

Die Verarbeitung unformatierter Daten am Beispiel der Beschreibung von wissenschaftlich-technischen Dokumenten

Harald Löhr

Zöllmener Straße 5
01157 Dresden
loehr@dd.sda.de

Abstract: Wissenschaftlich-technische Dokumente wie Zeitschriftenartikel, Patente und Bücher waren schon lange Gegenstand von Informationsmitteln wie Kataloge und kumulierte Inhaltsverzeichnisse. Im 20. Jahrhundert wuchs die Flut an Dokumenten und damit die Aufgabe, effektiver über ihren Inhalt zu berichten. Nach der weltweiten Anwendung der Dezimalklassifikation als Ordnungsmittel für Kurzreferate in Karteikartenform und von meist fachgebietsbezogenen Referateblättern stand von den 1960-er Jahren an die rechnergestützte Informationsverarbeitung für die Aufgaben der Information und Dokumentation zur Verfügung.

Bereits mit Magnetbändern konnten Dokumentennachweise effektiv gespeichert und rechnergestützte Informationsmittel preisgünstig und schnell bereitgestellt werden. Für die Rechanlage R300 wurde ein Typenfeinprojekt für die Information und Dokumentation bereitgestellt und DDR-weit angewendet. Das AIDOS war die Produktreihe für die Folge-Rechnerfamilien.

Am Beispiel von Rechnerpraktika auf der Basis dieser Softwaresysteme wurden bereits seit den 1970-er Jahren Verfahren zur effektiven Verarbeitung unformatierter Daten in die Informatik-Ausbildung einbezogen.

1 Speicher für wissenschaftlich-technische Dokumente - historischer Abriss -

1.1 Primär- und Sekundärdokumente. DK-Zahlen

Als **Primärdokument** wollen wir hier ein Dokument verstehen, in dem über einen (neuen) Sachverhalt, Zusammenhang o. ä. berichtet wird.

Ein **Sekundärdokument** ist ein Dokument, das über ein anderes Dokument berichtet, darüber hinaus aber keinen neuen Sachverhalt vermittelt.

Unter dem Oberbegriff wissenschaftlich-technische Primärdokumente wurden zuerst schriftlich fixierte Dokumente verstanden, die sachlich-fachlichen Inhalt haben und niedergelegt wurden, um die Mitwelt und die Nachwelt über die Gedanken des Verfassers zu informieren und ggf. auch einen Urheberanspruch zu begründen. Ohne Anspruch auf Vollständigkeit sind das Zeitschriftenartikel, Bücher, Patentschriften, Prospekte, Kurzmitteilungen, Gesetze und Standards, aber auch Gerichtsurteile und Festlegungen aller Art.

Dokumenten größeren Umfangs, wie z. B. Fachbücher und nicht alphabetisch geordnete Nachschlagewerke, wurde schon seit langem eine spezielle Ergänzung hinzugefügt, mit der man zu Einzelinformationen im Buch zugreifen kann: das Register/der Index. Diese 'Schlagwortliste' findet man (wenn vorhanden) meist am Ende des Werkes. In diesem Verzeichnis von Fachbegriffen sind zu jedem enthaltenen Stichwort die Nummern der Seiten genannt, auf denen der Fachbegriff (signifikant) auftritt. Der Gestalter des Registers legte fest, welche Inhaltskomponenten er für registerwürdig ansah. Gegenwärtig unterstützen Textverarbeitungssysteme das Registermachen technisch - die Entscheidung über die Aufnahme eines Wortes fällt aber in jedem Einzelfall der Mensch! Automatisierte Verfahren, die z. B. die Häufigkeit eines Fachbegriffs als Kriterium für die Aufnahme ins Register verwenden, verschieben das Problem nur auf eine andere Ebene, denn die Programme, die die Entscheidung treffen, werden auch von Menschen entworfen.

Mit dem Register wurde die zweistufige Recherche geboren. Einstufig soll hier die sequentielle Suche nach einer relevanten Textstelle im Gesamttext (Seite für Seite) bedeuten. Zusätzlich bietet ein Register die Möglichkeit, zu einem Fachbegriff relevante Seitennummern zu finden (Stufe 1) und dann zum gesuchten Kontext auf den in der Liste vermerkten Seiten zuzugreifen (Stufe 2). Die Register gestatten nur den Zugriff auf relevante Textseiten innerhalb eines Werks.

Seit etwa 200 Jahren nahm die Anzahl der wissenschaftlichen Zeitschriften ständig zu und damit die Schwierigkeit, zu den Primärdokumenten meist geringeren Umfangs zuzugreifen, die in den Zeitschriftenbänden/-jahrgängen abgelegt waren.

