren inkorporierter gammastrahlender Nuclide. Zusätzlich sollen ein- und zweidimensionale Ortsfunktionen der Aktivität aufgenommen werden. Weiterhin wird im Bereich des Ganzkörperzählers Rechenkapazität für Forschungsaufgaben benötigt.

Abb. 1: Aufgaben der Datenverarbeitung



6. Die Strahlentherapie im Neubau der Radiologischen Universitätsklinik in Kiel arbeitet mit einem 42 MeV Betatron und zwei Co-60 Bestrahlungsgeräten. Die Dosisverteilung im Patienten bei Stehfeld-, Pendel- und Kreuzfeuerbestrahlung soll mit Hilfe eines Rechners geplant und die Linien gleicher Dosis auf einem Plotter ausgezeichnet werden.

Datenanfall

Beide Scanner und der Ganzkörperzähler liefern täglich etwa zehn bis zwanzig Bilder mit jeweils 16 K Daten während 30 Minuten. Die Meßwerte müssen on-line übernommen und abgelegt werden. Der Abstand der Rohdaten liegt im Millisekundenbereich, die Matrixgröße sowie Zeilen- und Spaltenzahl sind variabel.

Die aktive Geschwindigkeitssteuerung der Scanner soll on-line als Funktion jedes einzelnen übernommenen Meßwertes geschehen. Beim Funktionsmeßplatz erscheinen bis zu zehn mal täglich maximal 9 x 1000 Meßwerte, deren Mindestabstand für eine Gruppe aus 9 Werten 0,1 sec beträgt.

Die Gammakamera kann täglich bis zu 40 Bilder mit jeweils 2 x 300.000 Einzeldaten und einem Mindestabstand von 6 µsec liefern. Die Abspeicherung der 600.000 Werte soll in zwei 128 x 128 Punkte Matrizen geschehen.

Die Berechnungen der Strahlentherapie benötigen mindestens 12 K 16 bit Worte im Kernspeicher bei segmentierten Programmen. Es fallen täglich etwa 10 bis 15 Berechnungen mit einer mittleren Rechenzeit von 5 Minuten an.

Datenauswertung und Archivierung

Die Meßwerte aller Geräte werden in Matrixform geordnet, wobei die Zeilen und Spalten unterschiedlicher Größe den Koordinaten des nuclearmedizinischen Bildes bzw. denen der Funktionsuntersuchung entsprechen müssen.

Zur Bildumrechnung wird ein Befehlsvorrat für Matrixoperationen benötigt, der jederzeit durch Anwenderprogramme erweiterbar sein muß.

Die Bildmatrizen sollen kurzzeitig (etwa 24 Stunden) auf einem Massenspeicher abgelegt werden. In diesem Zeitraum ist die Bildaufbereitung und endgültige Archivierung als Foto in der Krankenakte vorgesehen. Für wissenschaftliche Auswertung wird ein weiterer Massenspeicher mit einer Kapazität von einigen hundert Bildern zur Langzeitspeicherung notwendig. Zu sämtlichen Bildmatrizen gehören patienten- und untersuchungsbezogene Kommentare und Parameter. Das System soll die Datensätze unter Identifikationsnummern verwalten.

Datenausgabe

Die Ergebnisse der nuclearmedizinischen Untersuchungen sind dreidimensionale Verteilungsmuster der Radioaktivität, wo-

bei neben den beiden Ortskoordinaten die Intensität die dritte Koordinate ausmacht. Die Abbildung von x-y Koordinaten kann auf jedem Bildschirm geschehen, zusätzlich soll die z-Achse als Helligkeitsmodulation der Punkte bzw. als Farbmodulation auf den Sichtschirmen erscheinen, wobei ein "stehendes" Bild verlangt wird. Die Isodosen der Strahlentherapiepläne sind zweidimensionale graphische Darstellungen, zur Ausgabe eignet sich ein x-y Plotter.

Alphanumerische patientenbezogene Daten und Kommentare sollen über einen Schnelldrucker ausgegeben werden.

Radiologische Gesichtspunkte

Durch verbesserte Auswertung und Darstellung von Untersuchungsergebnissen und Therapieplänen soll die Qualität der radiologischen Diagnostik und Therapie gesteigert werden. Voraussetzung dafür ist ein Datenverarbeitungssystem, das von der Datenaufnahme über Auswertung und Ausgabe von eingearbeitetem medizinischem Personal zu handhaben und in seiner Wirkungsweise nachvollziehbar ist. Die Radiologie, insbesondere die Nuclearmedizin, ist einer der jüngsten Zweige der Medizin. Weiterentwicklungen auf technischen und medizinisch-wissenschaftlichen Gebieten sind in schneller Folge zu erwarten. Die Datenverarbeitung muß hardware- und softwaremäßig so modular aufgebaut sein, daß sämtliche Neuerungen des Gebietes mit einfachen Erweiterungen und Änderungen des Gesamtsystems Berücksichtigung finden können.

Hardwarekonfiguration

Der von allen Geräten einer nuclearmedizinischen Abteilung stammende Datenfluß ist so groß, daß sich der Einsatz eines einzigen Rechners zu dessen Erfassung und Verarbeitung nicht empfiehlt. Seit Kompaktrechner zu einem realistischen Preis angeboten werden, kann die Datenerfassung und Steuerung der verschiedenen Geräte über mehrere Rechner-Systeme erfolgen. Diese Systeme sollten weitgehend autonom arbeiten, jedoch alle zur weiteren Verarbeitung und Archivierung mit einem Zentralsystem verbunden werden. Aus diesen Gründen wurde die in Abbildung 2 aufgeführte Hardwarekonfiguration zur Realisierung der gestellten Aufgaben verwendet.

Die Verbindung der Rechner untereinander muß gewährleisten, daß jedes System die umfangreiche Ausstattung - Hardware und Software - des Zentralrechners benutzen kann.

Durch die Archivierung der aufgenommenen Patientendaten an einer Stelle hat der Benutzer die Möglichkeit, Vergleiche der mit verschiedenen Geräten aufgenommenen Daten anzustellen.

Programme für alle Peripherierechner können auf dem dafür besser ausgestatteten Zentralsystem zeit- und kostensparend erstellt und abgespeichert werden.

Das zentrale Rechnersystem übernimmt Daten von den fünf Stationen, in ihm wird ausgewertet und archiviert. Gleich-

zeitig ist Programmentwicklung und Therapieplanung möglich, ohne den Datenfluß von den externen Rechnern zu unterbrechen (Multiprogramming-Betriebssystem).

Das Gamma-Kamera-System arbeitet weitgehend autonom, auch bezüglich der Verarbeitung der Meßwerte. Bei Bedarf kann jedoch eine Datenübertragung zur weiteren Auswertung im Zentralsystem vom Bediener eingeleitet werden.

Der Rechner für den Doppelkopfscanner dient zur Steuerung des Scanners und Akkumulation der Zählraten von den beiden Detektoren. Die Meßergebnisse werden blockweise zum Zentralsystem übertragen.

Der Rechner für den Ganzkörperscanner hat im wesentlichen die gleiche Aufgabe wie der für den Doppelkopfscanner.

Das System am Ganzkörperzähler ist wie das Gamma-Kamera-System unabhängig vom Zentralrechner. Es arbeitet als Vielkanalanalysator und übernimmt zusätzliche Zählraten vom Ganzkörperzähler. Es hat sich jedoch in der Praxis gezeigt, daß dieses System autonom arbeitet und keine der Fähigkeiten des Zentralsystems benötigt. Daher wird die vorhandene Verbindung zum Zentralsystem nicht genutzt.

Der Funktionsmeßplatz ist über ein digitales Koppelgerät direkt mit dem Zentralrechner verbunden, der während einer Messung ständig die Zählraten vom Meßplatz übernimmt.

Systembeschreibung

Allgemeines:

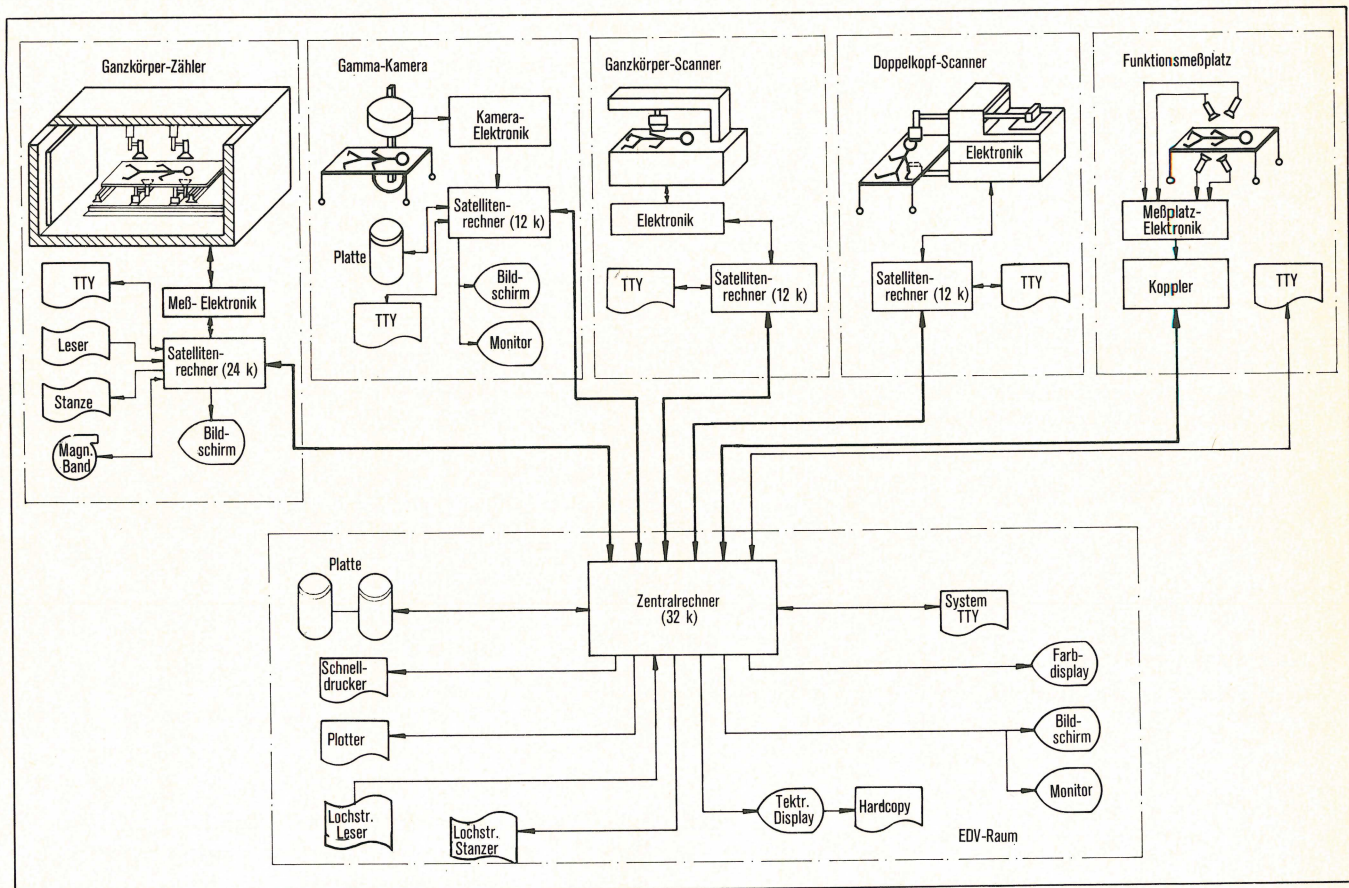
An jeder der peripheren Datenstationen werden vor Beginn einer Messung oder Meßwertübertragung Patientendaten und Meßparameter im Dialog über die Bedienungskonsole erfragt. Eingegeben werden Patientennummer, Name, Geburtsdatum, appliziertes Isotop, Dosis, Lage(n), Kollimatornummer(n), Anzahl der Analysatoren und zusätzlicher Kommentar.

Aus diesen Angaben wird auf einer Platte des Zentralsystems die Patientendatei aufgebaut und über den Schlüsselwert Patientennummer ein Zeiger zu den danach erzeugten Dateien mit den Szintigrammen und Zeitfunktionen des Patienten geschaffen. Weitere Fragen im Dialog beziehen sich auf den bevorstehenden Meßvorgang: Fahrgeschwindigkeit der Meßköpfe, Weginkrement zur Akkumulation der Zählimpulse und Bildgröße bei den Scannern; Gesamtmeßzeit und Zeitinkrement beim Funktionsmeßplatz.

Die Datenübertragung von den Satellitensystemen zum Zentralrechner und zurück erfolgt in allen Fällen über gleiche Leitungen und Schnittstellen bitseriell. Für je 16 bit werden zusätzlich drei Prüfbits generiert und überwacht. Es werden paarig abgeschirmte, verdrehte Leitungen verwendet, welche zur Potentialtrennung an den Schnittstellen auf beiden Seiten optisch entkoppelt sind.

Die erreichbare Übertragungsgeschwindigkeit bei den in der Radiologie Kiel vorhandenen Entfernungen von ca. 200 m be-

Abb. 2: Konfiguration des Gesamtsystems



trägt 1 Million bit/sec. Der mögliche Datendurchsatz von den Peripherierechnern zur Platte des Zentralsystems wird wesentlich von anderen Faktoren bestimmt: Zugriffszeit der Platte, Länge der Datenblöcke, Software-Prioritäten auf beiden Seiten, direkter Kernspeicherzugriff.

Gamma-Kamera-System

Die von der Kamera gelieferten Informationen über den Ort und die Zeit werden in einem schnellen Doppel-ADC mit je maximal 8 bit und einer Wandlerrate von 200 MHz (Konversionszeit: $2,8\mu$ sec) digitalisiert und über ein Formatierungs-Interface dem Datenfeld im Kernspeicher des 16bit-Computers zugeführt. Die Akkumulation der Daten kann wahlweise in Matrizen- oder Originalform geschehen; sie werden in jedem Falle, blockweise mit Kennungen versehen, zum Plattenspeicher geschafft (siehe Abbildung 3).

Ein Doppel-Digital-Analog-Wandler besorgt zusammen mit einer Spezial-Hardware die Darstellung der Information des Datenfeldes, z.B. Bilder oder Zeitfunktionen. Ein Bedienungsfeld für Einstellung von Marken in X, Y oder C (counts) ist vorhanden. Mit diesen Marken können bis zu 16 verschiedene "interessierende Gebiete" gewählt werden. Der Massenspeicher hat im wesentlichen 2 Funktionen: er speichert die Daten in verschiedenen Betriebsarten (Originaldaten, Histogramme, Zeitfunktionen, "Regions-of-Interest"), und er enthält die Programme.

Bei der Programmspeicherung wurde von der Tatsache ausgegangen, daß an einem solchen System im allgemeinen kein Multiprogramming bewältigt wird, d.h. jeder Bediener des Systems kann annehmen, daß der Rechner nicht zu einem Zeitpunkt mehrere Aufgaben ausführt. Wenn das der Fall ist, kann auf ein aufwendiges Organisationssystem (ein Betriebssystem, wie z.B. im Zentralrechner) verzichtet werden. Da der Plattenspeicher nun einmal vorhanden ist, werden die Programme in Form von Kernspeicherabzügen (je einer für

jede Funktion) auf ihm abgelegt. Im Kernspeicher verbleibt lediglich ein Restmodul als Monitor, welcher das Tastenfeld überwacht und jeweils bei Betätigung einer Taste den notwendigen Kernspeicherabzug von der Platte holt und das Programm ausführt.

Dieses Betriebssystem bringt einen weiteren Vorteil: es ist leicht möglich, zusätzliche Kernspeicherabzüge herzustellen und einzubauen, um weitere Aufgaben bewältigen zu können.

Der Plattenspeicher ist aufgeteilt in:

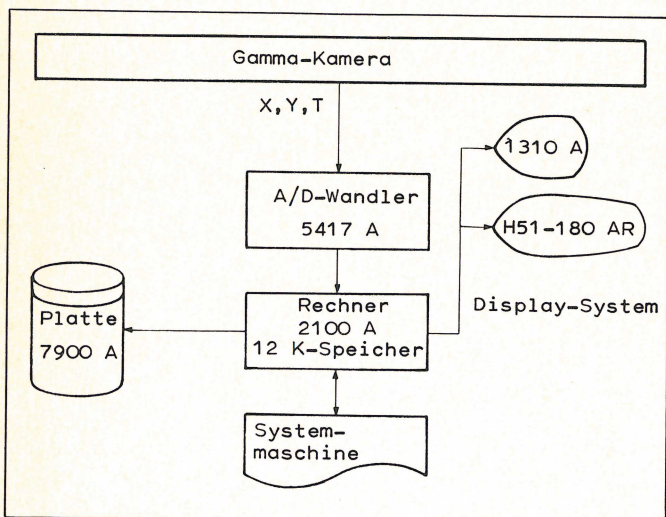
- Daten
- hergestellte Bilder
- hergestellte Zeitfunktionen
- einmal definierte "interessierende Gebiete"
- Patienten-Daten
- Programme

Die Systemkonsole und die dazugehörigen Ein/Ausgaberroutinen sind so aufgebaut, daß der Aufruf einer Betriebsart durch einen einzigen Tastendruck erfolgt. Die Tasten sind durch ein abnehmbares Deckblech entsprechend gekennzeichnet.

Mittels Tastendruck können folgende Operationen durchgeführt werden:

- Datenakkumulation,
- Herstellung von Bildern,
- arithmetische Operationen mit und zwischen den Bildern,
- Eliminierung der Nichtlinearitäten des Detektors (Normalisierung),
- Glättung der Bilder,
- Definition von interessierenden Bereichen mittels Lichtgriffel oder Marken,
- Herstellung von Zeitfunktionen,
- Glättung der Zeitfunktionen,
- Weiterverarbeitung der Zeitfunktionen durch mitgelieferte oder vom Benutzer geschriebene BASIC-Programme,
- Kommentare zu den Daten abspeichern,
- Abspeicherung von Bedienerbefehlen als Sequenzen auf der Platte und Abruf zum automatischen Ablauf einer Messung mit Auswertung,
- Einleitung eines Datentransfers zwischen Gamma-Kamera-System und Zentralsystem.

Abb. 3: Blockschaltbild Gamma-Kamera



Ein Beispiel soll die Arbeitsweise des Systems verdeutlichen: In der Betriebsart "Speicherung der Originaldaten" werden zwei Pufferspeicher im Ping-Pong-Betrieb hin- und hergeschaltet. In jeweils einem Speicher wird akkumuliert, im anderen auf Massenspeicher ausgelesen. Daher entsteht kein Informationsverlust. Die Folgefrequenz bei der Herstellung von Bildern aus diesen Daten ist nur begrenzt durch die statistisch limitierte Aussagekraft. Zum anderen kann die Wahl der Parameter nach der Aufnahme, d.h. nach einer ersten Beobachtung des Verlaufs der Speicherung geschehen. Da der Rechner eine Wortlänge von 16bit hat, können bei einer Auflösung von z.B. 128×128 Bildpunkten, d.h. 2×7 bit, noch maximal vier Radionuklide unterschieden werden.