Die Hauptaufgabe der Weiterarbeit mit den gespeicherten **Primärdokumenten** besteht darin, die zu einer bestimmten Fragestellung relevanten Dokumente wieder zu finden. In den ersten Jahrhunderten des wissenschaftlichen Austauschs waren dazu die persönlich formulierten Briefe zu analysieren und die entsprechenden Zeitschriften gezielt zu durchsuchen. Welche Verluste bei dieser Rechercheart entstehen, ist u. a. seit der Nichtveröffentlichung vieler Erkenntnisse von Carl Friedrich Gauss selbst oder seit der (teilweisen) Vernichtung der Aufzeichnungen von Bernhard Riemann bekannt.

Aus diesen Gründen spielte die Veröffentlichung der Arbeitsergebnisse in wissenschaftlich-technischen Zeitschriften und Jahrbüchern, also in Periodika, eine hervorragende Rolle. Schnell wurde das so gespeicherte Wissen unübersehbar: Es musste aufbereitet werden. Deshalb entstanden verschiedene Formen von **Sekundärliteratur**, deren Inhalt nicht in originär neuem Wissen besteht, sondern darin, auf Originalquellen hinzuweisen. Der wichtigste Typ eines Sekundärdokuments sind durch Fachleute verfasste Kurzfassungen, so genannte Kurzreferate von 10 - 20 Zeilen á ca. 40 Zeichen. Sie sind außer in gebundener Form gut auf Karteikarten im A6-Format unterzubringen. Damit war eine neue Aufgabe entstanden: Die Sekundärdokumente mussten für eine erfolgreiche Suche sinnvoll gespeichert werden.

In der Mitte des 20. Jahrhunderts waren weltweit zwei Hauptformen von Speichern für Kurzreferate verbreitet. Es waren nach der international standardisierten universellen Dezimalklassifikation (UDK) geordnete **Karteikartenspeicher** und **Referatzeitschriften** mit einer jeweils eigenständigen fachlichen Gliederung des Fachgebiets.

Wichtigstes Kennzeichen einer **Referatekarte** war die DK-Zahl, die das Referat einem Fachgebiet zuordnet. Bei mehreren Fachgebieten mussten Duplikate der Karte angefertigt werden, damit das Sekundärdokument unter allen relevanten Fachgebieten wieder gefunden wird.

Das Prinzip einer **Dezimalklassifikation** besteht darin, dass das gesamte Wissen der Menschheit in Gruppen unterteilt wird, deren jede durch eine Dezimalziffer von 1 bis 9 gekennzeichnet wird. Jede Gruppe wird nun bei Bedarf in 9 Untergruppen unterteilt. Die Unterteilung kann beliebig oft fortgesetzt werden, bis die notwendige Feinheit der Fachgebiete erreicht ist. Nach jeweils 3 Ziffern wird als Gliederungshilfsmittel ein Punkt eingefügt. DK-Zahlen von 15 und mehr Ziffern sind keine Seltenheit, weil manches Wissensgebiet bei der Erst-Definition des Systems der DK-Zahlen in seinem künftigen Umfang nicht richtig eingeschätzt werden konnte. Das trat besonders bei den Naturwissenschaften und in der Technik auf. Im praktischen Einsatz orientierte man darauf, hinter der Leitkarte für eine DK-Zahl zwischen 5 und 30 Referatekarten stehen zu haben, um leicht zu den Einzelkarten zugreifen zu können. Wurde diese Richtzahl bleibend überschritten, dann musste die nächste Untergliederungsebene eröffnet werden. Automatisch ergab sich dadurch die notwendige Feingliederung der DK-Zahlen in einer Informationsstelle. Einige Beispiele für DK-Zahlen sind in Tabelle 1 zusammengestellt. Man erkennt gut die unterschiedliche Größe der durch die ausgewählten vierstelligen DK-Zahlen repräsentierten Wissensgebiete.

Tabelle 1 Beispiele für DK-Zahlen nach [De05]

DK-Zahl	Bedeutung
621	Angewandte Physik
621.3	Elektro-, Magnettechnik, technische Optik, Kommunikations-, Computertechnik, Elektronik, Lichttechnik
621.381	Elektronik
621.381.528.7	Thyristoren
621.39	Computertechnik
621.4	Antriebsmotoren und Wärmetechnik
621.6	Gebläse, Lüfter, Pumpen

Parallel dazu, dass die Abteilungen der Speicher für Sekundärinformationen immer größer und damit schlechter überschaubar wurden, ertönte zunehmend die Forderung danach, nicht nur nach Fachgebieten, sondern auch außerhalb von Fachbüchern nach Inhaltskomponenten der Dokumente suchen zu können.