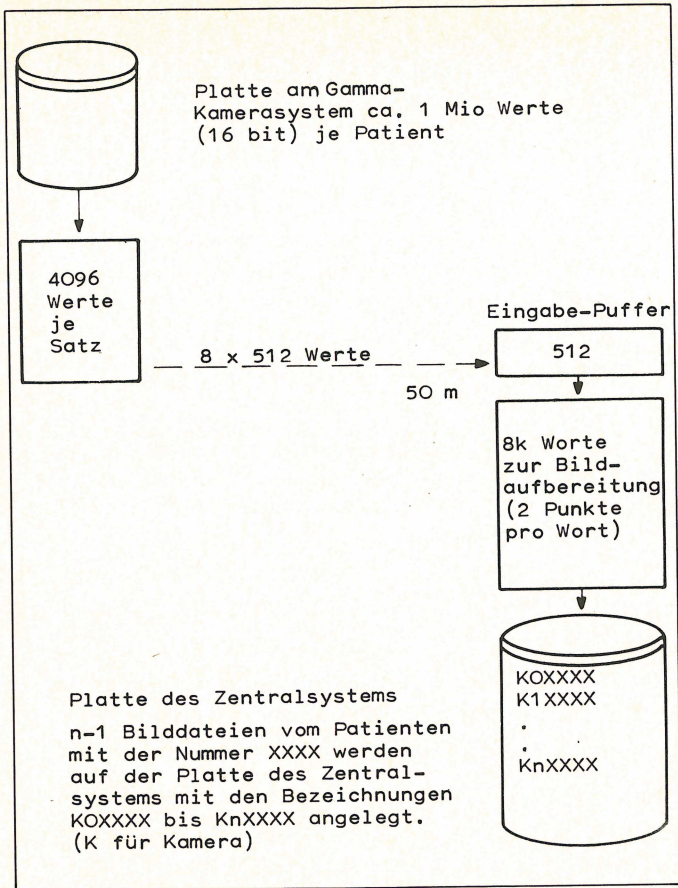
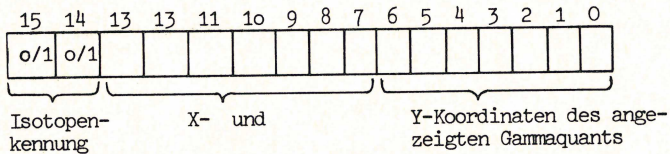


Abb. 4: Datenfluß der Gamma-Kamera

Datenwort Typ "Originaldaten":



Die Originaldaten einer Aufnahme (ca. 300.000 Werte) werden zu einem darstellbaren Szintigramm verarbeitet, in dem die X- und Y-Koordinaten eines Ereignisses als Indizes eines zweidimensionalen Feldes aufgefaßt werden. Der Inhalt der jeweiligen Feldelemente ist zunächst null und wird so oft um eins erhöht, wie das entsprechende Koordinatenpaar in den Originaldaten auftritt.

Auf diese Weise können im Rechnersystem an der Gamma-Kamera Bilder mit 64 x 64 Punkten erzeugt werden. Überträgt man die Originaldaten jedoch zum Zentralsystem, so können dort Bilder mit der max. sinnvollen Auflösung von 128 x 128 Punkten erzeugt werden, weil dort ein größerer Arbeitsspeicher zur Verfügung steht. Außerdem ist damit gewährleistet, daß die Meßergebnisse, welche zu einem Patienten gehören, auch dann zu einer Datenbank zusammengefaßt werden, wenn sie an verschiedenen Stationen aufgenommen wurden (siehe Abbildung 4).

Die Übertragung der Originaldaten zum Zentralsystem löst der Bediener der Gamma-Kamera durch einen speziellen Systembefehl

aus. Sie ist nicht zeitkritisch, denn i.a. wird während der Umrüstzeit für den nächsten Patienten übertragen.

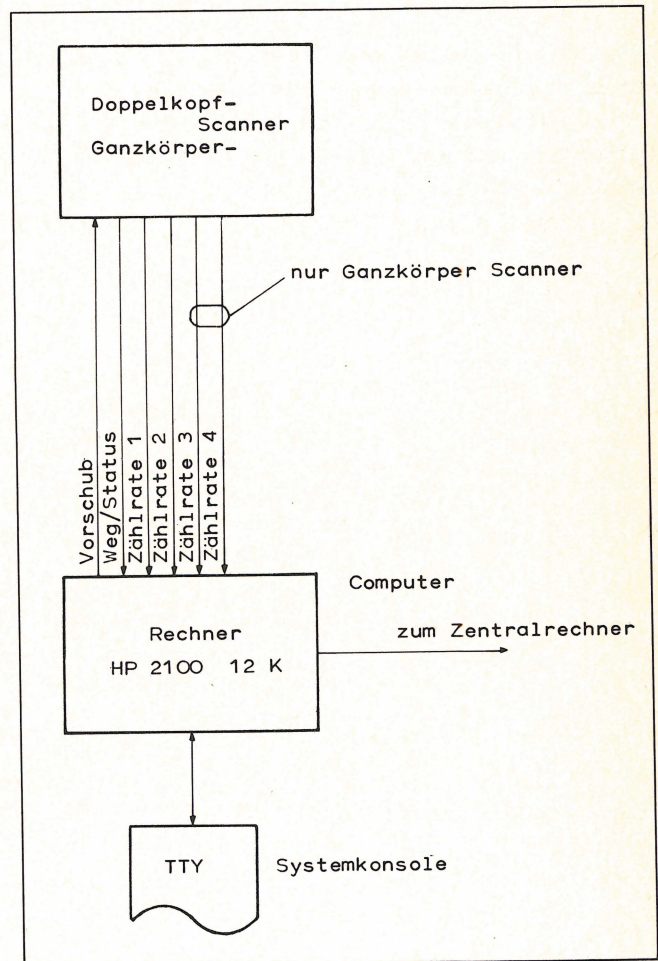
Doppelkopfscanner-System

Der Scanner tastet ein Feld zeilenweise, mäander- oder kammförmig ab und liefert Meßwerte (Impulse) und Statusinformationen (siehe Abbildung 5). Vier Interface-Karten besorgen die Datenübernahme und die Steuerung des Scanners durch den zugehörigen Rechner:

- eine 16bit-Duplex-Register-Karte erledigt die Übernahme der benötigten Statusinformationen aus dem Scanner (Vorschubinformation, Rechts/Links-bewegung, Betriebsart mäander- oder kammförmig, Zeilenende, Scanner Ein/Aus)
- zwei einfache Interrupt-Karten übermitteln dem Rechner, wann ein Impuls von einem der beiden Detektoren kommt.
- eine 16bit-Relais-Karte stellt die vom Programm vorzuziehenden Geschwindigkeiten am Scanner ein.

Der Rechner zählt über die Interrupt-Karten die Impulse und speichert sie nach dem Durchlaufen eines wählbaren Weginkrements, versehen mit Ortsinformationen, ab. Dadurch entsteht ein Abbild der Aktivitätsverteilung innerhalb des abgetasteten

Abb. 5: Blockschaltbild Doppelkopf- und Ganzkörperscanner



ten Feldes. Dieses Bild wird jedoch erst im Zentralsystem hergestellt, weil nur dort ausreichend Kernspeicherplatz dafür zur Verfügung steht und die fertigen Bilder auf dem dort vorhandenen Großbildsichtgerät und Farbsichtgerät dargestellt werden.

Die Geschwindigkeit des Scanners kann entweder fest vorgegeben werden, oder sie wird vom Rechner in Abhängigkeit von der letzten aufgenommenen Zählrate für die nächste Meßstrecke neu eingestellt. Das Ergebnis ist ein Bild der Aktivitätsverteilung mit in etwa gleichverteilter statistischer Genauigkeit. Gebiete mit zu geringen Zählraten werden mit maximaler Geschwindigkeit durchfahren, dadurch verringert sich die Meßzeit erheblich.

Ganzkörperscanner-System

Die Hardware- und Software-Struktur ähnelt der des Doppelkopf-Scanners. Im Gegensatz zum Doppelkopf-Scanner, bei dem im Normalfall Szintigramme mit Abmessungen von bis zu 35 x 40 cm erstellt werden, besteht beim Ganzkörper-Scanner die Möglichkeit, Ganzkörper-scans mit einer Aufnahme-Fläche von ca. 60 x 190 cm in Verbindung mit der Datenverarbeitung durchzuführen. Da die maximale Bildgröße für displayfähige Szintigramme auf 16 k (128x128 Bildpunkte) begrenzt ist, ergibt sich daraus zwangsläufig eine Zerlegung der Ganzkörperaufnahme in Teilbilder, die im Zentralsystem abgespeichert sind. Diese Teilbilder können wie die Aufnahmen der anderen Peripheriesysteme behandelt und modifiziert werden. Nach dem Glätten und der Hintergrundsubtraktion wird jedes Teilbild fotografiert. Die Zusammensetzung dieser Fotos ergibt eine Ganzkörperaufnahme mit den modifizierten Bilddaten, die gegenüber einer konventionell erstellten Aufnahme eine wesentlich höhere Aussagekraft besitzt.

Ganzkörperzähler-System

Der Ganzkörperzähler besteht in diesem Fall aus einer Stahlkammer, in der sich vier paarig gegenübergestellte Szintillationssonden in Längsrichtung des Patienten bewegen können. Eine umfangreiche Elektronik sorgt für die Steuerung der Bewegung der Köpfe, für die Hochspannung und die Weiterver-

arbeitung der Impulse durch vier Verstärker-/Analysator-/Zählerkanäle. Ausgabe der Zeitfunktionen und Profils cans auf Streifenschreiber sind ebenfalls vorgesehen.

Das Rechnersystem am Ganzkörperzähler arbeitet in vier verschiedenen Betriebsarten (siehe Abbildung 6):

1. Vielkanalanalyse
Für Kontaminationsuntersuchungen an Patienten und Personal der Klinik werden Energiespektren erzeugt. Ein mitgeliefertes, umfangreiches Paket für die Erzeugung, Verarbeitung und Ausgabe dieser Spektren auf typische Peripheriegeräte ermöglicht diese Betriebsart.
2. Zeitfunktionen
Der zeitliche Verlauf einer Aktivitätsspeicherung wird dargestellt.
3. Profils can
Die Verteilung der Aktivität im Patienten in Längsrichtung wird untersucht.
4. Ganzkörperscan
Durch Querbewegung des Patientenlagerungstisches zusätzlich zur Längsbewegung der Detektoren können Bilder der Aktivitätsverteilung im Körper des Patienten hergestellt werden.

Vielkanalanalyse

Das System quantisiert die analogen Signale der Detektoren am Ganzkörperzähler im Analog / Digital-Wandler (ADV). Die digitale Information wird in einer der Interface-Steckkarten im ADC solange gespeichert, bis der Rechner bereit ist, die Daten zu übernehmen.

Dieses Datenwort benennt die Adresse einer Kernspeicherzelle im Rechner. Nach Transfer dieser Adresse in den Rechner inkrementiert dieser den Inhalt der entsprechenden Speicherzelle. Das Ergebnis einer Anzahl derartiger Vorgänge stellt ein Histogramm dar - z.B. ein Gamma-Spektrum - welches im Rechner gespeichert ist. Das Histogramm wird auf dem Sichtgerät des Systems dargestellt.

Die Systemsoftware ist modular aufgebaut. Die Einzelmodule können mittels eines Frage- und Antwortverfahrens auf die

der jeweiligen Anwendung zugeschnittenen Betriebssysteme zusammengebunden werden. Folgende Aufgabenbereiche werden von der Software bewältigt:
Daten-Akkumulation,
Datendarstellung,
Drucken von Daten auf Systemkonsole,
Ein- bzw. Ausgabe auf Lochstreifen,
Ein- bzw. Auslesen von Daten auf Magnetband,
Datenmanipulationsroutinen,
Hilfsroutinen für selbstgeschriebene Programme,
Speichern und Verbessern von Systembefehlsfolgen zum späteren automatischen Ablauf.

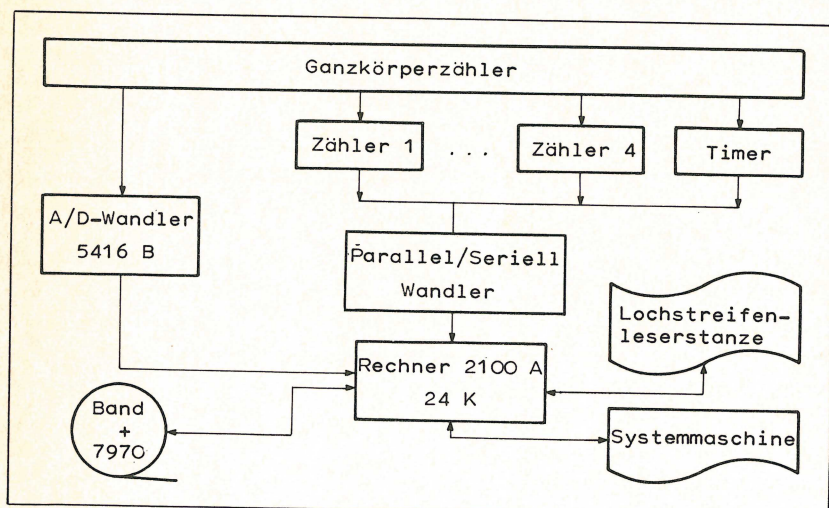
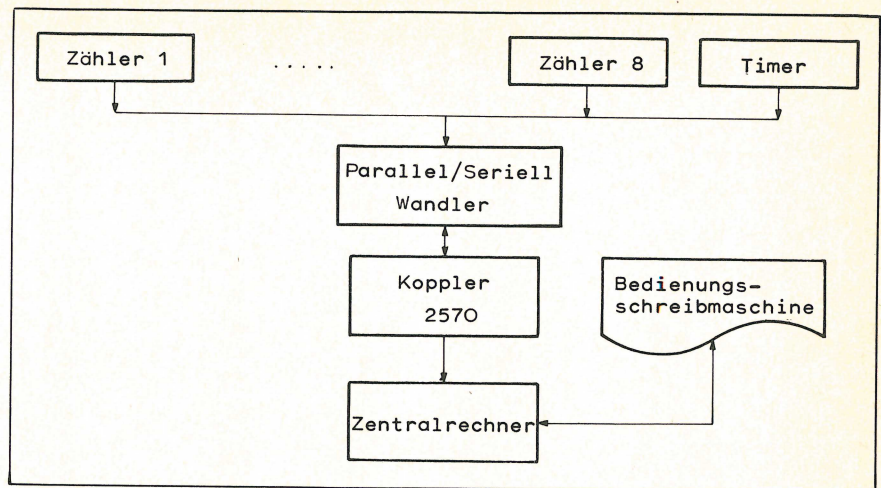


Abb. 6: Blockschaltbild Ganzkörperzähler

Abb. 7: Funktionsmeßplatz

Der Standard-Software-Umfang des Vielkanalanalysators wurde mit einem Paket erweitert, welches über die Elektronik des Ganzkörperzählers Daten übernimmt. Über jeden der vier mit Verstärker, Analysator und Zähler bestückten Kanäle werden die Daten (Zählraten) über einen Parallel-Seriell-Wandler und ein 8bit-Eingabe-Interface zeichenweise übernommen. Das Programmpaket GANZ arbeitet in drei verschiedenen Betriebsarten.



Zeitfunktionsmessung:

Ein Verlauf der Aktivitätsverteilung über die Zeit wird durch vom Bediener vorgegebene Parameter (Zeitinkrement, Anzahl der Meßpunkte, Anzahl der Zeitfunktionen) erzeugt. Die Funktion wird wie ein Spektrum dargestellt und kann mit den vorher erwähnten Analyseroutinen weiterverarbeitet werden.

Profilschan:

Die geometrische Verteilung der Aktivität über die Längsachse des Patienten wird dargestellt. Dabei gibt der Ganzkörperzähler die Vorschubimpulse (und somit die Weginkremente) für den Profilschan. Die Lage des Profilschans innerhalb der im Ganzkörperzähler vorgegebenen Länge, die Anzahl der Bildpunkte, die räumliche Auflösung (in cm/Bildpunkt) sowie die Anzahl der Bilder kann vorgegeben werden. Auch die Profilschans werden wie die Spektren und Zeitfunktionen vom Softwarepaket ausgewertet.

Ganzkörperpschan:

Bei Ganzkörperpschan gibt der Bediener folgende Parameter vor: den Anfangspunkt des Bildes, Anzahl der Bildpunkte in einer Zeile, Anfang der Zeilen, räumliche Auflösung in cm/Bildpunkt und Anzahl der Bilder. Die Vorschubimpulse innerhalb einer Zeile liefert wieder die Elektronik des Ganzkörperzählers, das Zeilenende wird durch Zurücksetzen eines Positionszählerstandes erkannt.

Funktionsmeßplatz

Am Funktionsmeßplatz werden von vier Zählsonden über je zwei Analysatoren mit acht Zählern bis zu acht Zeitfunktionen aufgenommen. Die Zählerstände werden über einen Koppler ins Zentralsystem transferiert und abgespeichert (siehe Abbildung 7) und zu displayfähigen Kurven aufbereitet. Die Patienten- und Untersuchungsdaten, die Gesamtmeßzeit sowie das gewünschte Zeitinkrement - während dessen die Zählimpulse integriert werden - sind vom Bediener im Dialogverkehr auf der TTY einzugeben. Die Steuerung des Kopplers für die Datenübernahme sowie der Dialogverkehr erfolgen vom Zentralrechner.

Betriebssysteme der Satellitenrechner und des Zentralrechners

Die Peripheriesysteme verfügen über ein sogenanntes Basic-Control-System, das hauptsächlich für den Input-Output-Verkehr, wie Dialog, Rechnerkopplung usw., zuständig ist.

Das zentrale Rechnersystem ist der Kopf der gesamten Anlage. Betrachtet man jedoch das Master-Slave-Verhältnis zwischen Zentralrechner und Satelliten, so ist in fast allen Betriebsarten der Zentralrechner Slave.

Sein RTE-Betriebssystem gestattet Multiprogramming in Vordergrund/Hintergrundbetrieb. Im Kernspeicher steht ein Laufbereich für das sogenannte Swapping und ein anderer für Segmentieren von Programmen zur Verfügung (siehe Abbildung 8).

Das Betriebssystem ermöglicht den simultanen Ablauf der Programme für die Patientenverwaltung, der Meßwert-Erfassung und -Verarbeitung aller Peripheriesysteme sowie der Operator-Kommunikation. Voraussetzung für den Simultanbetrieb ist die sofortige Unterbrechbarkeit des gerade im Vordergrund laufenden Programms (z.B. Meßwert-Erfassung) von Programmen höherer Priorität. Das unterbrochene Programm wird im aktuellen Zustand auf die Platte gespeichert und nach Ablauf des Programms höherer Priorität wieder in den Kernspeicher transferiert, um an der Unterbrechungsstelle fortzufahren.

Durch das Swapping-Verfahren kann der Benutzer alle Peripheriesysteme gleichzeitig für Untersuchungen verwenden. Das Swapping setzt jedoch für die Vordergrund-Programme Transferpuffer voraus, die im Commonbereich liegen.

Im Hintergrund des Kernspeichers werden hauptsächlich von den im Vordergrund laufenden Meßwert-Verarbeitungsprogrammen die Bildmatrizen aufbereitet. Ist der Hintergrund durch die Bildaufbereitung nicht belegt, können in diesem Kernspeicherbereich Programme laufen, z.B. für die Dosisplanung der Strahlentherapie. Dieses Programm benötigt jedoch mehr Kernspeicherplatz, als es der Hintergrund zuläßt. Das Programm wird deshalb in mehrere Segmente aufgeteilt, die nacheinander im Hintergrund ablaufen, wobei ein gemeinsamer Programmteil ständig im Kernspeicher steht.