Die Abhilfe stellten vorerst fachbezogene Bibliographien dar, in denen für vorher bekannten Nutzerbedarf relevante Auszüge aus Referatespeichern zusammengestellt wurden. Der Begriff Bibliographie wurde hier über seinen Ursprung im Bibliothekswesen hinaus erweitert.

Als ergänzender Schritt wurde dann das Prinzip der Register für eine neue Speicherform verwendet und erweitert. So wurde die Forderung an die Recherche als Suche nach für eine Aufgabe relevanten Dokumenten über ihre Inhaltskomponenten erfüllt. Dabei entstand die **Sichtlochkartei**.

Sie bildete die erste Möglichkeit zur freien Recherche nach relevanten Dokumenten - nicht nur nach Textstellen in *einem* Gesamtdokument - entsprechend der gekennzeichneten Inhaltskomponenten.

1.2 Sichtlochkartei

Diese Speicherungsform für Informationen war nicht nur originell, sondern begründete auch eine wissenschaftlich fundierte Recherchestrategie. In diesem Sinn ist die Sichtlochkartei eine direkte Vorstufe der rechnergestützten zweistufigen Informationsrecherche über Inhaltskomponenten. Das sah anfangs allerdings niemand so deutlich.

Das **Prinzip der Sichtlochkartei** ist denkbar einfach. Zu jedem Fachbegriff (Deskriptor) wird eine Sichtlochkarte angelegt, in die an den entsprechenden Positionen der Nummern der Primärdokumente, für die der Deskriptor vergeben wurde, ein Loch gestanzt wird. Sucht man nun alle Dokumente, für die der Deskriptor 'Recherche' als bedeutsam erklärt wurde, so greift man die Sichtlochkarte zu diesem Deskriptor 'Recherche' und schaut gegen eine Lichtquelle. Alle Löcher, die hell erscheinen, weisen auf die relevanten Dokumentennummern hin - die Rechercheaufgabe ist gelöst. (s. dazu [Gü71])

1.3 Logische Grundoperationen in Rechercheanfragen

Mit solchen Sichtlochkarten lassen sich die Grundoperationen der Aussagenlogik auf einfache Weise für Rechercheanfragen (Anfragen) realisieren.

Die **Konjunktion**, das logische UND mehrerer Deskriptoren, ergibt sich durch das Übereinanderlegen der zutreffenden Sichtlochkarten. Das Licht scheint nur für die Nummern der Dokumente hindurch, für die **sowohl** der eine Deskriptor **als auch** der zweite Deskriptor **als auch** der dritte ... vergeben wurde.

Die **Disjunktion**, das logische ODER, besteht in der **Vereinigung** der Dokumentennummern, die auf den Sichtlochkarten für den ersten Deskriptor mit denen für den zweiten Deskriptor mit denen für den dritten ... abgelocht wurden. (Logisches ODER bedeutet hier: Relevant sind die Dokumente, die mit dem ersten Deskriptor **oder** mit dem zweiten **oder mit beiden** usw. gekennzeichnet worden waren.)

Beide Operationen ergeben die Liste der Nummern der relevanten Dokumente. Die **Negation** sollte in Recherchefragen tunlichst vermieden werden.

2 Nutzung eines Programmsystems für die Speicherung von Sekundärdokumenten (R300-Projekt)

Am Beispiel des R300-Typenprojekts für die Information und Dokumentation soll der Anfang der 1970-er Jahre in der DDR erreichte Stand der Arbeit mit wissenschaftlich-technischen Dokumenten dargestellt werden, die in Form von weitgehend unformatierten Daten programmgerecht aufbereitet und dann in R300-Dateien gespeichert und recherchiert werden konnten. Derartige Programmsysteme können heute noch genutzt werden. Allerdings sind die zu lösenden Aufgaben infolge des Fortschritts sowohl der Speicherverfahren als auch der Recherchestrategien auf moderneren effektiveren Wegen zu lösen. Ob dabei immer die gewünschte Treffergenauigkeit erreicht wird, steht allerdings in Frage. Hier muss ein Kompromiss zwischen Wirtschaftlichkeit einschließlich Zeitersparnis und Genauigkeit gefunden werden.