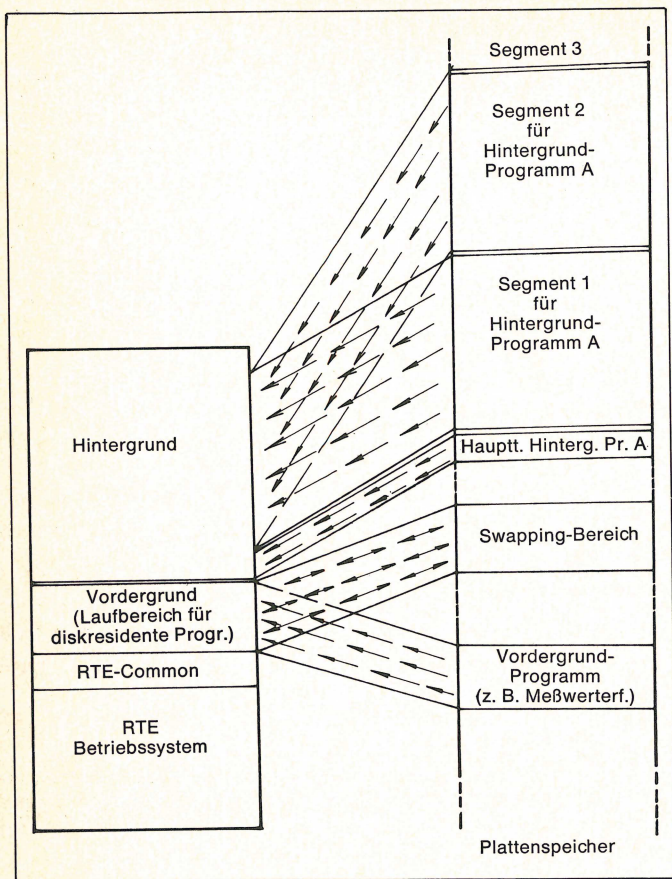


Abb. 8: Kernspeicher-Aufteilung im Zentralrechner

Auf den Plattenspeichern des Zentralsystems können Daten und Programme als Dateien (Files) abgelegt werden. Der Benutzer kann über vereinbarte Namen per Programm oder per Bedienerbefehl auf diese Files, deren maximale Anzahl nur durch die Plattengröße limitiert wird, zugreifen. Darüber hinaus können sie auch von Peripheriegeräten genutzt werden.

Organisation der Patienten- und Bilddateien

Die von den Peripheriesystemen transferierten patienten- und untersuchungsspezifischen Daten werden im Zentralsystem in entsprechende Dateien abgelegt. Der oberste Suchbegriff, unter dem Patientendateien zu finden sind, ist die 4-stellige Patienten-Identifikations-Nummer, kurz IDNR genannt. Die Zuordnungsdatei dient als Inhaltsverzeichnis aller im System eingetragenen Patienten und kann auf dem Schnelldrucker aufgelistet werden. Über diese Liste kann auf die gewünschte Patientendatei zurückgegriffen werden, in der alle Bildnamen mit den zugehörigen Untersuchungsdaten abgelegt sind. Der Diagnostiker kann über den Bildnamen auf die Bilddatei zugreifen und die gewünschte Patientenaufnahme auf dem Display darstellen. Die Zusammenhänge zeigt die Abbildung 9.

Meßwert-Erfassung und -Verarbeitung

Anhand eines Beispiels der Meßwert-Erfassung und -Verarbeitung für den Doppelkopf-Scanner sollen die entscheidenden

Vorteile eines Rechnerverbundsystems klar herausgestellt werden. Während der Satellitenrechner den Dialogverkehr, die Steuerung der Scanner und die Meßwert-Erfassung mit anschließender Datenkomprimierung übernimmt, wird der Zentralrechner innerhalb der Meßwerterfassungs-Phase durch die Übernahme der Rohdaten nur minimal belastet. Werden im Hintergrund zu dieser Zeit keine Bilddaten für andere Peripheriesysteme aufbereitet, kann das Programm für die Dosisplanung die Rechnerkapazität auslasten.

Vor der Untersuchung des Patienten werden vom Bediener über Dialogverkehr die Patientendaten, die untersuchungsspezifischen Daten (Angaben über die applizierten Isotope, die verwendeten Kollimatoren, die Lage des Patienten), ein freier Kommentar sowie die Betriebsdaten für den Scanner, wie Scangeschwindigkeit, obere und untere Zählratengrenze für Geschwindigkeitssteuerung, auf der TTY eingegeben. Am Ende des Dialogs werden die Patienten- und Untersuchungsdaten zum Zentralsystem transferiert und in die Zuordnungs- und Patientendatei abgelegt. Die Eintragung in die Zuordnungsdatei erfolgt nur, wenn der Patient im Zentralsystem zu diesem Zeitpunkt noch nicht erfaßt war.

Anschließend erfolgt die Einleitung des Scanvorganges durch den Satellitenrechner. Der Scanner tastet mit den beiden Meßköpfen den Patienten zeilenweise ab und liefert je nach Aktivität des gerade durchlaufenen Feldabschnittes entsprechende Impulsraten.

Das Weginkrement Δx , über das die ankommenden Impulse akkumuliert werden, ist abhängig von dem vorgewählten Zeilenabstand ΔY (z.B. $\Delta Y = 2 \text{ mm}$). Für eine maßstabsgerechte Darstellung sollte $\Delta X = \Delta Y$ sein.

Die Summe der Aktivität über der Meßstrecke X erscheint später als helligkeitsgesteuerter Bildpunkt auf dem Display (Abbildung 10). Auf dem Display sind Punkte mit hoher Aktivität heller ausgesteuert als angrenzende Bildpunkte mit niedriger Aktivität. Die über der Meßstrecke X erfaßten Impulsraten der beiden Szintillationsdetektoren werden über die tatsächliche Scangeschwindigkeit der zuletzt durchlaufenen Meßstrecke normiert und in einem Wechsellpuffer abgespeichert. Vor dem Überlauf der einen Hälfte dieses Wechsellpuffers erfolgt der Datentransfer dieser Rohdaten zum Zentralsystem in 256 Wort-Blöcken. Während des Transfers steht die 2. Hälfte des Wechsellpuffers zum Abspeichern von neu anfallenden Meßwerten zur Verfügung. Das im Zentralrechner gestartete Meßwerterfassungsprogramm für den Doppelkopf-Scanner übernimmt diese Rohdaten in den Common und speichert sie in einem Rohdatenfile auf der Platte ab.

Während der Meßwerterfassung am Scanner werden die Betriebszustände des Scanners laufend durch den Satellitenrechner überwacht und bei Fehlbedienungen entsprechende Fehlermeldungen sowohl auf der Bedienerkonsole als auch auf der System-TTY des Zentralsystems ausgegeben. Durch den Rechneranschluß besteht die Möglichkeit, die Geschwindigkeit des Scanners in Abhängigkeit von der gerade gemessenen Aktivität zu steuern. Über uninteressanten Gebieten mit niedriger Akti-

vität kann mit maximaler Scangeschwindigkeit, über Gebieten höherer Aktivität mit einer der Impulsrate proportionalen Geschwindigkeit gefahren werden.

Die Untersuchungen werden einerseits zeitlich verkürzt, zum anderen verbessert sich die Zählstatistik der Aufnahme. Am Ende der Untersuchung ist die Meßwertfassung abgeschlossen, das Meßwertverarbeitungs-Programm im Zentralrechner transferiert die auf der Platte abgespeicherten Rohdaten in den Hintergrund des Kernspeichers und ordnet diese in eine Bildmatrix ein.

Für die spätere Darstellung des Bildes werden Displayparameter (z.B. X und Y-Dimension) errechnet und als Vorspann an den Anfang der Bilddaten abgespeichert. Zur Bestimmung der tatsächlichen Größenverhältnisse bei der Wiedergabe des Bildes werden Abstandsmarken für die X- und Y-Koordinate eingebildet. Anschließend wird das displayfähige Bild unter dem vorher festgelegten Filenamen auf der Platte abgespeichert. Die Datenerfassung und Meßwertverarbeitung mit Bildaufbereitung im Zentralsystem laufen für die restlichen Peripheriesysteme ähnlich ab.

Operator-Kommunikation

Über ein Bedienungsprogramm hat der Arzt Zugriff zu den Patientendaten mit den zugehörigen Aufnahmen. Das Bedienungsprogramm kann in folgende Aufgabengebiete unterteilt werden:

- Ausdrucken der Patientenverzeichnis-Liste und der Patientendateien auf Schnelldrucker, sowohl vom Systembereich als auch vom Archivbereich
- Darstellung von Szintigrammen (128x128 Bildpunkte) und Funktionsuntersuchungen auf dem Display
- Manipulation der auf dem Display wiedergegebenen Aufnahmen, wie:

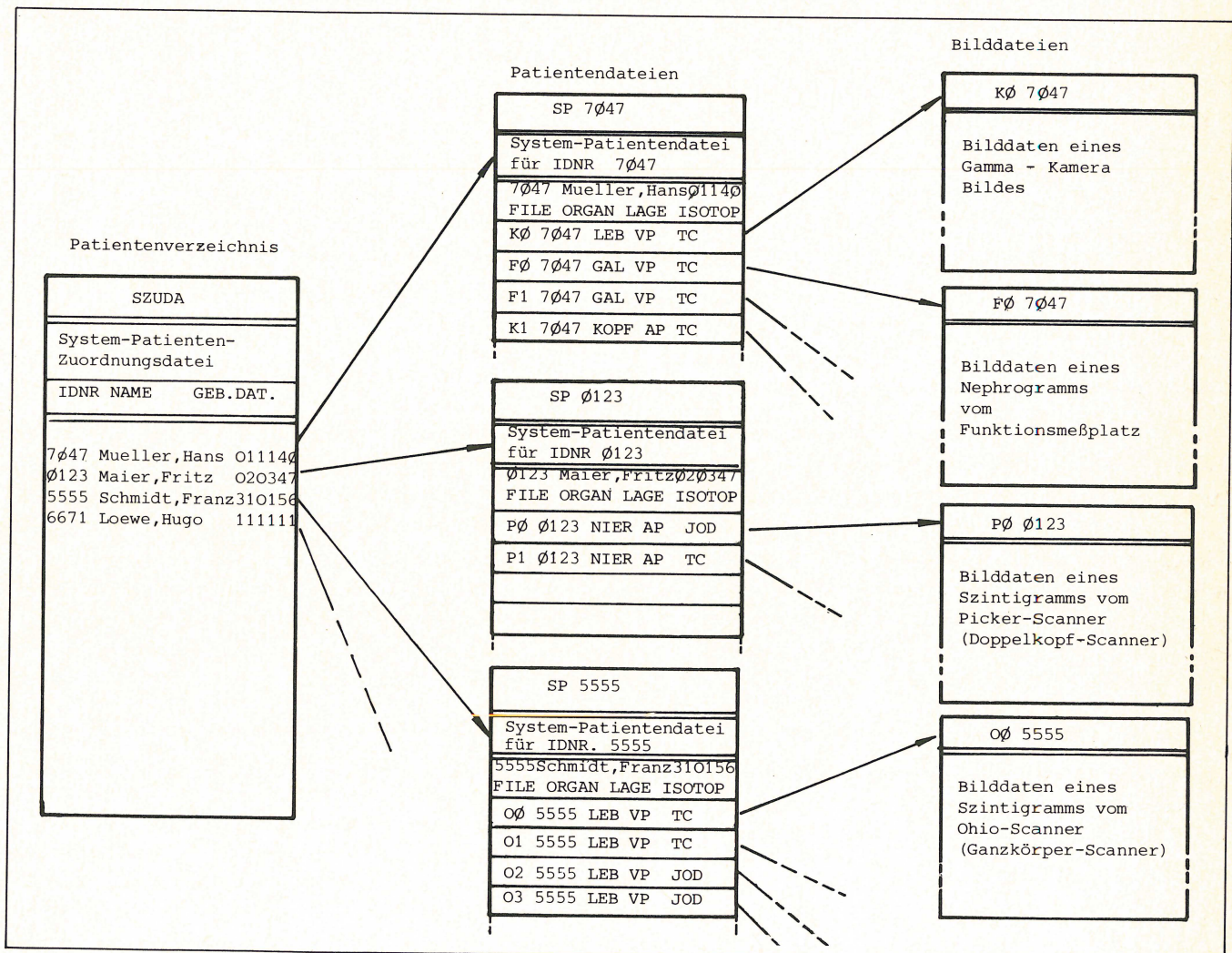
a) Bildarithmetik

Das dargestellte Bild wird einem Glättungsverfahren zum Ausgleich von statistischen Schwankungen unterzogen. Ferner können mathematische Operationen wie Multiplikation, Division, Addition und Subtraktion einer Konstanten mit jedem Matrixpunkt des Bildes ausgeführt werden. Eine praktische Anwendung wäre die Hintergrundsubtraktion zur Kontrastverbesserung.

b) Bildverknüpfung

An einem Beispiel soll der praktische Nutzen dieser Operation demonstriert werden. Die bereits auf dem Display

Abb. 9: Vereinfachte Darstellung der Datenorganisation



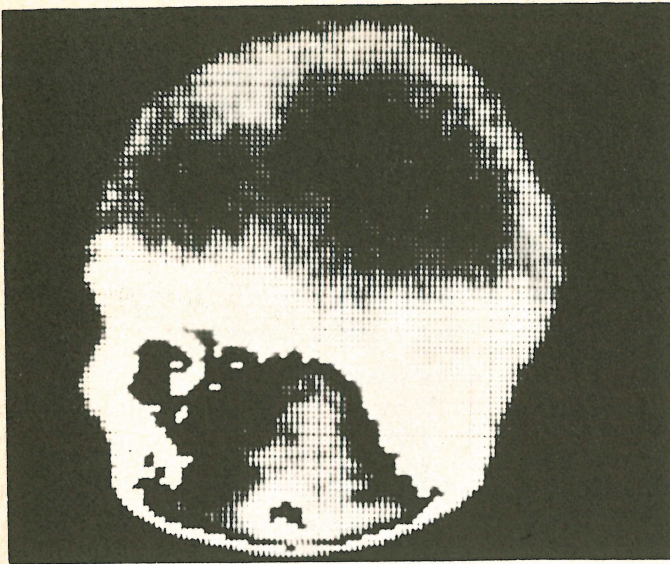


Abb. 10: Hirnszintigramm mit Befund

dargestellte Aufnahme des oberen Meßkopfes eines Doppelkopf-Scanners kann mit einem gleichzeitig entstandenen Bild des unteren Meßkopfes mathematisch verknüpft werden. Als Ergebnis der Applikation einer zweifach markierten Substanz wird durch die Addition der beiden Bildmatrizen des oberen und des unteren Meßkopfes auf einfache Weise die Summenaktivität ermittelt. Ein weiteres Beispiel wäre die Verknüpfung einer Gammakamera-Aufnahme mit einem sogenannten "Normalbild" zur Kompensation von Abbildungsfehlern bei der Gammakamera.

- c) Variation der Betrachterperspektive
Das auf dem Display wiedergegebene Bild kann für eine erforderliche Betrachtungsweise um 90 Grad und 180 Grad gedreht oder spiegelbildlich dargestellt werden.
- d) Bildausschnitt
Besonders interessante Gebiete auf dem Display werden von dem Betrachter durch "Marker" eingegrenzt und als Teilgebiet formatfüllend auf dem Display wiedergegeben.

Archivierung

Patienteneintragungen und die zugehörigen Aufnahmen können jederzeit auf der Archivkassette abgelegt werden. Die Löschung einer Aufnahme im Systembereich erfolgt nach der Archivierung automatisch. Modifizierte Bilder (Bildarithmetik) können unter Vergabe eines neuen Namens archiviert werden. Die mit der neuen Aufnahme zu archivierende

Patienteneintragung wird entweder direkt von der System-Patientendatei in die Archiv-Patientendatei übernommen oder aber mit einem neuen Kommentar versehen. Das Originalbild und die zugehörige Patienteneintragung bleiben in diesem Fall im Systembereich unverändert.

Der Anschluß von weiteren Peripheriesystemen ist durch den modularen Aufbau der Programmabusteine für die Meßwert-erfassung, die Meßwertverarbeitung und die Patientenver-waltung einfach zu realisieren.

An der Entwicklung und Implementierung der Anwendersoftware für das hier beschriebene Verbundrechnersystem haben noch mitgearbeitet: Frau Dr. Sanda Ornen, Herr Vijay Misra und Herr Jürgen Uhl (alle von INTERDIGIT GmbH, 7507 Pfinztal/Karlsruhe). Die Verfasser danken allen Beteiligten für ihre Mitarbeit und Beiträge.

Literaturverzeichnis

GREMMEL, H und WENDHAUSEN, H (Hrsg.) (1973):
Radiologische Klinik der Universität Kiel.
Verlag T.H. Mann, Hildesheim

Anschrift der Verfasser:

Prof. Dr.med. Helmut Gremmel
Dr.rer.nat. Henning Wendhausen

Radiologische Klinik
der Universität Kiel

23 Kiel

Arnold-Heller-Str. 9

Jürgen Pomplun
Klaus Stamer
Hewlett-Packard

2 Hamburg 1

Beim Strohause 26

Udo Günther

INTERDIGIT Gesellschaft f. Automation

7507 Pfinztal/Karlsruhe

Wesostr. 96

Entwicklung eines modular aufgebauten Programmsystems zur optimalen Erfassung, Speicherung und Auswertung von Daten bei klinischen Arzneimittelprüfungen*)

1. Mitteilung : Grundlagen und Ziele**)

Von E. Hartmann, G. Laudahn, J. Lehnert, H. Röpke, E. Rosenfeld

Zusammenfassung

Nach dem Stand der Wissenschaft ist zu fordern, daß Arzneimittelprüfungen im Sinne der klinischen Pharmakologie durchgeführt werden. Das bedeutet die biometrische Planung der Untersuchungen und ihre Durchführung als kontrollierte therapeutische Studien. Nur auf diese Weise ist eine Objektivierung von Arzneimittelwirkungen sowie eine quantifizierbare Abschätzung des Risikos möglich.

Die Zielsetzung unseres Vorhabens besteht in der Entwicklung und Erprobung eines der Allgemeinheit zur Verfügung zu stellenden, modular aufgebauten Programmsystems zur optimalen Erfassung, Speicherung und Auswertung der großen Datenmengen, die bei klinischen Arzneimittelprüfungen gewonnen werden.

Die dazu notwendigen Arbeiten umfassen die Problem- und Systemanalyse, die Programmentwicklung und die praktische Erprobung des zu erstellenden Softwaresystems.

In dieser 1. Mitteilung werden die Notwendigkeit eines solchen Programmsystems erläutert sowie die wissenschaftlichen Voraussetzungen und die Zielsetzung präzisiert.

Summary

The present state of science demands that the testing of drugs be encouraged under the aspect of clinical pharmacology. This means the biometric planning of investigations and their implementation as controlled therapeutic studies. Only in this way an objectivation of drug-effects and a quantifiable evaluation of the risks can be made possible.

Our project is aimed at developing and testing a modular program system - to be made generally available - for the optimal acquisition, storage and evaluation of the large amounts of data which are collected during the clinical testing of drugs.

The necessary work for this includes problem and system analysis, program development and the practical testing of the software system to be provided.

In this first report the necessity for such a program system is explained and the objectives and scientific requirements are defined.

Allgemeiner Stand auf dem Gebiet der klinischen Pharmakologie

Bis in die erste Hälfte dieses Jahrhunderts beruhte - von einigen Ausnahmen abgesehen - die "Prüfung" und Anwendung von Arzneimitteln im wesentlichen auf den individuellen Erfahrungen des Arztes. Somit unterlagen sie auch den Einflüssen der jeweils herrschenden "Schulmeinung". Damit war die Beurteilung jeglicher Pharmakotherapie überwiegend durch subjektive Kriterien bestimmt; ein objektiv begründetes und

*) Gefördert durch das Bundesministerium für Wissenschaft und Technologie, Projekt Datenverarbeitung in der Medizin, Teilvorhaben DVM 116.

**) Herren Prof.Dr. v. Eickstedt, Prof.Dr. Fuchs, P.Dr. May, Dr. Lorenz, Prof.Dr. Schneider und Prof.Dr. Westermann danken wir für ihre konstruktive Mitwirkung bei der Überarbeitung der Zielsetzung.

erst damit verallgemeinerungsfähiges Urteil über Vor- und Nachteile eines bestimmten Arzneimittels konnte auf diese Weise nicht zustande kommen.

Obwohl diese Situation schon lange Zeit als unbefriedigend erkannt war, wurden erst etwa 1950 aktive Maßnahmen zu ihrer Überwindung ergriffen. Überwiegend waren es folgende Gründe, die zur Entwicklung der klinischen Pharmakologie im heutigen Sinne führten:

- Die Einsicht, daß eine sinnvolle, wirksame und sichere Arzneimitteltherapie Kenntnisse voraussetzt, die nur durch systematische wissenschaftliche Untersuchungen zu erwerben sind.
- Die zunehmende Zahl der von der pharmazeutischen Industrie hergestellten Arzneimittel mit immer stärkerer und differenzierterer Wirkung.

- Das damit im Zusammenhang stehende Auftreten schwerwiegender Arzneimittelnebenwirkungen.
- Das zunehmende Interesse der Öffentlichkeit an Fragen der Arzneimittelwirkung und -sicherheit.

Gegenwärtig ist die Institutionalisierung der klinischen Pharmakologie in den USA, in Großbritannien und Schweden am weitesten fortgeschritten. Insbesondere in den USA wurde die Notwendigkeit einer Zusammenarbeit von industrieller und klinisch-akademischer Arzneimittelforschung auf dem Gebiet der klinischen Pharmakologie schneller erkannt als in Deutschland und dementsprechend intensiviert.

Es muß jedoch betont werden, daß in den genannten Ländern auch die steigenden Forderungen der Gesundheitsbehörden die Entwicklung wesentlich forciert haben. So ist besonders auf die 1962 in den USA erlassenen gesetzlichen Bestimmungen (KEFAUVER-HARRIS-Amendments) hinzuweisen, durch welche die pharmazeutischen Firmen zum Wirkungsnachweis ihrer neuen Arzneimittel gezwungen wurden.

Dagegen verlief die Entwicklung der klinischen Pharmakologie in Deutschland aus verschiedenen Gründen wesentlich langsamer:

- Die bereits in den 30er Jahren von P. MARTINI entwickelte Methodenlehre der klinisch-therapeutischen Forschung hatte keinen nennenswerten Widerhall gefunden.
- Impulse, die in der Nachkriegszeit durch die angloamerikanische wissenschaftliche Literatur gegeben wurden, führten kaum zu praktischen Konsequenzen; es blieb weiterhin bei einer Überbewertung der individuellen therapeutischen Erfahrung als Grundlage der Arzneimittelbeurteilung.
- Die Prüfung neuer Arzneimittel erfolgt in der Mehrzahl der Fälle auch heute noch ohne genügende Berücksichtigung biometrischer Planungs- und Auswertungsverfahren sowie ohne den Einsatz angemessener DV-Methoden.
- Im deutschen Arzneimittelgesetz besteht gegenwärtig noch keine Auflage, die Wirkung neuer Arzneimittel zu beweisen.

Innerhalb der deutschen Pharma-Industrie mußten sich diejenigen Firmen, die mit ihren Präparaten bereits im angloamerikanischen Markt vertreten waren oder sich diesen Markt für neuentwickelte Präparate erschließen wollten, als erste auf die verschärften Bedingungen einstellen. Das Bestreben der klinischen Forschungsabteilungen der Industrie, die Qualität der Arzneimittelprüfungen in Klinik und Praxis zu verbessern, konnte jedoch erst ab 1965 zu greifbaren Erfolgen führen.

Im Juni 1965 gab die Deutsche Gesellschaft für Innere Medizin Richtlinien für die klinische Prüfung von Arzneimitteln heraus. Die Richtlinien wurden zwar kein Bestandteil des Arzneimittelgesetzes, hatten jedoch normativen Charakter und konnten im Zweifelsfall konkrete rechtliche Bedeutung erlangen. Damit mußte auch in Deutschland der klinischen Pharmakologie mehr Beachtung geschenkt werden, und die Bereit-

schaft zur Durchführung kontrollierter Arzneimittelprüfungen nahm allmählich zu.

Die Industrie versuchte ihrerseits, diese Entwicklung zu stimulieren. Als Ansatz dazu sind die Bemühungen der PAUL-MARTINI-Stiftung (PMS) der Medizinisch-Pharmazeutischen Studiengesellschaft zu sehen. Die dieser Stiftung aus der deutschen Pharma-Industrie zufließenden Mittel wurden ab 1965 vorzugsweise dazu verwendet, jährlich 4 jungen Wissenschaftlern durch Stipendien eine klinisch-pharmakologische Fortbildung in den USA und England zu ermöglichen. Seit 1970 wird durch die PMS zusätzlich der Aufbau klinisch-pharmakologischer Abteilungen an deutschen Universitäten mitfinanziert. Der Nutzen dieser begrüßenswerten Initiative der PMS blieb jedoch bisher relativ gering, da Privatinitiative für die Durchsetzung des Fachgebietes Klinische Pharmakologie allein nicht ausreicht. Hier ist vielmehr festzustellen, daß Forschung und Lehre in der klinischen Pharmakologie nicht so gefördert wurden, daß die seit Jahren aus der internationalen Entwicklung erkennbaren Forderungen auf dem Arzneimittel-sektor auch in Deutschland nach Umfang und Qualität hätten erfüllt werden können.

Vom Bundesminister für Jugend, Familie und Gesundheit wurden am 11.6.1971 Richtlinien über die Prüfung von Arzneimitteln herausgegeben (H III 1 - 4410 - 11 - 7368/71 und H III 2 - 104.3 - 3233/71; Bundesanzeiger 23 Nr. 113 vom 25.6.71).

Von diesem Zeitpunkt an verfährt das Bundesgesundheitsamt nach dieser Richtlinie in der Weise, daß Arzneimittelspezialitäten nach § 21, Abs. 1a und Abs. 1b des Arzneimittelgesetzes vom 16.5.61 in der geänderten Fassung vom 23.6.70, die zur Anwendung beim Menschen bestimmt sind, nur dann in das Spezialitätenregister eingetragen werden, wenn sie nach diesen Richtlinien geprüft worden sind. Der Inhalt dieser Richtlinien wird in das in Vorbereitung befindliche neue Arzneimittelgesetz aufgenommen, das jedoch voraussichtlich nicht vor 1975/76 in Kraft treten wird.

Die Richtlinien fordern "kontrollierte klinische Untersuchungen", um die therapeutische Wirksamkeit und Unbedenklichkeit eines Arzneimittels objektiv feststellbar zu machen.

Allein durch die Herausgabe dieser Richtlinien für die klinische Prüfung von Arzneimitteln ist das gesteckte Ziel keineswegs zu erreichen.

Seine Realisierung setzt vielmehr voraus, daß ein standardisiertes und praktikables System zur optimalen Erfassung, Speicherung und Auswertung von Daten bei Arzneimittelprüfungen entwickelt wird, das dem Bundesgesundheitsamt Entscheidungskriterien für eine sachrelevante Beurteilung von Registrierungsunterlagen der pharmazeutischen Industrie in die Hand gibt.

Eine solche Möglichkeit ist durch die Entwicklung eines modular aufgebauten Programmsystems gegeben. Abgesehen von den BMD-Programmen (1), dem Statistical Program for Social Sciences (2) und dem statistischen Programmsystem "STATSYS"

(3), die zum Teil in unser Programmpaket eingearbeitet werden können, sind uns auf diesem Gebiet bisher keine weiterreichenden Systeme aus der einschlägigen Fachliteratur bekannt geworden. Dagegen können wir auf eigene Überlegungen (4,5) und auf die im folgenden genannten eigenen Vorarbeiten verweisen, die im Zusammenhang mit der Planung, Durchführung und Auswertung von Arzneimittelprüfungen geleistet wurden.

Durch konsequente Anwendung der Biometrie auf die spezifischen Versuchsbedingungen für die von uns bearbeiteten Präparategruppen und Indikationsgebiete, wurden Prototypen von Versuchsprotokollen entwickelt und praktisch erprobt. Diese Versuchsprotokolle haben die Form von dv-gerechten Fragebögen, in die der klinische Pharmakologe alle für die Arzneimittelprüfung relevanten Daten einträgt. Sie dienen zugleich als Einzelfallberichte, die der Gesundheitsbehörde zusammen mit der Gesamtauswertung zur Verfügung gestellt werden.

Beschreibung des Teilvorhabens DVM 116

Im folgenden werden die einzelnen Teilaspekte zur Erreichung dieses Zieles detailliert dargelegt. Die Gliederung erfolgt nach allgemeinen, biometrischen und programmierungstechnischen Gesichtspunkten.

Allgemeine Grunddaten der klinischen Prüfung

1. Verantwortliche Gesamtleitung (z.B. Firma, Abteilung)
2. Angabe der prüfenden Ärzte und der Prüfkliniken mit Ortsangaben
3. Prüfphase sowie Beginn und Dauer der Prüfung
4. Charakterisierung des Prüfpräparates
 - a) Bezeichnung (Nummer, Name, chemische Wirkstoffe)
 - b) Galenische Zubereitungsform mit Begleit- und Hilfsstoffen
 - c) Dosierung
 - d) Applikationsart
 - e) Applikationsfrequenz

Biometrische Erfordernisse für die klinische Prüfung

1. Zielsetzung: Beschreibung der Zielsetzung der jeweiligen Arzneimittelprüfung mit Angabe der Indikation (gegebenenfalls mit Lokalisationen) nach dem WHO-Code.

Beispiele:

- a) Definierte Anforderung an ein neues pharmakologisches Wirkungs- oder Anwendungsprinzip
- b) Definierte Verbesserung der Arzneimittelwirkung
- c) Verbesserung des therapeutischen Wirkungsspektrums
- d) Definierte Verbesserung der Verträglichkeit bei gleicher Arzneimittelwirkung (Vergrößerung der therapeutischen Dosierungsbreite durch Verringerung der Nebenwirkungsrate)

2. Begründung: Epikritische Bewertung der Ergebnisse früherer Untersuchungen und gegebenenfalls fremder Veröffentlichungen

zur medizinischen Begründung der Zielsetzung (Stand des Wissens einschließlich vorklinischer Befunde).

Beispiele:

- a) Erschließung einer neuartigen Therapie
- b) Medizinische Verbesserung bekannter Therapien
- c) Wirtschaftliche Verbesserung von Therapien (ambulante statt stationärer Behandlungsmöglichkeit; Senkung der Therapiekosten)

3. Grundgesamtheit: Definition der Grundgesamtheit, aus der die Stichproben gezogen werden. Festlegung von:

- a) Einschlußkriterien, wie z.B. Anamnese, Diagnose mit spezifizierten Kriterien, Rasse, Alter (geschlechtsspezifisch), Gewicht, Begleiterkrankungen und obligate Begleittherapien, Mindest- und Gesamtdauer der Behandlungen und Beobachtungen vor und nach der Therapie.
- b) Ausschlußkriterien, wie z.B. kontraindizierte Begleittherapien, Vorbehandlungen, vermutete Kontraindikationen und Behandlungsabbruch vor der Mindestbehandlungsdauer aus Gründen, die nicht mit der Prüfung im Zusammenhang stehen, prüfungsspezifische Angaben der sonstigen zulässigen Gründe für den nachträglichen Probandenausschluß von der Auswertung.
- c) Sonstige Angaben, wie z.B. geografische oder gegebenenfalls klimatische Angaben.

4. Wirkungskriterien: Definition der biologisch-medizinischen Hauptwirkungen und der tolerierbaren Begleitwirkungen sowie der möglichen schädigenden Nebenwirkungen, gegebenenfalls je Indikation:

- a) Bestimmungsmethode, wie z.B.: Visuelle Beurteilungen oder Messungen.
- b) Typ des Merkmals, wie z.B.: Häufigkeiten, multinominale Größen, Ranggrößen, Scoring-Systeme, Meßwerte.

5. Kontrollvariable: Angabe der Kontroll- und Kovariablen zur Adjustierung von Wirkungen mit Angabe der Bestimmungsmethode (z.B. Dauer der Erkrankung; Körpergewicht). Für nicht abzufragende Nebenwirkungen soll der WHO-Schlüssel verwendet werden.

6. Stichprobenumfänge: Festlegung von Gruppenumfängen auf der Basis der Fehler 1. und 2. Art sowie der zu fordernden relevanten Reaktionsunterschiede bezüglich der ausschlaggebenden Meßgrößen und Häufigkeiten. Außerdem sind hierbei Lage- und Streuungsparameter (z.B. Mittelwerte, Standardabweichungen, Häufigkeiten) von evtl. bekannten Kontrastbehandlungen (mit Quellenangaben) zu berücksichtigen. Bei kurzzeitigen Behandlungen kommen auch optimale Stichprobenumfänge (Sequenzanalyse oder Stufenpläne) in Betracht.

Angaben über die angestrebte Anzahl der prüfenden Ärzte und die geforderte Mindestanzahl der jeweils zugeordneten Probanden.

7. Therapiespezifikation: Definition der Applikationsfrequenz und der Dosis bzw. falls erforderlich der Dosen. Angabe des Applikationsbeginns, der Mindest- und Gesamtbehandlungsdauer und gegebenenfalls Festlegung der Vorbeobachtungsphase. Nur wenn die Dosis eindeutig über den Effekt ermittelt und gesteuert werden muß (z.B. Blutglucosespiegel), bleibt die Wahl der Dosis dem prüfenden Arzt überlassen.
8. Prüfanordnung: Aufstellung einer Prüfanordnung, die in der Regel als kontrollierte klinische Untersuchung geplant sein sollte, d.h. mit Kontrastgruppe und streng zufälliger Zuordnung der Behandlungen zu den Patienten (Randomisierung). Maßnahmen zur Ausschaltung dominierender Störungseinflüsse (z.B. durch Schichtung, Block- oder Paarbildung nach Vorbehandlung, Alter, Geschlecht). Die Prüfanordnung ist zu begründen.
9. Kontrastbehandlung: Definition der Kontrastbehandlung unter Berücksichtigung von Wirkstoffen, ihren Zubereitungen und Begleitstoffen, den Applikationsformen, dosierungen und Applikationsintervallen mit jeweiliger Begründung. Als Kontrastbehandlung kommen z.B. in Betracht: Vergleichspräparate, Spezialdiät, Placebo (falls ethisch vertretbar; insbesondere bei neuen pharmakologischen Wirkprinzipien).
10. Doppelblindprüfung: Planung der Prüfungen als Doppelblindversuche, sofern die Beurteilungen der angestrebten therapeutischen Wirkungen und der möglichen Nebenwirkungen durch die Kenntnis der Behandlungsart beim prüfenden Arzt bzw. beim Patienten beeinflussbar sind (bias of reporting; Vorurteile).
11. Datenerhebung: Zeitlicher Ablauf der Datenerhebung, orientiert an einem Bezugszeitpunkt (z.B. Behandlungsbeginn oder einem Zeitpunkt, der durch Zyklen oder andere Intervalle bedingt ist).
Entwurf von dv-gerechten und eindeutig beantwortbaren Fragebögen zur fixierten Erfassung der notwendigen Daten.
12. Auswertungsmethodik: Weitestgehende Festlegung der Auswertungsmethodik, bei der folgende Prinzipien zu berücksichtigen sind:
- Grundtypen der deskriptiven Tabellen und Grafiken
 - Art der Datenverdichtung im Längs- und Querschnitt
 - Angaben über vorgesehene Teilauswertungen (z.B. nach Ländern, prüfenden Ärzten, Vorbehandlungen) und Zwischenauswertungen
 - Angaben zu den statistisch zu testenden Hypothesen und Gegenhypothesen sowie Festlegung der relevanten Differenzen zwischen zwei bzw. der relevanten maximalen Differenzen bei mehreren Behandlungen.
- Angabe der akzeptablen Irrtumsrisiken je Indikation und Merkmal im Hinblick auf die Zielsetzung.

Beispiele:

Das Prüfpräparat soll bezüglich der einzelnen Wirkungskriterien im Vergleich zur Kontrasttherapie entweder mindestens um einen bestimmten Betrag besser bzw. verschieden oder ebenso gut, aber nicht mehr als um einen bestimmten Betrag ungünstiger sein.

13. Handlungsalternativen: Festlegung der Handlungsalternativen in Form einer Entscheidungsmatrix, die in Abhängigkeit vom Verlauf bzw. Ausgang der Untersuchungen zu ergreifen sind. (z.B. Wiederholung der gesamten Studie bei unerwartet hohem drop out unter Berücksichtigung der Ursachen.)

Angabe von Gründen, die zur vorzeitigen Beendigung der Prüfung und damit zur Ablehnung des Präparates führen können.

Programmentwicklung und Dateikonzept

1. Generelle Aspekte

Das zu entwickelnde Programmsystem soll weitgehend modular aufgebaut sein und dadurch eine Verknüpfung beliebiger Programme mit flexibler Datenübergabe gewährleisten. Die Möglichkeit der Integrierung bereits vorhandener Auswertungsprogramme soll durch die für das Datenhandling zu entwickelnden Routinen berücksichtigt werden.

Die Organisation der Programme soll klar erkennbare Schnittstellen aufweisen, um dem Benutzer zu ermöglichen, sowohl Programme, Unterprogramme als auch spezielle Routinen nach seinen Wünschen durch eigene Software zu ersetzen. Die Notwendigkeit dazu wird sich insbesondere ergeben, wenn maschinenorientierte Ein-/Ausgaberroutinen eingesetzt werden sollen bzw. wenn die Struktur einer Anlage (Wort/Byte) berücksichtigt werden soll.

Die Programmierung soll grundsätzlich in Fortran erfolgen. Der Einsatz von weitergehenden Fortranmöglichkeiten eines DV-Herstellers soll so weit wie möglich vermieden oder über-sichtlich im Programm gekennzeichnet werden.

Die Dokumentation des Programmsystems wird neben der Beschreibung der einzelnen Programme bzw. Module eine einfü-hrende Legende zur Zielsetzung und Handhabung mit Angabe der Voraussetzungen für die Verwendung der einzelnen Bausteine enthalten. Für jedes einzelne Programm werden Angaben über Organisation, Unterprogramme, Schnittstellen, Methode und Datenteil mitgeteilt. Ferner werden Testdaten und Ergebnisse zur Verfügung gestellt.

Zur Gewährleistung einer ständigen Anpassung an neue Methoden und Erkenntnisse ist ein permanenter Änderungsdienst vorgesehen, der die Information aller Benutzer sicherstellen wird.

2. Dateien

a) Erfassung und Kontrolle der Daten

Die Datenerfassung basiert auf den dv-gerecht entwickelten Fragebögen der klinischen Prüfung. Das Format und die Aufteilung der Bögen sind bei den vorgesehenen Eingabemedien (Lochkarte, Terminal, Bildschirm) frei wählbar, jedoch sollten verarbeitungstechnische Erfordernisse (z.B. Ordnungsbe-griffe, Sortiermerkmale) sowie spezielle Anforderungen für die biometrische Auswertung berücksichtigt werden.

Für das Erzeugen, Erweitern und Ergänzen der Dateien werden Updating-Programme entwickelt, die unabhängig von der Auf-teilung des Fragebogens sind. Mit diesen Programmen können jedoch lediglich formale Prüfungen der Daten auf Vollstän-digkeit und Sortierfolge vorgenommen werden. Für die logische Vollständigkeit und Validitätsprüfung wird ein spezielles Programm entwickelt, das dem Benutzer eine variable Formu-lierung der durchzuführenden Datenprüfungen gestattet. Eine Speicherung dieser Angaben mit dem zugehörigen Fragebogentyp ermöglicht dann einen Bezug bei Arzneimittelprüfungen der gleichen Indikationsgruppe und/oder Fragebogentypen. Die zu untersuchenden Daten können hierbei sowohl alpha-numerisch als auch numerisch sein.

b) Datenspeicherung

Zur Gewährleistung des Einsatzes vorhandener Auswertungs-software wird die Datenorganisation auf den vorgesehenen Datenträgern (Bänder, Platten) grundsätzlich sequentiell sein, und es wird eine konstante Satzlänge pro Patient vor-ausgesetzt. Durch entsprechende Dateihandling-Routinen wird dem Benutzer freigestellt, Satzlänge und Aufbau zu modifi-zieren.

Jede Datei kann sowohl komplett als auch als Teilmenge zu-sätzlich in anderer als sequentieller Form gespeichert wer-den (inverted, random; - s. auch Dateihandling). Die Notwen-digkeit hierfür wird sich insbesondere für die Beantwortung von speziellen Fragestellungen über ein Terminal (z.B. bei Nebenwirkungen) ergeben.

c) Dateihandling

Ein Schwerpunkt der zu entwickelnden Programme wird bei dem Datenhandling liegen. Diese Programme sollen die Untersu-chung, Umstrukturierung und Transformation der vorliegenden Daten ermöglichen, ohne jedoch die nach der Durchführung der Plausibilitätsprüfungen vorliegenden Dateien zu verändern. Weiterhin soll mit diesen Programmen die Bildung von Teil-kollektiven und Untermengen einer Datei sowie die Anpassung an die Eingabeanforderungen bereits vorhandener Programme ermöglicht werden. In diesem Zusammenhang muß untersucht werden, ob die Makrobefehle oder Unterprogrammaufrufe die für den Benutzer geeignetste Form des Programmeinsatzes sind. Die in den jeweiligen Betriebssystemen zur Verfügung stehen-den Hilfsprogramme (SORT/MERGE, COPY etc.) sowie die Edi-torbefehle des entsprechenden Timesharing-Systems (ADD, DELETE, REPLACE etc.) müssen ebenfalls einsetzbar sein.

3. Auswertungsprogramme

Bei den Auswertungsprogrammen handelt es sich in der Regel um Programme der analytischen und deskriptiven Statistik. Die vorhandenen Programme für die Durchführung statistischer Analysen und Tests werden insoweit benutzt, als sie sich in das vorgesehene Datenhandling einfügen lassen. Bei allen notwendig werdenden Modifikationen bereits vorhandener Pro-gramme soll die Verwendung numerisch einwandfreier Algorithmen garantiert sowie auf optimalen Speicherplatz und Rechen-zeitbedarf besondere Sorgfalt verwendet werden.

Die Programme für die Häufigkeitsanalysen müssen dagegen größtenteils neu entwickelt werden. Es handelt sich hierbei um Programme für die Schätzung bedingter und nichtbedingter Wahrscheinlichkeiten nach Auszählung der erfragten Informa-tionen sowie um deren tabellarische Zusammenstellung. Die Form der Tabellen soll für den Benutzer frei wählbar sein, so daß der Computerausdruck direkt als Registrierungsunter-lage für die Arzneimittelbehörde verwendet werden kann. Durch eine permanente Speicherung von Tabellentexten soll der Benutzer von Schreibearbeit entlastet werden.

Auch bei konsequenter Planung der Versuche und Auswertungen werden sich gelegentlich neuartige Fragestellungen ergeben, die mit den vorgesehenen Programmen nicht beantwortet werden können. Die problemlose und kurzfristige Integration von zu-sätzlichen Programmen muß daher gewährleistet werden.

Abschließende Bemerkungen

Zur Vermeidung von Fehlinterpretationen erscheint ein Hin-weis angebracht, welche Leistungen das entwickelte System nicht erbringen kann und soll.

Wegen der ständigen Weiterentwicklung des Standes der klini-schen Pharmakologie kann das System nicht alle speziellen biometrischen Problemanalysen von vornherein berücksichtigen. Es erübrigt sich also nicht die Denkarbeit zu neuen Frage-stellungen.

Vor allem soll das System keinerlei "Standard-Prüfvorschrif-ten" für bestimmte Indikationen oder Therapien liefern.

Es geht vielmehr von den Grundarten der klinischen Prüfme-thoden aus. Die angestrebten Prüfungstypen können sogar für verschiedene Indikationen und Therapien identisch sein. Bei der Handhabung des Systems wird daher nur auf allgemeine biometrische Begriffe, nicht aber auf Indikationen, Thera-pien, Applikationsarten und Präparatetypen Bezug genommen.

Das System kann auch nicht sämtliche Details aller denk-baren Prüftypen enthalten; es ist jedoch so konzipiert, daß es bausteinartig fortentwickelt werden kann.

Das System soll außerdem keine fertigen Fragebögen produ-zieren oder gar vorschreiben.

Die Art der Variablen kann je nach Art des speziellen Prüf-problems vom Benutzer beliebig gewählt werden. Auswertungs-methoden sollen nicht starr an die Variablen bzw. Skalen-arten angeknüpft werden, sondern im Rahmen des biometrisch Sinnvollen wählbar sein.

Literaturverzeichnis

- (1) DIXON, J.W. (1968): BMD, Biomedical Computer Programs, Los Angeles; University of California, Health Sciences, Computing Facility
- (2) NIE, N.H., BENT, D.H. and HULL, C.H. (1970): Statistical package for the social sciences (SPSS). McGraw Hill Book Comp., New York
- (3) VICTOR, N. und HÖRMANN, A. (1974): Das statistische Programmsystem "STATSYS". EDV in Medizin und Biologie 5, 27-30
- (4) HARTMANN, E. (1973): Zur Problematik der Interpretation von Studien über Arzneimittelnebenwirkungen. Methoden der Information in der Medizin 12, 91-97
- (5) RÖPKE, H. (1970): Einsatz des Digitalcomputers in der Forschung der pharmazeutischen Industrie. IBM Nachrichten 201

Anschrift für die Verfasser:

Dr. Horst Röpke

Fachbereich Physikochemie u. Informatik
Schering AG

1 Berlin 65

Postfach 65 03 11

Textverarbeitung in der Medizin

Von F. Wingert

Zusammenfassung

Die Notwendigkeit und die Bedeutung der Entwicklung automatischer Verfahren zur Textverarbeitung in der Medizin werden erläutert. Verschiedene Ansätze, vor allem das Verfahren von PRATT, werden dargestellt.

Wichtige Komponenten automatischer Verfahren zur Textverarbeitung sind Lexika und Regelsysteme als Teile einer Meta-sprache. Es wurde ein Segment-Lexikon und ein Regelsystem zur Segmentierung zusammengesetzter Wörter der deutschen und englischen medizinischen Terminologie entwickelt. Die Struktur dieses Lexikons und der Segmentierungsalgorithmus werden beschrieben, und es wird ein Ausblick auf Prinzipien und Möglichkeiten der Erweiterung der semantischen Informationen in einem solchen Segment-Lexikon gegeben.

Summary

The need and the value of automatic language processing methods in medicine are explained. Several approaches, especially the approach of PRATT, are demonstrated.

The main components of automatic language processing methods are dictionaries and a system of rules as parts of a meta-language. A root dictionary and a system of rules have been developed for segmentation of compound word forms in English and German medical terminology. The structure of this dictionary and the segmentation algorithm are described.

Principles and possibilities of additional semantical informations in such a root dictionary are pointed out.

1. Einleitung

Die Medizin ist eine datenintensive Wissenschaft, und das Spektrum reicht von numerischen Daten über Bilder bis hin zu Textdaten, die den größten Teil medizinischer Befunde beschreiben. Viele numerische Daten sind nur paraphrasierte Textdaten. So wird die Ungleichung {Blutzuckergehalt $>x$ } paraphrasiert als "Hyperglykämie". Dabei ist x ein numerisches Datum, das von verschiedenen Parametern wie Bestimmungsmethode, Alter des Probanden, Tageszeit, vorangegangene Nahrungsaufnahme und Arzt abhängt. Invariant ist nur der Typ der Ungleichung. Der als Zahlenwert ausgedrückte Blutzuckergehalt besitzt für sich allein einen geringeren Informations-

gehalt. Erst ein gemeinsamer Wissensstand bei Sender und Empfänger der Nachricht {Blutzuckergehalt = y } über die verschiedenen Einflußgrößen macht die Nachricht zu einer verwertbaren Information. In der sprachlichen Form sind diese Einflußgrößen dagegen bereits berücksichtigt.

Die Verarbeitung von Texten kann auf verschiedenen Niveaus geschehen, die sich vor allem in den Paraphrasiemöglichkeiten unterscheiden. Semantische Inhalte medizinischer Aussagen gehören im allgemeinen mehreren Klassen semantischer Informationen an. Diese Klassen sind nicht eindeutig festgelegt, müssen aber die Forderungen erfüllen, daß sie alle medizinischen Informationen umfassen und daß sie disjunkt

sind. Die Kriterien für Klasseneinteilungen hängen von den Retrieval-Bedürfnissen ab. Dabei steht die hierarchische Struktur der Klassen im Vordergrund, die der Hierarchie beim Retrieval entspricht. Innerhalb der gleichen Klassen dürfen nur Texte vorhanden sein, deren semantische Inhalte nicht wesentlich verschieden sind. Was wesentlich verschieden ist, hängt dabei wieder von der speziellen Anwendung ab. Grundlage eines Textverarbeitungssystems ist daher nach [12] ein Lexikon, das

- alle Träger semantischer Informationen enthält,
- Synonyma (formal verschiedene, aber semantisch gleichwertige Texte) zueinander oder zu einer Vorzugsbenennung in Beziehung setzt,
- jedem Element seine morphologischen, syntaktischen und semantischen Informationen zuordnet,
- zusätzlich alle Informationen enthält, die in einem speziellen System zum "Verständnis" einer Aussage notwendig sind. So kann ein lexikalisches Element der semantischen Kategorie "Erkrankungen" auf alle Elemente der Kategorie "Lokalisationen" verweisen, in denen diese Erkrankung auftreten kann.

Technische Systeme bestehen aus einer endlichen Menge vorgegebener Formulierungen, die durch Ja-/Nein-Entscheidungen ausgewählt werden (Checklisten). Sie haben den Vorteil der einfachen Speicherung und des einfachen Retrieval, schränken jedoch stark die Freiheit der Formulierung ein. Gewisse Freiheiten gibt es in manchen technischen Systemen durch freie Wahl von Modifikatoren.

Am stärksten verbreitet sind Systeme, in denen jedes Dokument ersetzt wird durch eine Menge von Deskriptoren. Die Deskriptoren sind Wörter oder Phrasen mit zentraler semantischer Bedeutung. Die Auswahl der ein Dokument beschreibenden Deskriptoren erfolgt so, daß der Informationsinhalt eines Dokuments möglichst gut durch den Informationsinhalt der Deskriptoren wiedergegeben wird. Systeme, die das Vorhandensein des Deskriptors im Text des Dokuments erfordern, sind descriptor-in-context-Systeme [1,2,3], Systeme, bei denen die Deskriptoren dem Text hinzugefügt werden, sind descriptor-out-of-context-Systeme (ICD [4], KDS [5], SNDO [6], MESH [7]).

Zur Zeit gibt es zwei Terminologien, die sowohl als descriptor-in-context-System als auch als descriptor-out-of-context-System verwendet werden können (SNOP [8] und SABIR-C [9]).

ICD und SNDO sind logisch geordnete Listen von Diagnosen mit einem starren numerischen Code. Der Vorteil numerischer Codes liegt auf der Hand: Die Nummern spiegeln die gewünschten Implikationen wider, und ihre Arithmetik ist sehr einfach. So können die Auswertungslogiken von technischem Personal erstellt werden, das keine Kenntnis der medizinischen Nomenklatur besitzt. Ein solcher Code läßt jedoch kaum Erweiterungsmöglichkeiten zu, wie sie gerade in der modernen

Medizin notwendig sind. Die Möglichkeiten von ICD und SNDO sind:

- Auflisten der erlaubten Phrasen innerhalb des Fachgebiets und ihrer korrekten Schreibweise,
- primitives Äquivalent für den semantischen Inhalt im Code,
- einige Thesaurus-Funktionen:
Synonyma werden zusammen aufgelistet, primitive Struktur der semantischen Bedeutung eines Wortes unabhängig vom Kontext.

Der LAMSON- [1,2] und der AGK-Thesaurus [3] sind einander sehr ähnlich. Der Benutzer kennt nur ein Texteingabe- und Textausgabesystem, da nur intern verschlüsselt wird. Die Wortelemente des Thesaurus sind in wenige semantische Klassen (beim AGK-Thesaurus 6) eingeteilt. Implikationen auf Wortebene werden im Thesaurus hinzugefügt (Pneumonie → Entzündung, Lunge). Der umgekehrte Weg, die Erkennung von Mehrwortphrasen, ist dagegen nur unvollkommen möglich und muß vom Benutzer vorgesehen werden. Paradigmatische Informationen auf Wortebene können ebenso entweder über Pointer oder direkt addiert werden. So wird die Information

OESOPHAGUS \subset GASTROINTESTINALES

addiert. Synonyma können in der gleichen Weise gehandhabt werden. Auf dieser Ebene sind die Konventionen der Synonymklassen und der hierarchischen Verkettungen die Äquivalente der starren Codestruktur von ICD, haben jedoch den Vorteil, daß diese Thesaurus-Organisation ein offenes System ist, das durch Erweiterungen der Liste der Terme nicht zerstört wird. Ein Nachteil ist auch die völlige Unterdrückung kontextabhängiger und grammatikalischer Einflüsse.

SNOP unterscheidet sich von den genannten Terminologien dahingehend, daß es eine echte Nomenklatur ist, die aus den vier Listen TOPOGRAPHIE (T), MORPHOLOGIE (M), AETIOLOGIE (E) und FUNKTION (F) besteht. Das Konzept der SNOP ist, daß jede pathologisch-anatomische Aussage sich in Elementaraussagen zerlegen läßt, die einer der vier Listen angehören. Jedem Element ist ein fünfstelliger Code zugeordnet. Das erste Zeichen des Codes bezeichnet die Liste (T, M, E, F), die restlichen vier Zeichen markieren die Stellung in der hierarchisch strukturierten Liste. So geht aus dem Code für LEBER (T5600) die Zugehörigkeit zum GASTROINTESTINALTRAKT (T5) hervor. Synonyma werden unter der gleichen Code-Nummer aufgelistet. Da es Wörter bzw. Phrasen mit Informationen aus verschiedenen Kategorien gibt, ist eine Implikationsstruktur vorgesehen, in der die fehlenden Informationen addiert werden.

Die formale Struktur der Elemente der SNOP ist

$w_1 \dots w_m$	IMP	CODE	.
-----------------	-----	------	---

Dabei sind w_1, \dots, w_m die Wörter des englischen Textes, IMP ist ein eventuell vorhandener Implikationscode und CODE ist der SNOP-Code für diesen Text.

Beispiele:

w_1 w_m	IMP	CODE
UTERUS LOWER SEGMENT	Ø	T823o
PHARYNX	Ø	T6o1o
PHARYNGITIS	T6o1o	M4o0o
INFLAMMATION BLENORRHAGIC	Ø	M4o4o

Das Ziel der automatischen Textanalyse ist die Abbildung einer Aussage in eine Folge von sogenannten TMEF-Statements, deren Information der Information der sprachlich formulierten Aussage äquivalent ist.

Die Aussage CARCINOM DES RECHTEN HAUPTBRONCHUS wird abgebildet in

T261o M81o3 Eoo0o F0oo0

(T261o: HAUPTBRONCHUS, RECHTS), (M81o3: CARCINOM).

Diese explizite semantische Struktur ist ein großer Fortschritt, wenn auch noch viele Wünsche offenbleiben. So kann nicht allen denkbaren Ordnungs- und Bezugskriterien der Topographie Rechnung getragen werden. SNOP enthält zwar die Terme ARTERIOLE (T4oo1) und CORTEX DER NIERE (T71o4). Die Lokalisation ARTERIOLEN DES CORTEX DER NIERE kann jedoch nicht durch einen einzigen topographischen Code repräsentiert werden. In der Topographie ist der systematischen deskriptiven Anatomie der Vorzug gegeben vor rein organbezogenen Anordnungen. Die ARTERIA RENALIS ist unter dem System KREISLAUF aufgelistet, und der Bezug zur Niere ist verlorengegangen. Diese Problematik tritt fast immer auf, wenn Teilstrukturen eines Organs beschrieben werden, die zu einer systematischen Struktur gehören. Sie kann jedoch statt direkt in der Nomenklatur auch über die Datenstruktur gelöst werden. Hier überschreiten wir die fließende Grenze von Semantik und Pragmatik. SNOP stellt eine Nomenklatur zur Verfügung, die

- atomare semantische Einheiten in
- möglichst weitgehender semantischer Strukturierung auf der Basis einer
- TMEF-Aussagenstruktur enthält.

Die TMEF-Struktur kann durch die Definition von Relationen zwischen verschiedenen TMEF-Statements sehr flexibel wechselnden Bedürfnissen angepaßt werden. Dies ist Aufgabe des pragmatischen Teils der Textanalyse.

3. Automatische Verarbeitung medizinischer Texte

Die Bedeutung des Computers in der modernen Medizin liegt weitgehend in der Verbesserung der Kommunikationsmöglichkeiten. Entwicklung und Anwendung solcher Kommunikationsverfahren setzen jedoch einen Dialog mit dem Computer in freier Sprache voraus. Dies bedeutet nicht, daß von einem Computer menschliche Intelligenz bei der Generierung und beim Verständnis von Texten erwartet werden kann oder soll. Der Dia-

log muß mit einem Benutzer geführt werden, der solche Texte erzeugen und verstehen kann. Die Aufgabe des Computers in diesem Dialog ist die Benutzung expliziter logischer Strukturen bei der Analyse von Texten, bei ihrer Einordnung und Paraphrasierung. Dazu gehört eine wohl definierte Nomenklatur als Referenz. Nach PRATT [10] muß die Anarchie eines vollkommen freien Textsystems vermieden werden, da sonst die gesamte Last der Strukturbildung dem menschlichen Benutzer aufgebürdet wird. Diese Forderungen führen hin zur Notwendigkeit der Definition einer Kunstsprache. Vor allem in der Pathologie sind Methoden der automatischen Textverarbeitung entwickelt worden [10, 14, 16 - 24], die aber meist noch unbefriedigende Resultate besitzen.

Die drei Hauptkomponenten eines Textverarbeitungssystems sind: Benutzer, Regelsystem, Referenz-Lexikon. Ziel ist die Erfassung von Dokumenten, ihre Verarbeitung und das Wiederfinden solcher Dokumente, die vom Benutzer gegebene semantische Relationen erfüllen. Die dabei zu leistenden Teilaufgaben sind

- (1) Erkennung und Validierung eines Textes durch morphologische und syntaktische Analyse;
- (2) Erkennung und Registrierung einer Dateneinheit und ihrer semantischen Bedeutung;
- (3) Abbildung der Daten in eine Struktur, die den Informationsgehalt der Originaldaten enthält.

Zur morpho-syntaktischen Analyse gehört die Erkennung der lexikalischen Einheiten (Roots, Suffixe, Wörter) und ihrer Wortkategorien (Substantiv, Adjektiv, Artikel, Verb, Pronomen, ...). Die morpho-syntaktische Analyse führt also zur Abgrenzung möglicher semantische Informationen tragender Einheiten und der sie in Beziehung zueinander setzenden Operatoren. Durch die semantische Analyse werden die in der morpho-syntaktischen Phase analysierten möglichen Informationsträger zu Dateneinheiten durch Addition semantischer Informationen. Die Bildung dieser Einheiten ist vom Text abhängig und berücksichtigt die syntaktischen Gegebenheiten. In stärkerem Maße gehen hier jedoch auch Sprachgewohnheiten ein, die sich über lexikalische Forderungen hinwegsetzen. So werden die Lokalisationen in

ENTZÜNDUNG DER APPENDIX UND DES HODENS bzw.

ENTZÜNDUNG DER APPENDIX DES HODENS

zu differenzieren sein und die Dominanz semantischer Informationen über lexikalische Korrektheit in einer Phase wie

CORTICALE NIERENZYSTE

berücksichtigt werden müssen. Kontextabhängige Relationen müssen hier ebenfalls analysiert werden (BLINDHEIT DURCH UNFALL - UNFALL DURCH BLINDHEIT, BECKENNIERE - NIERENBECKEN).

Bei der Abbildung der Daten in eine formale Struktur werden Regeln einzusetzen sein, die die spezifische Verkettung semantischer Informationen in medizinischen Aussagen beschreiben. Diese Phase kann zur Auflösung noch bestehender Mehrdeutigkeiten benutzt werden.

Die Verwendung morpho-syntaktischer und semantischer Informationen setzt Listen der Träger solcher Informationen und Regeln für die Kombination der Informationen voraus. Die Aufgabenverteilung auf Listen und Regeln ist nicht starr, sondern hängt von der jeweiligen Realisierung ab. So kann theoretisch jede erlaubte Aussage Element der Liste sein. Diese einfachste Stufe setzt umgekehrt voraus, daß der Benutzer nur Formulierungen benutzt, die in der Liste enthalten sind (Checklist-Verfahren, Komposition von Texten aus vorgegebenen Bruchstücken nach fixen Regeln).

Das andere Extrem wäre eine Liste, die nur das erlaubte Alphabet enthält und ansonsten jede Aussage über Regeln analysiert und verarbeitet. Wegen der Starrheit eines geschlossenen Systems mit erschöpfender Liste wird vom Benutzerstandpunkt her das optimale Verfahren mehr auf seiten regelintensiver Verfahren liegen, die um so komplizierter sind, je größer die Freiheit beim Benutzer ist.

Zur Verdeutlichung sei die SNOB zugrunde gelegt. Gegeben seien Listen aller vorkommenden Lokalisationen (T) und aller morphologischen Veränderungen (M). Erlegt man dem Benutzer die Restriktion auf, pro Aussage nur einen T-Begriff oder nur einen M-Begriff zu benutzen, dann sind alle Aussagen der Form

- M (T_k) k = 1,2,... und
- M_j (T) j = 1,2,...

zugelassen. Die Verarbeitung einer dieser Beschränkungen unterliegenden Aussage gliedert sich in die Stufen

- Erkennung der topographischen Einheiten,
- Erkennung der morphologischen Einheiten,
- Kombination aller morphologischen mit allen topographischen Einheiten.

Wird diese Beschränkung jedoch aufgehoben, dann treten bei der gegebenen Verarbeitungsvorschrift Kombinationen auf, die nicht der ursprünglichen Aussage entsprechen.

Diese Erzeugung falscher Informationen wird von PRATT unter dem Aspekt, daß keine echte Information verlorengelht, in Kauf genommen. Dabei wird Vollständigkeit auf Kosten der Präzision angestrebt. Jede Reduktion der Menge falscher Informationen führt sofort zum Problem der Mehrdeutigkeit von Wörtern wie UND, MIT, BEI usw. .

Die Erkennung von Synonyma ist relativ einfach, solange diese auf Wortebene liegen, da syntaktische Gegebenheiten praktisch keine Rolle spielen (MUSKEL - MUSCULUS, GELENK - ARTICULATIO, ...). Da Wörter als Bausteine von Texten in jedem System analysiert werden müssen, ist der Ersatz von Wortsynonyma anhand von Wortlisten lösbar. Dies geschieht schon in den bereits zitierten Thesaurus-Systemen. Bei Mehrwortsynonyma wird das Problem jedoch sehr viel schwieriger. Die zitierten Thesaurus-Systeme bürden die Last der Rekombination von Mehrwortsynonyma dem Benutzer auf. Da zudem syntaktische Einflüsse verlorengelhten, sinkt vor allem die Präzision. So wird in solchen Systemen bei der Frage nach

APPENDIX oder HODEN auch APPENDIX DES HODENS gefunden werden, und die Differenzierung ist erst beim Lesen des Originaltextes möglich. Andererseits muß der Benutzer bei der Suche nach Erkrankungen der Nierenrinde beachten, daß diese Begriffe unter CORTEX DER NIERE, NIERENCORTEX, NIERENRINDE, CORTEX RENALIS usw. gefunden werden können.

4. Ein Beispiel

PRATT [10] beschreibt ein sehr wirkungsvolles Verfahren zur Verarbeitung pathologisch-anatomischer Texte auf der Basis der SNOB, dessen Erfolge es wert erscheinen lassen, auf dem eingeschlagenen Weg weiterzugehen.

Es werden drei Phasen unterschieden:

- Vorbereitung einer Aussage,
- Grammatik finiter Zustände,
- Generierung der Datenstruktur.

Die Vorbereitung einer Aussage umfaßt:

- Umwandlung des Textes in ein Feld von Wörtern bzw. Kunstwörtern und Sonderzeichen,
- Ersatz von Abkürzungen durch ihre volle Form (L4 → vierter Lendenwirbel),
- Einführung von Kunstwörtern für obligate Formen ("consistent with" → cw),
- Aufbau eines Symbolfeldes parallel zum Wortfeld. Dieses Symbolfeld enthält pro Wort morphologische und syntaktische Informationen für alle Transformationen: Plural-Singular, Adjektiv-Substantiv, Ersatz von Wortsynonyma, Markierung als allgemein sprachliches Wort ohne spezielle semantische Informationen, Klassifizierung als möglicher Index für die nächste Phase, Operatoren.

Die Informationen in dieser Vorbereitungsphase sind in mehreren speziellen Listen enthalten, auf die über eine Endungsanalyse der Wörter zugegriffen wird. In der zweiten Phase werden mögliche semantische Einheiten durch einen Verifikationsalgorithmus anhand der SNOB zusammengestellt, bei dem Kontextabhängigkeiten durch ein "principle of longest match" berücksichtigt werden. In der dritten Phase werden redundante Informationen eliminiert und ein Feld formaler Aussagen der Form TMEF generiert.

5. Aspekte einer Kunstsprache

Die Notwendigkeit einer Kunstsprache beim Dialog mit dem Computer ist durch die Redundanz und die mangelnde Präzision der natürlichen Sprache begründet. Die durch die Kunstsprache zu erfüllenden Forderungen sind:

- Präzision,
- Redundanzfreiheit,
- axiomatischer Aufbau,
- Möglichkeit zur automatischen Verarbeitung,
- explizite Regeln von Sprachstruktur und Inhalt,
- formale Notation zur Erleichterung der Automatisierung.

Dafür muß der Reichtum der natürlichen Sprache an Paraphrasiemöglichkeiten geopfert werden.

Die Anzahl der in der medizinischen Nomenklatur verwendeten verschiedenen Begriffe mit semantischen Informationen ist wesentlich geringer als in der Umgangssprache, vor allem, wenn man sich auf Diagnosen beschränkt. Andererseits sind auch Endungen sehr häufig Träger semantischer Informationen. Ein Beispiel hierfür ist die Endung -ISMUS bei: ALDOSTERONISMUS, PLUMBISMUS, VAGINISMUS. Hier wird durch diese Endung oder durch Austausch einer neutralen Endung aus einer chemischen Substanz oder aus einem topographischen Begriff eine Krankheitsbezeichnung.

Umfangreiche Gruppen finden wir vor allem auch zwischen den Kategorien "Erreger" und "Erkrankungen", von denen hier nur einige Beispiele erwähnt seien [12]:

- AMOEBA - AMOEBIASIS
- GIARDIA - GIARDIASIS
- MALEOMYCES - MALEOMYCOSIS
- GEOTRICHUM - GEOTRICHOSIS

Eine Übertragung des PRATT'schen Verfahrens auf die deutsche Sprache ist zwar grundsätzlich möglich, setzt jedoch

eine starke Erweiterung des Referenz-Lexikons voraus, da in der deutschen Sprache wesentlich häufiger zusammengesetzte Wörter benutzt werden. Weitere Nachteile wären aus dem gleichen Grund durch die kompliziertere Implikationsstruktur und die notwendig höhere Änderungsfrequenz gegeben.

Zur Vermeidung dieser Probleme wird daher mit Morphemen statt mit Wörtern gearbeitet. Auf dieser Ebene ist das Verfahren von PRATT ohne weiteres übertragbar. Die einzige Änderung in einem Pilotprojekt wird der Austausch der ersten Phase der Textvorbereitung gegen eine Wortsegmentierung sein, durch die die morpho-syntaktischen Informationen bereitgestellt werden. Zusätzliche Vorteile sind:

- Verwendung semantischer Informationen schon in der Vorbereitungsphase. Die Verifikation in der Phase der Grammatik finiter Zustände erübrigt sich, wenn die Information in einem SNOP-Term gleich der Summe der Informationen seiner Morpheme ist.
- Die Benutzung von Positionsinformationen für Morpheme bezüglich der verwendeten Nomenklatur reduziert den Suchbereich in der Verifikationsphase.

Abb. 1: Teil eines Eingangs in das Root-Lexikon D_p und in das Suffix-Lexikon D_s auf der linken Bildseite. Einige Suffix-Familien auf der rechten Bildseite.

Ein vielleicht noch wichtigerer Aspekt ist die Multilingualität. Das zur Wortzerlegung verwendete Segment-Lexikon ist parallel für deutsche und englische Diagnosetexte auf-

Example:

root: CARCIN \$ 13 84 85 104 120 150 152 171

SSs: O \$ 15 17 104 . . . connector

OM \$ 85 . . .

OMA \$ 85 . . .

OMATA \$ 85

OMATICA \$ 171

OMATICI \$ 171 (inflectional form)

OMATICO \$ 171

OMATICUM \$ 171

OMATICUS \$ 171

OMATISCH \$ 171

OMATOES & 120 German adjective

OMATOSA \$ 120

OMATOSE & 152

OMATOUS / 120

Sample of strings which are valid:

CARCINO... { match 104
 must be followed by another root because letter O
 is a connector

CARCINOM match 85

CARCINOMA match 85

CARCINOMATOES match 120

gebaut worden. Da nur graduelle Unterschiede bei den Formulierungsgewohnheiten zwischen der deutschen und englischen Medizin bestehen, ist zu vermuten, daß das gleiche Codierungssystem für Deutsch und Englisch benutzt werden kann und der einzige Unterschied in einem sprachspezifischen Teil des Segment-Lexikons besteht.

PRATT verwendet drei Typen von Lexika:

- Liste produktiver Endungen bzw. Vorsilben,
- Liste einzelner vollständiger Wörter,
- SNOP in alphabetischer Sortierung.

Der Zweck der Liste produktiver Endungen bzw. Vorsilben ist die Kopplung zwischen Endung und syntaktischer Information. Dies gilt auch für die rudimentäre Wortliste. SNOP wird in der gleichen Form, in der sie publiziert ist, zur Verschlüsselung benutzt.

In diesen drei Listen manifestiert sich der Versuch, an Texte bzw. Textteile gebundene Informationen mit möglichst geringer Redundanz zu speichern. Informationen, die fest an Wortendungen gebunden sind, sollen also nicht in Wortlisten geführt werden. So genügt die Information

- ITIS = ENTZÜNDUNG,

eine Wortliste mit allen möglichen Vorkommen dieser Endung wäre wesentlich umfangreicher und enthielte nicht mehr Informationen. Der Nachteil liegt in der Erzeugung von Mehrdeutigkeiten, deren Grad mit der Reduktion des Informationsträgers steigt. So gilt die oben angeführte Regel nicht uneingeschränkt, wie die Beispiele

LINITIS (PLASTICA) und CAPITIS (+ CAPUT) zeigen.

Solche Fälle müssen mit speziellen Ausnahmelisten abgefangen werden, die ihrerseits Verarbeitungszeit und Speicherplatz benötigen. Bei der Reduktion der Informationsträger muß also ein Kompromiß zwischen dem Grad der Reduktion und dem Grad der dadurch erzeugten Mehrdeutigkeit geschlossen werden.

6. Wortsegmentierung

Bei der Segmentierung von Wörtern wird die Bildung zusammengesetzter Wörter durch Zerlegung in Wortteile (Segmente) rückgängig gemacht, die Träger semantischer und/oder syntaktischer Informationen sind. Die wichtigsten Vorteile eines Segment-Lexikons sind:

- Reduzierter Speicherplatzbedarf,
- einfachere Übersetzung medizinischer Texte von einer Sprache in eine andere Sprache,
- höherer Grad an Vollständigkeit bei geringerem Aufwand, verglichen mit einem Wortlexikon,
- Basis eines effektiven Alphabets zur Entwicklung einer Metasprache,
- Basis semantischer Regeln zur Interpretation und Paraphrasierung zusammengesetzter Wörter,
- Reduktion des Referenz-Lexikons.

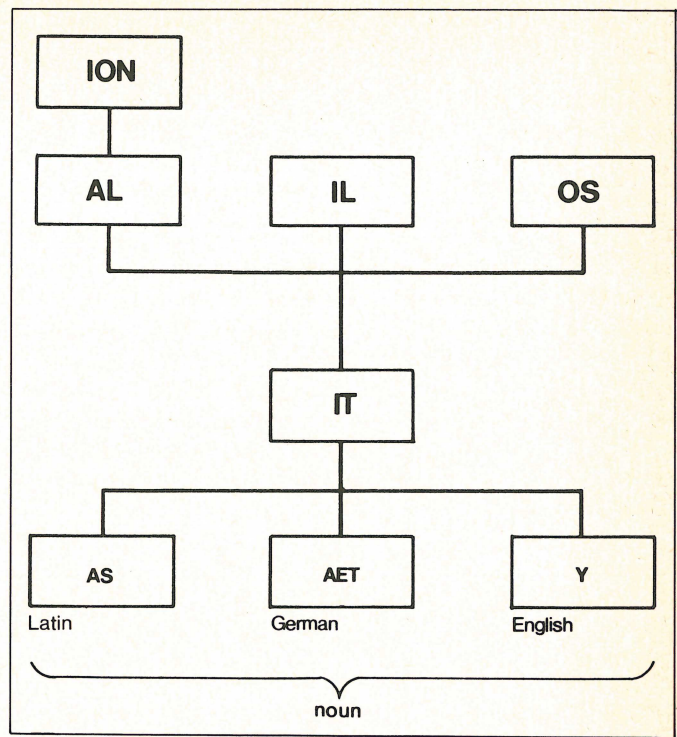


Abb. 2: Teil eines Suffix-Baumes. Jeder Kasten gehört zu einer anderen Familie.

Im Idealfall sind die semantischen Informationen eines Wortes gleich der Summe der semantischen Informationen seiner Segmente und die syntaktischen Informationen eines Wortes gleich den syntaktischen Informationen seines letzten Segmentes:

HEPAT ...	ITIS	NEPHR ...	EC ...	TOM ...	IE
↓	↓	↓	↓	↓	↓
LEBER	ENTZÜNDUNG	NIERE	HERAUS	SCHNEIDEN	Substantiv
	↓				
	(Substantiv)				

Da die Bildung zusammengesetzter Wörter vor allem in den europäischen Sprachen sehr ausgeprägt ist, ist auch in der allgemeinen Sprachforschung über dieses Problem gearbeitet worden [11]. Bedingt durch den Sprachumfang sind rein manuelle Verfahren sehr aufwendig. Erst der Einsatz des Computers macht eine Untersuchung auf breiter Basis möglich. Die allgemeine linguistische Literatur bietet bisher bei der Verarbeitung medizinischer Texte kaum Hilfe. Die Gründe sind:

- Beschränkung auf spezielle Wortfamilien,
- Spezialsprache der Medizin (deutsch, lateinisch, griechisch und ihre Kombinationen),
- Reduktion der sprachlichen Möglichkeiten, vor allem in Diagnostexten (Telegrammstil mit Vermeidung von Verben, Überwiegen von Substantiven und Adjektiven),
- kürzere Formulierungen bei semantischer Eindeutigkeit auf Kosten allgemeinsprachlicher Regeln.

Wichtige Arbeiten im Zusammenhang mit der Segmentierung medizinischer, vor allem pathologisch-anatomischer Texte, sind bereits geleistet worden [12, 13]. Der dort verwendete

Algorithmus behandelt jedes zusammengesetzte Wort als Folge X_1, X_2, \dots, X_n von Segmenten, die mit semantischen Informationen assoziiert werden können. Das Lexikon umfaßt etwa 1300 Segmente mit 500 unterschiedlichen Morphemen. Die Segmentierung wird jedoch sehr viel effektiver, wenn von der formalen Gleichwertigkeit der Komponenten abgegangen wird. Aus pragmatischen Gründen haben wir in der ersten Phase als kanonische Form eines Wortes die Form

$$w = R_1 s_1 R_2 s_2 \dots R_n s_n$$

gewählt. Jedes Wort ist alternierend aus einer Root (R) und einem Suffix (s) zusammengesetzt. Das Suffix kann auch fehlen, d.h. die Zeichenkette, die aus keinem Zeichen besteht (\emptyset), ist Teil der Menge der Suffixe. Beispiele für derart zerlegte Wörter sind:

NIERE ... N ... RINDE ... N ... CARCIN ... OME
 $R_1 \quad s_1 \quad R_2 \quad s_2 \quad R_3 \quad s_3$

LEBER ... ENT ... ZUEND ... UNG
 $R_1 \quad \emptyset \quad R_2 \quad \emptyset \quad R_3 \quad s_3$

7. Entwicklung des Segment-Lexikons

Zur Entwicklung eines Segment-Lexikons sind sehr umfangreiche Wortlisten nötig. Um den manuellen Arbeitsaufwand zu reduzieren, wurde daher ein vorwiegend computerorientiertes Verfahren entwickelt. Mittels eines Computerprogramms wurden die im AGK-Thesaurus [3] enthaltenen Wörter daraufhin untersucht, ob sie mit einer Root begannen, die selbst Wort in diesem Thesaurus war. In diesen Fällen wurde die längste, am Anfang eines Wortes stehende Root gestrichen. In vielen Fällen entstanden dabei Wörter, die ihrerseits bereits in der Liste enthalten waren oder wieder mit einem bereits im Thesaurus enthaltenen Wort begannen. Die in diesem Iterations-

prozeß neu entstandenen Wörter wurden in die Liste aufgenommen oder eliminiert, wenn sie bereits vorhanden waren. Die nach diesem Prozeß verbliebene Restliste wurde manuell von Artefakten gesäubert.

Danach wurde eine Liste der häufigsten Endungen erstellt und die Form des jetzt vorliegenden Lexikons gebildet, das aus einer Menge von Roots und einer Menge von Suffixen besteht, die über spezielle Regelklassen miteinander verkettet sind. Diese Regelklassen beinhalten:

- Syntaktische Informationen für Roots und Suffixe:
 Substantiv, Adjektiv, Adverb, Präposition, Artikel, ...
 Singular, Plural,
 Deklinationsklasse,
- Morphologische Informationen:
 R-s-Kombination, Markierung möglicher terminaler Segmente,
- Semantische Informationen:
 Chemische Substanzen (Salz, Alkohol, Keton, ...),
 Tumor (-OM), Entzündung (-ITIS), Negation, ...

Die erste Version des Lexikons umfaßte etwa 3.500 Segmente. Sie wurde ergänzt anhand der Wortlisten aus unserer deutschen SNOP-Übersetzung und aus dem Klinischen Diagnoseschlüssel [5].

Da wegen der gemeinsamen "Muttersprachen" Griechisch und Latein die deutsche und die englische medizinische Fachterminologie sehr ähnlich sind, lag der Verdacht nahe, daß ein großer Teil der Segmente identisch für eine deutsche und eine englische Version ist. So wurde die englische SNOP-Version probeweise mit dem deutschen Lexikon segmentiert. Diese Segmentierung war bei etwa 70 % der Wörter erfolgreich. Die Addition von etwa 20 speziellen englischen Regelklassen führte zu einer Erfolgsquote von über 80 %.

In der derzeitigen Version des Lexikons sind die deutschen und englischen Teile gemischt. Die Zugehörigkeit eines Segments zu einem der beiden Teile wird durch ein spezielles Steuerzeichen markiert. Das Lexikon enthält jetzt etwa

Abb. 3: Suffix-Familien S_{31} und S_{86} .

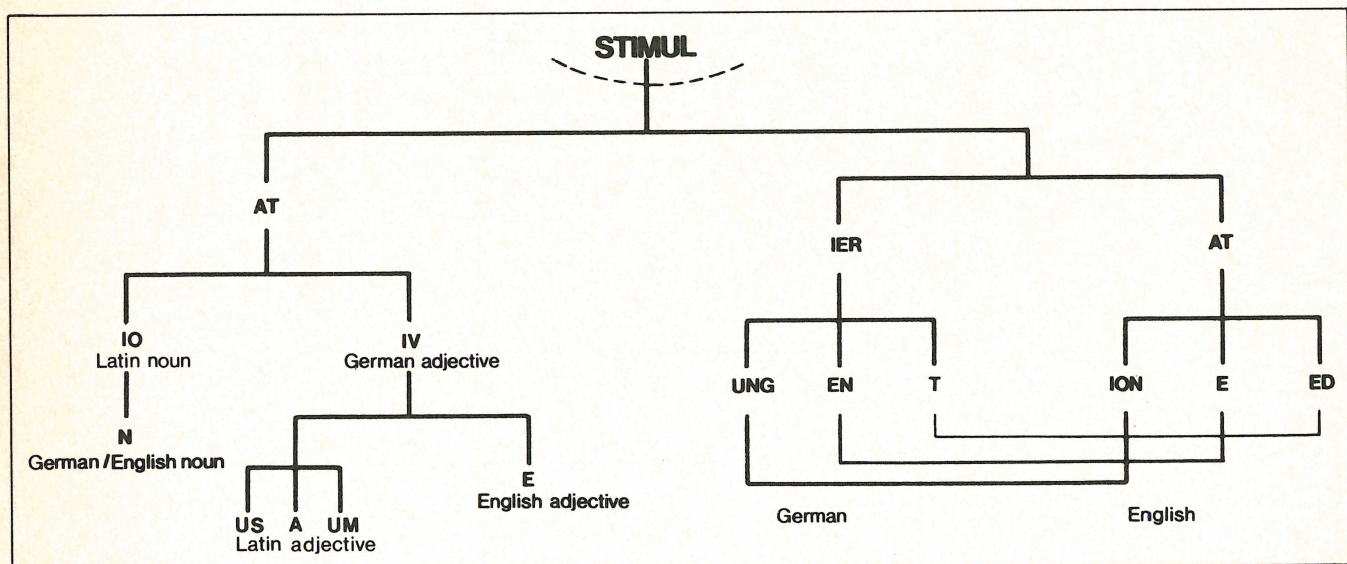
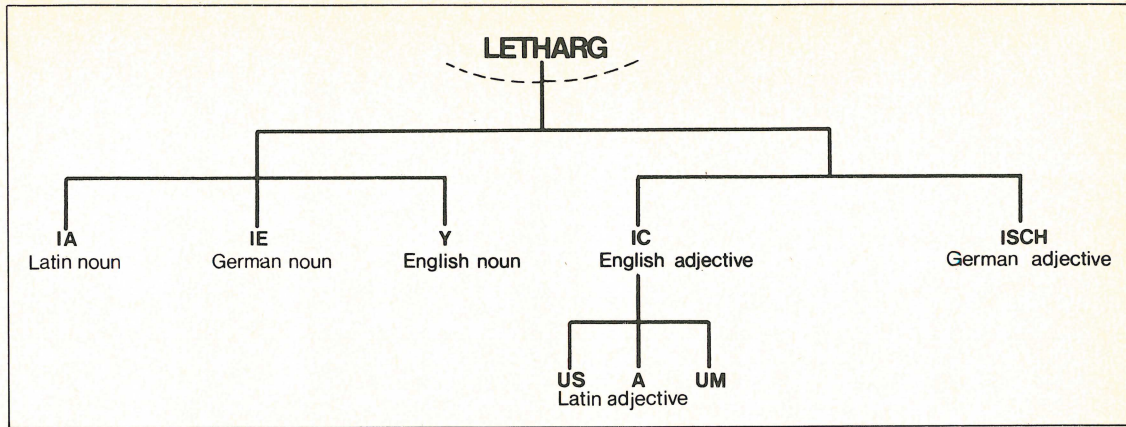


Abb. 4:
Suffix-Familien
S₃₃ und S₆₆.



8000 Segmente, wobei etwa zwei Drittel identisch für Deutsch und Englisch sind.

8. Struktur des Lexikons

Das Segment-Lexikon besteht aus einer Liste D_R von Roots und einer Liste D_S von Suffixen.

Die Grundform eines Elements (siehe Bild 1) ist

$$\left. \begin{matrix} R \\ s \end{matrix} \right\} L \quad i_1 \dots i_e \dots i_k \quad \begin{matrix} k \leq 14; \\ 0 \leq i_e \leq 255, e=1, \dots, k. \end{matrix}$$

Dabei bedeutet

R = Root

s = Suffix

L = Steuerzeichen für die Zugehörigkeit zur deutschen, zur englischen oder zu beiden Versionen

i_e = Nummer einer Regelklasse

Eine allgemeine Regelklasse mit der Nummer n enthält eine Familie S_n von Suffixen

$$S_n = \{s_{n1}, s_{n2}, \dots, s_{nm}\}, \quad 0 \leq n \leq 255,$$

wobei jedem Element s_{ni} eine spezifische morpho-syntaktische und/oder semantische Information zugeordnet ist. Ist mit einer Root R die Regelklasse mit der Nummer n assoziiert, dann bedeutet dies, daß alle Kombinationen

$$R * S_n = \{Rs_{n1}, Rs_{n2}, \dots, Rs_{nm}\}$$

erlaubte Sequenzen in einem Wort sind. Ist ein Suffix s mit einer Regelklasse n assoziiert, dann bedeutet dies, daß s Element der Familie S_n ist:

$$s \in S_n$$

(siehe Bild 1 bis 6).

Zur Zeit ist die maximale Anzahl von Regelklassen pro Root bzw. Suffix gleich 14.

9. Segmentierungsalgorithmus

Bei der Segmentierung zusammengesetzter Wörter können Mehrdeutigkeiten auftreten. So kann z.B. das Wort ECTOPIA als

$$\begin{matrix} EC \dots TOP \dots IA & \text{oder als} & ECT \dots O \dots PI \dots A \\ R_1 \emptyset R_2 & s_2 & R_1 & s_1 & R_2 & s_2 \end{matrix}$$

segmentiert werden. Diese Mehrdeutigkeiten sind dadurch bedingt, daß der Algorithmus, der sich im System der Regelklassen manifestiert, mehrdeutig ist und Eindeutigkeit in manchen Fällen nur durch den Kontext erzielt werden kann. Der eindeutige morphologische Träger der Information ist dann nur das ganze Wort oder eine ganze Phrase und Eindeutigkeit wird im beschriebenen Verfahren in solchen Fällen auch nur durch Aufnahme des zusammengesetzten Wortes in das Lexikon als Pseudo-Root oder durch Anwendung besonderer Regeln erzielt. Diese Fälle sind relativ selten, so daß sie das Lexikon kaum belasten. Ihr Auftreten ist jedoch einer der Gründe dafür, daß das System nicht statisch sein kann, sondern in einem ständigen Dialog und unter Benutzerkontrolle weiterentwickelt werden muß. Formale Ursachen von Mehrdeutigkeiten sind vor allem "Reaktionen" an der Nahtstelle zweier Segmente:

- Vokalelision (MYOOPIA → MYOPIA),
- Unterdrücken von Konsonanten, wenn die vorangehende Root mit dem gleichen Doppelkonsonanten endet, mit dem die nachfolgende Root beginnt und darauf ein Vokal folgt (STOFFFARBE → STOFFARBE),
- Geminatio (TAR → TARRING).

9.1 Formalalgorithmus

Alternierend wird von links nach rechts entsprechend der kanonischen Form eines Wortes nach einer Root bzw. einem Suffix gesucht. Die dabei benutzten Prinzipien sind:

- Es gibt ein Root-Lexikon D_R und ein Suffix-Lexikon D_S ,
- den Segmenten sind Regelklassen zugeordnet,
- der Leerstring ist Element von D_S ,
- die kanonische Form eines Wortes ist $R_1s_1R_2s_2\dots R_ns_n$, wobei jede Kombination R_1s_1 "erlaubt" ist,
- "erlaubt" ist eine Kombination R_1s_1 , wenn R_1 mit der Regelklasse S_k assoziiert ist und s_1 Element von S_k ist,
- zusätzliche Regeln entscheiden darüber, ob ein Segment letztes Segment eines Wortes sein kann.
- Gegeben sei die Sequenz

$$w = d_1d_2\dots d_iX,$$

wobei die Sequenz $d_1d_2\dots d_i$ erlaubt ist und X der verblei-

bende Reststring ist. Es wird nun nach dem längsten Segment d_{i+1} gesucht, so daß

$$w = d_1 d_2 \dots d_i d_{i+1} Y$$

ist und die Sequenz $d_1 d_2 \dots d_{i+1}$ erlaubt ist (Prinzip des längsten erlaubten Segments).

Wird ein solches d_{i+1} gefunden, dann wird das Verfahren mit Y fortgesetzt. Wird ein solches d_{i+1} nicht gefunden, dann wird ein Fehler in der bereits gefundenen (formal erlaubten) Sequenz angenommen. Zunächst wird angenommen, daß d_i falsch ist, daß also die korrekte Folge

$$w = d_1 d_2 \dots d_{i-1} t_i Z$$

ist, wobei t_i um mindestens ein Zeichen kürzer ist als d_i . Wird ein solches t_i gefunden, dann wird der Algorithmus mit Z fortgesetzt. Wird ein solches t_i nicht gefunden, dann wird angenommen, daß d_{i-1} falsch ist, etc. .

Der Algorithmus endet erfolgreich, wenn

- (1) das ganze Wort in eine erlaubte Sequenz, also in kanonische Form zerlegt ist, oder erfolglos, wenn
- (2) kein R_1 gefunden wird.

9.2 k- und z-Filter

Zur Reduktion des Lexikons werden formal verschiedene Varianten eines Segments, die sich nur in der k-, z-, c-Schreibweise unterscheiden, unterdrückt. Sofern eine c-Variante existiert, ist nur diese c-Variante im Lexikon enthalten. Wir benutzen hier das gleiche Prinzip wie im AGK-Thesaurus [3]. Die Übersetzung der anderen Varianten wird vom Programm dann vorgenommen, wenn ein Segmentierungsversuch erfolglos war und das Wort mindestens ein Zeichen "k" oder "z" enthält. Die durch dieses Verfahren bedingten Mehrdeutigkeiten werden in der bereits beschriebenen Weise behandelt.

10. Addition zusätzlicher semantischer Informationen

Zur automatischen Textanalyse genügt ein Segment-Lexikon nicht, da es viele Einheiten aus mehreren Segmenten gibt, deren semantische Informationen nicht gleich der Summe der semantischen Informationen ihrer Komponenten ist. Solche Einheiten können Wörter oder auch Mehrwortphrasen sein. Für sie benötigt man besondere Lexika, und die Segmentierung ist hier von vorwiegend technischem Wert. Da als eigentliche Verschlüsselungsgrundlage für Diagnosetexte die SNOP dienen soll, können zusätzliche semantische Informationen in das Segment-Lexikon aufgenommen werden. Wir unterscheiden dabei semantische Positionsvektoren und eigentliche semantische Informationen. Dabei soll in diesem Zusammenhang der Begriff "Semantik" bezüglich der SNOP verstanden werden, d.h. der semantische Inhalt eines Wortes oder eines Strings ist repräsentiert durch die Stellung, die dieses Wort bzw. dieser String in der Baumstruktur der SNOP besitzt. Die Stellung in der Struktur der SNOP ist codiert durch den SNOP-Code.

10.1 Positionsvektoren

Ein Positionsvektor besteht aus vier Komponenten, die den vier semantischen Kategorien der SNOP entsprechen. Jede Komponente enthält einen vierstelligen Code, der einen Knoten in der Baumstruktur der SNOP bezeichnet. Sind alle vier Stellen des Codes besetzt, dann entspricht der Code einem echten SNOP-Code und damit einem Eingang in die SNOP. Sind nur die ersten drei Stellen besetzt, dann entspricht der Code dem Knoten, von dem aus alle SNOP-Eingänge entspringen, deren vierte Stelle besetzt ist. Analog sind Codes mit besetzten ersten beiden Stellen bzw. Codes mit besetzter erster Stelle bzw. Codes, die nur aus dem Listensymbol bestehen, definiert. Konkret bedeutet dies, daß die Positionsvektoren eine spezielle Form einer nach dem Segment invertierten Liste ist, da jede Komponente den

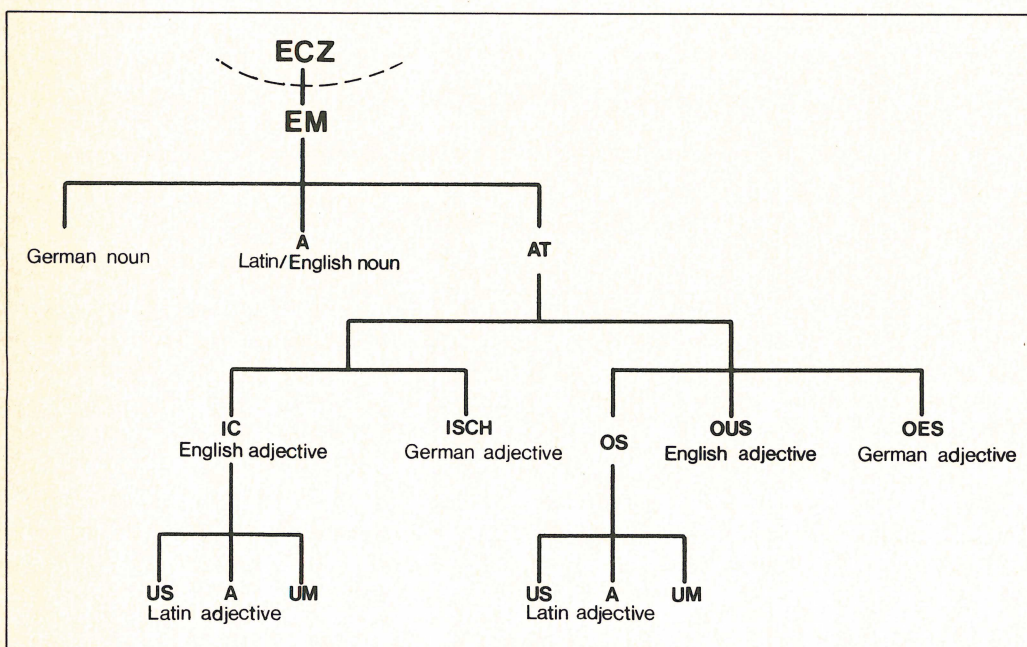
Knoten bezeichnet, unterhalb dessen das Segment in mindestens einem SNOP-Eingang vorkommt und außerhalb dessen das Segment in keinem SNOP-Eingang der gleichen Kategorie vorkommt.

Beispiel für die Baumstruktur der SNOP:

- T 5 Verdauungssystem
- T 56 Leber
- T 561 Leber, Glisson'sches Dreieck
- T 5611 Leber, intrahepatischer Gallengang

Abb. 5:

Suffix-Familie S_{80} .



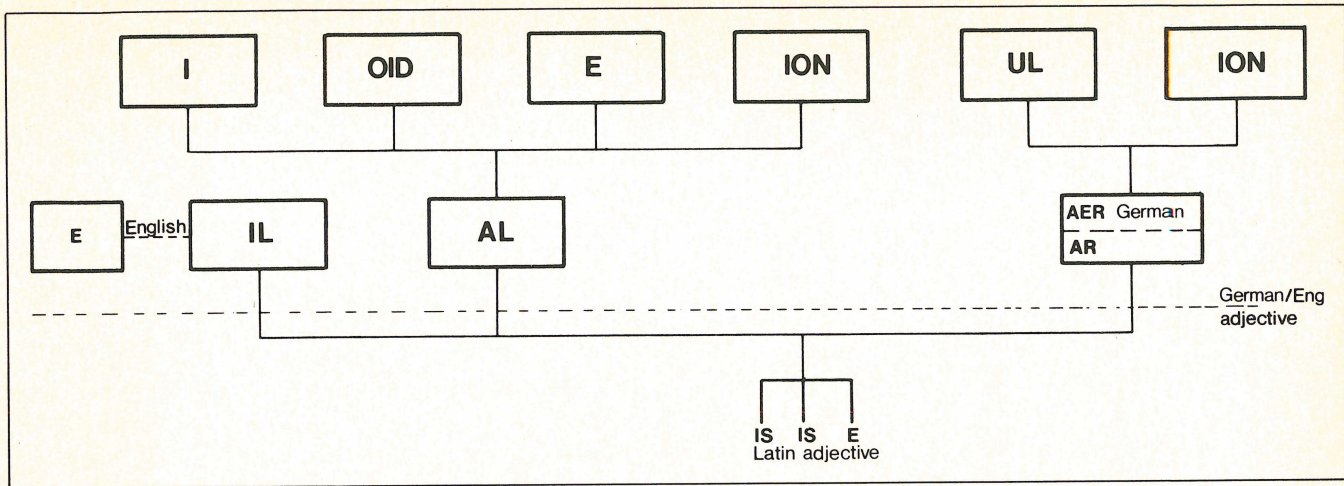


Abb. 6: Teil eines Suffix-Baumes. Jeder Kasten gehört zu einer anderen Familie.

10.2 Semantische Informationen

Eine einem Segment zugeordnete semantische Information ist ein TMEF-Statement, das 1 bis 4 SNOP-Codes der vier semantischen Kategorien enthält. Diese Information wird nur solchen Segmenten zugeordnet, die in einem SNOP-Eingang ohne zusätzliche semantisch signifikante Segmente vorkommen. Sie kann in der Verschlüsselungslogik benutzt werden als Code für nicht in der SNOP gefundene Segmente. Daher können alle SNOP-Eingänge gestrichen werden, die nur aus einem semantisch signifikanten Segment bestehen. Zusätzlich können jedoch alle SNOP-Eingänge gestrichen werden, deren semantische Information gleich der Summe der semantischen Informationen seiner Segmente ist.

Beispiel:

Segment	Semantische Information			
APPEND	T6600	M-	E-	F-
ECTOMIE	T-	M1520	E-	F-
ITIS	T-	M4000	E-	F-

Mit diesen semantischen Informationen im Segment-Lexikon können z.B. die Informationen der SNOP-Eingänge

M4000 APPENDICITIS (T6600)
 M1520 APPENDECTOMIE (T6600)

als Summe der semantischen Informationen ihrer Segmente generiert und daher aus der SNOP gestrichen werden. Dies führt zu einer Reduktion des Umfangs der SNOP um etwa 20 %. Nach dieser Reduktion enthält jeder Eingang in SNOP nur echte fixed phrases.

10.3 Wortsegmentierung als Basis eines semantischen Regelsystems

PACAK [25] hat mit einem weniger umfangreichen Segment-Lexikon Anordnungsmuster von Segmenten in zusammengesetzten Wörtern anhand der semantischen Struktur der SNOP untersucht. Jedes Segment trägt eine Information T bzw. M bzw.

E bzw. F - je nachdem, welcher der vier semantischen Kategorien der SNOP dieses Segment zuzuordnen ist. Zusätzlich ist vermerkt, ob dieses Segment ein Eingang in die SNOP ist oder nicht. So bedeutet "TS", daß dieses Segment zur Kategorie TOPOGRAPHY gehört und ein Eingang in die SNOP ist (z.B. GLOMERULO) und "EN", daß dieses Segment zur Kategorie ETIOLOGY gehört und kein Eingang in die SNOP ist (z.B. TOXIN). PACAK fand so folgende Liste von Anordnungsmustern:

Semantische Kategorien der Segmente	Wortform	Semantische Kategorie der Wortform
TN TS	MYO, CARDIUM	T
TS MS	BRONCH, ITIS	M
TS ES	MENINGO, COCCUS	E
TS FS	VASO, SPASM	F
MS TS	VARICO, BLEPHARON	M
ES FN	AMEB, IASIS	F
ES EN	BACTERIO, TOXIN	E
FS FS	ALGO, SPASM	F
TS TS MS	GLOMERULO, NEPHR, ITIS	M
FN TS FN	HYPHER, TRICH, OSIS	M
TS EN FN	THYRO, TOXIC, OSIS	F
FN TS MS	MENO, METRO, RRHAGIA	F
TS TS TS TN MS MS	PERINEO, COLPO, RECTO, MY, OM, ECTOMY	M

Ein möglichst umfangreicher Regelsatz dieser Art ist für eine automatische Textanalyse von großem Wert, da mit seiner Hilfe nicht nur das Referenz-Lexikon verkleinert werden kann, sondern auch Wörter oder Phrasen verarbeitet werden können, die im Referenz-Lexikon nicht enthalten sind. Der zweite Effekt ist besonders wichtig, weil Wortschöpfungen in der lebenden Sprache häufiger sind als die Änderung semantischer Inhalte von Segmenten.

Die Analyse von Anordnungsmustern dient vor allem auch bei polysemantischen Segmenten zu einer eindeutigen Klassifikation. So wurde nachgewiesen [13], daß Wörter mit dem Suffix -IASIS meist zur Kategorie MORPHOLOGIE gehören, wenn diese Endung auf ein T-Segment folgt (z.B. ... LITHIASIS), Aus-

nahme MYIASIS, andernfalls gehört das Wort zur Kategorie FUNKTION (z.B. AMOEBIASIS, ASCARIASIS,...).

Wir untersuchen zur Zeit diese Gesetzmäßigkeiten. Die Ergebnisse werden gesondert publiziert werden.

11. Projektierte Änderungen des Systems

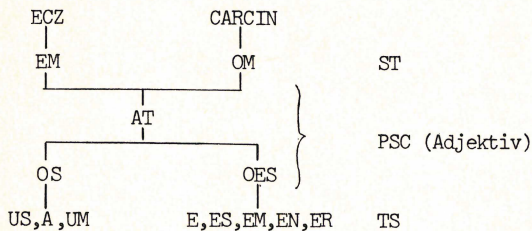
Eine genauere Analyse der Suffixe zeigt, daß diese einen im allgemeinen dreiteiligen Aufbau besitzen:

- Suffix-Stamm (ST),
- Träger der Wortkategorie (PSC),
- terminales Suffix (TS).

Das terminale Suffix kann eine Flexionsendung sein, wenn es am Ende eines Wortes steht, oder es kann ein Verbindungsstring sein, wenn danach noch eine weitere Root folgt. Die drei Teile eines Suffix müssen nicht alle vorhanden bzw. verschieden sein:



Teile eines Suffix können vorangehende Teile in ihrer Bedeutung modifizieren:



In vielen Fällen besteht ein Suffix zwischen zwei Roots nur aus einem Verbindungszeichen

LARYNG...O...TRACHE...O...BRONCH...ITIS .

Der Zweck dieser Aufspaltung der Suffixe ist mehrfach.

Sieht man von sprachspezifischen Ausnahmen ab

dann gilt (siehe Bild 2 bis 6):

- (1) PSC determiniert sowohl das Verbindungszeichen als auch die möglichen Flexionsendungen.
- (2) Bei vielen Substantiv-Adjektiv-Transformationen genügt allein der Austausch des PSC.
- (3) Die meisten Roots stammen aus der lateinischen bzw. griechischen Sprache. Viele sind verenglischt bzw. verdeutsch worden. In diesen Fällen ist zur Übersetzung aus einer Sprache in die andere meist nur der Austausch sprachspezifischer Teile des Suffix notwendig.

Literaturverzeichnis

[1] DIMSDALE, B. (1969): A Thesaurus Processor. IBM Los Angeles Scientific Center, IBM-Form 320-2628

[2] UCLA HOSPITAL - IBM (1970): Natural Language Retrieval System. Pathology Thesaurus. Univ. of Calif., Los Angeles

[3] RÖTTGER, P., WINGERT, F. et al. (1974): Konzeption und Organisation des AGK-Thesaurus. In: Symposium über Klartextanalyse in der Medizin, Siemens AG

[4] WORLD HEALTH ORGANIZATION (Ed.) (1967 und 1969): International Classification of Diseases. Vol. I, II (8th revision, Genf)

[5] IMMICH, H. (1969): Klinischer Diagnoseschlüssel (KDS). Stuttgart: Schattauer

[6] AMERICAN MEDICAL ASSOCIATION (Ed.) (1964/65): Standard Nomenclature of Diseases and Operations. Chicago

[7] NATIONAL INSTITUTES OF HEALTH (1972): Medical Subject Headings, Index Medicus. NIH-Public. No. 72-265, Vol. 13, Part 2, National Library of Medicine, US Gov. Printing Office, Wash. DC

[8] WELLS, A.H. (chm.) (1965): Systematized Nomenclature of Pathology. Committee on Nomenclature and Classification of Diseases, Chicago, Ill. College of American Pathologists

[9] WOLFF-TERROINE, M. (1970): SABIR-C: Automatical System of Bibliography, Information and Research in Cancerology. Bull. Bibliograph. FR 15, 169-176

[10] PRATT, A.W. (1973): Medicine, Computers and Linguistics. Advances in Biomedical Engineering 3, 97-140

[11] KELLER, H.H. (1973): Establishing a German Root System by Computer. Computers and the Humanities 7, 199-208

[12] PACAK, M.G., COUSINEAU, L., WHITE, W. (1972): The Segmentation Approach to Dictionary Construction. Annual Meeting of the Ass. of Canadian Pathologists, Sherbrooke, Canada, NIH-Form 32801

[13] PRATT, A.W., COUSINEAU, L., PACAK, M.G. (im Druck): Medical Language and the Computer.

[14] WONG, R.L., GAYNON, P. (1971): An Automated Parsing Routine for Diagnostic Statements of Surgical Pathology Reports. Meth. Inform. Med. 10, 168-175

[15] BARNETT, G.O., GREENES, R.J., GROSSMAN, J.A. (1969): Computer Processing of Medical Text Information. Meth. Inform. Med. 8, 177-182

[16] O'CONNOR, M.L., SMITH, J. (1973): A Computer-assisted Information System for Anatomic Pathology. Am. Journ. of Clinical Pathology 59, 357-364

[17] CROCKER, D.W. (1972): Automation in Anatomic Pathology. Human Pathology 3, 1-3

[18] PAPLANUS, S.H., STRANDBERG, J.D., SHEPARD, R.H. (1970): A Computer-based Means of Storage and Retrieval of Veterinary Pathology Diagnosis Without Numerical Coding. Laboratory Animal Care 20, 1123-1130

- [19] PAPLANUS, S.H., SHEPARD, R.H., ZVARGULIS, J.E. (1969): A Computer-based System for Autopsy Diagnosis Storage and Retrieval Without Numerical Coding. Laboratory Investigation 20, 139-146
- [20] BECKER, H. (1972): Automatic Processing of Surgical Biopsy Reports. Beitr. Path. 146, 301-312
- [21] SMITH, J., MELTON, J. (1964): Manipulation of Autopsy Diagnoses by Computer Technique. Journ. of the Am. Medical Ass. 188, 958-962
- [22] SMITH, J., MELTON, J. (1963): Automated Retrieval of Autopsy Diagnoses by Computer Technique. Meth. Inform. Med. 2, 85-90
- [23] RÖTTGER, P. (1970): Die Klartextanalyse pathologisch-anatomischer Befundberichte durch Elektronenrechner. Verh. Dtsch. Ges. Path. 54, 582-588
- [24] RÖTTGER, P., REUL, H., SUNKEL, H., KLEIN, I. (1970): Neue Auswertungsmöglichkeiten pathologisch-anatomischer Befundberichte. Klartextanalyse durch Elektronenrechner. Meth. Inform. Med. 9, 35-44
- [25] PACAK, M.G. (im Druck): Automatic Morphosemantic Analysis of Compound Word Forms in Medical Language.

Anschrift des Verfassers:

Prof. Dr. F. Wingert
 Institut für Medizinische
 Informatik und Biomathematik
 der Universität
44 Münster (Westf.)
 Hüfferstraße 75

Nachrichten und Berichte

DVM - Bericht 1

"Das Demonstrations-Datenverarbeitungs-Projekt für das Allgemeine Krankenhaus (DEPAK)"

Die Gesellschaft für Strahlen- und Umweltforschung mbH (GSF) gibt im Rahmen ihrer Projektträgerschaft für das Projekt Datenverarbeitung in der Medizin (DVM) den DVM-Bericht 1 "Das Demonstrations-Datenverarbeitungs-Projekt für das Allgemeine Krankenhaus (DEPAK)" heraus.

Mit dem DVM-Bericht 1 wird ein Überblick über das Konzept und die fortschreitende Verwirklichung eines Vorhabens gegeben, dessen Ziel die praxisnahe Entwicklung eines EDV-Modellsystems für das Allgemeine Krankenhaus mittlerer Größe ist. Orientiert an den realen Möglichkeiten und Grenzen des EDV-Einsatzes im klinischen Routinebetrieb soll diese Entwicklung, welche von der Integration der medizinisch-pflegerischen und administrativen Bereiche ausgeht, mittelfristig zu konkreten, verwertbaren Ergebnissen führen und auf Bundesebene ohne wesentliche Änderungen auf eine Vielzahl von Krankenhäusern übertragbar sein. Zugleich soll damit für Krankenhäuser der ersten Versorgungsstufe eine ausbaufähige Basis für weiterreichende, längerfristige Zielsetzungen im öffentlichen Gesundheitswesen geschaffen werden.

Das Vorhaben der DEPAK wird mit Förderungsmitteln aus dem 2. DV-Programm der Bundesregierung am Stadt- und Kreis-krankenhaus Kulmbach durchgeführt.

In der Reihe der DVM-Berichte, deren erstes Heft jetzt vorliegt, sollen Unterlagen zur Verfügung gestellt werden,

die einer raschen und breiten Anwendung von Forschungsergebnissen aus den über die GSF geförderten Vorhaben dienen.

Im Rahmen derselben Förderungsmaßnahme erscheint auch eine Dokumentationsreihe über das Gemeinschaftsvorhaben von Bund und Ländern "Einheitliche maschinelle Buchhaltung und Betriebsabrechnung in Krankenhäusern", das vom Statistischen Landesamt Rheinland-Pfalz in Bad Ems durchgeführt wird. Bisher liegen vor: Heft 1: Zielsetzung und Organisation des Projekts, Heft 2: Istanalyse, Heft 3: Kontenrahmen mit Musterkontenplan und Erläuterungen und Heft 4: Feinkonzept Finanzbuchhaltung in Krankenhäusern.

Anfragen an: Gesellschaft für Strahlen- und Umweltforschung mbH
 Projekt Datenverarbeitung in der Medizin
8000 M ü n c h e n 81
 Arabellastraße 4/II

C O M M O N ,

eine europäische Vereinigung von IBM-Benutzern kleiner und mittlerer Datenverarbeitungsanlagen.

Bereits auf der Jahrestagung in Berlin, aber in noch viel stärkerem Maße ist man sich 1974 auf der Common-Jahrestagung in Rom darüber klar geworden, daß die ständige Weiterentwicklung des Computers und der Software mehr denn je die Zusammenarbeit der Benutzer von Computern erfordert.

Die ersten Ansätze, Informations- und Erfahrungsaustausch auf internationaler Ebene zu betreiben, kamen 1960 aus

einer Gruppe von 1130-Benutzern und führten so zur Gründung der in Genf registrierten COMMON-Vereinigung. Sie vertritt ausschließlich die Interessen von IBM-Benutzern und ist mit COMMON North America liiert.

Betreut werden in der Hauptsache Benutzer der Anlagenreihe 1130, 1800, System /3, System /7 und aller /360 und /370-Anlagen, mit Ausnahme jener, die das Betriebssystem OS einsetzen.

Um die Vielfältigkeit der derzeitigen Anwendungen einigermaßen in Griff zu bekommen, wurden folgende sieben Themenkreise geschaffen:

1. Hardware
2. Betriebssysteme
3. Datenbanksysteme
4. Teleprocessing
5. Software
6. Datensicherung
7. Dokumentation

Die Koordinierung der einzelnen Fachrichtungen ist ein weiteres Ziel von COMMON, durch das der Benutzer die Möglichkeit erhält, auf fachorientierten Tagungen mit interessengleichen Gesprächspartnern zusammenzuarbeiten.

Zur Zeit beschäftigt sich COMMON bei den regionalen Treffen hauptsächlich mit folgenden aktuellen Themen:

1. Job-Accounting allgemein und unter POWER.
2. Sicherheit der Programme und die Möglichkeiten eines Verwaltungssystemes.
3. Dokumentation in der EDV.

Alle Nichtmitglieder, die sich für COMMON interessieren, sind zu einer ersten Kontaktaufnahme herzlich eingeladen, an einer COMMON-Tagung teilzunehmen. Die Mitgliedschaft kann von jedem erworben werden, der eine IBM-Datenverarbeitungsanlage betreibt. Der Beitrag beträgt derzeit 260 SFR p.a. und dient zur Finanzierung der Verwaltungsarbeiten und des Literaturdrucks.

Interessenten wenden sich bitte an den Projektleiter für Deutschland, der unter folgender Adresse zu erreichen ist:

Dietmar Schrenk
7023 Echterdingen
Bergstraße 38 Tel.: 0711/795562

Biometrie und Umweltforschung

Die Region Österreich/Schweiz der Biometrischen Gesellschaft führt in der Zeit vom 29. September bis zum 3. Oktober 1975 in Chur ein Seminar mit dem Thema
Biometrie und Umweltforschung
durch.

Vorläufiger Themen- und Zeitplan:

29. 9.75 Biometrische Methoden in der Oekologie;
Dynamische Modelle

30. 9.75 Stichprobentheorie, Erhebungen in Theorie und Praxis
1.10.75 Kurzvorträge und Ausflug
2.10.75 Biometrische Methoden bei präsentativ medizinischen Untersuchungen; Planung und Auswertung von Markierungsversuchen an freilebenden Tieren
3.10.75 Kurzvorträge; Biometrische Bearbeitung kartographisch dargestellter Ergebnisse von Erhebungen

Programmgestaltung:

Prof. Dr. S. Riedwyl, Sidlerstr. 5,
CH - 3012 Bern

Organisation und Anmeldungen:

Dr. F.H. Schwarzenbach, Giacomettistraße 96,
CH - 7000 Chur

ISIS Software Report

Im Augenblick liegt die 9. Ausgabe dieses Kataloges vor. Ca. 750 Programme werden hier angeboten. Dabei sind leider nur wenige technisch-wissenschaftliche Programme. Für alle Anwender von Interesse sind aber vor allem die sogenannten System-Programme. Die zusammengestellten Aufgaben sind übersichtlich und ausreichend.

Der ISIS Software Report erscheint zweimal jährlich und liefert einen umfassenden Überblick über das Software-Angebot im deutschsprachigen Wirtschaftsraum.

Anfragen an: Infratest Informations Services
8000 München 21
Landsberger Straße 338

Rheinische Friedrich-Wilhelms-Universität Bonn

Die Landwirtschaftliche Fakultät sucht für den Aufbau einer Arbeitsgruppe Mathematik-Statistik:

1 Wissenschaftlichen Rat und Professor (H 3) 1 Studienprofessor (A 15)

Die Stelleninhaber sollen in Zusammenarbeit mit einem weiteren wissenschaftlichen Mitarbeiter die Lehrverpflichtungen „Mathematik und Statistik“ im Grund- und Hauptstudium für Agrar- und Ernährungswissenschaftler wahrnehmen und bei der Forschungsbetreuung die Bereiche Biometrie (pflanzliche und tierische Produktion) sowie Ökonometrie bearbeiten.

Bewerbungen mit den üblichen Unterlagen bis zum 15. März 1975 erbeten an den
**Dekan der Landwirtschaftlichen Fakultät,
53 Bonn 1, Meckenheimer Allee 174.**