2.1 Unformatierte Daten. R300 als rechentechnische Basis

Die bis Anfang der 1960-er Jahre verarbeiteten Daten waren entweder Zahlen (für reine Berechnungsaufgaben) oder zumindest fest formatierte Datensätze (für ökonomische Aufgaben und Berechnungen). Derartige **formatierte Daten** lassen sich relativ einfach in Dateien schreiben und wieder auslesen. Auch Änderungen innerhalb der Datensätze bzw. das Löschen und Hinzufügen ganzer Sätze war mit den verfügbaren Standardverfahren leicht möglich. Die wichtigsten davon arbeiteten mit den Schlüsselwörtern der Datensätze.

Bei der Abbildung wissenschaftlich-technischer Informationen auf Datenstrukturen entstehen Datenelemente und Datensätze, deren Länge in sehr weiten Grenzen variieren können und die sich deshalb nur mit großer Speicherplatzvergeudung auf Strukturen mit festem Format des bisherigen Verständnisses abbilden lassen oder nur jeweils eine vordefinierte begrenzte Anzahl von Werten in einem Datenelement zulassen. Diese Daten waren mit den bis in die PL/1-Zeit hinein gebräuchlichen Verfahren nicht speicherplatzeffektiv zu verarbeiten. Für diese Daten bürgerte sich die Bezeichnung '**unformatierte Daten**' ein. So musste z. B. auch die bisherige Festlegung aufgegeben werden, dass ein Datenelement nur einen (1) Wert enthalten darf. Als Beispiele seien genannt, dass ein Buch häufig nicht nur einen einzigen Verfasser hat oder dass der Inhalt eines Dokuments i. allg. mit mehr als genau einem Deskriptor gekennzeichnet werden muss.

Am Beispiel des Dokumentennachweises, im R300-Projekt [Ro69] Nachweis genannt, wird gezeigt, wie wissenschaftlich-technische Sekundärdokumente speichergünstig und recherchéangepasst auf Datenstrukturen abgebildet werden können (s. Abschnitt 2.2).

Der ersten Digitalrechenanlage der DDR, dem ZRA1 (Zeiss Rechenautomat 1) folgte aus der Produktion des Kombinats Robotron der R300 (Robotron 300), der mit seinen 8 Magnetbandgeräten zur externen Speicherung und Verarbeitung von Massendaten geeignet war. Damit konnten die bei der Verarbeitung von beliebig vielen Datensätzen entstehenden Datenmengen extern gespeichert werden. Das war für formatierte Datensätze im Datenbanksystem Robotron (DBS/R) realisiert.

Für Datensätze mit unformatierten Daten wurde die R300-gestützte Verarbeitung im Typenfeinprojekt für die Information und Dokumentation [Ro69]) angeboten. Im praktischen Betrieb standen für die Nutzerdateien jeweils maximal 6 Magnetbändeinheiten zur Verfügung. Die Daten wurden über 8-Kanal-Lochband eingegeben, das standardmäßig auf dem Organisationsautomaten Optima 528 kodiert wurde. Die Datenausgabe erfolgte über einen 80spaltigen Paralleldrucker.

2.2 Struktur des Dokumentennachweises im R300-Projekt

An der Gestaltung des Nachweises und seinem Inhalt lässt sich der damalige Wissensstand auf dem Gebiet der rechnerunterstützten Verarbeitung unformatierter Daten charakterisieren.

Die zitierten Quellen außer [Ro69] wurden aus der praktischen Arbeit mit wissenschaftlich-technischen Dokumenten in einer betrieblichen Informationsstelle abgeleitet und sind als Arbeitsberichte anzusehen, z. B. [Gü71], [Gü72], [Lö72] und [Rö74]. An zahlreichen Stellen war die dargestellte theoretische Fundierung erforderlich, um zu einer stabilen und erfolgreichen Arbeit zu gelangen. Der Nachweis ist im R300-Projekt in 10 Eingabesätze, Kategorien genannt, gegliedert (s. Bild1). Intern ist der Gesamtnachweis ein Datensatz. Die zehn Kategorien sind dann Datenelemente mit Werten beliebiger Länge und z. T. auswertbarer Struktur.

0 ZAC01407

1 F.3.4.; F.3.5.;

2 Ferrit, Eigenschaft; hoechst, Frequenz; Teil, magnetisieren;

Berechnung; Theorie; Magnet, Eigenschaft; Permeabilitaet; Literatur;

4 Schloemann, Ernst;

5 =Microwave behavior of partially magnetized ferrites.

(=Verhalten teilmagnetisierter Ferrite bei Mikrowellenfrequenzen).

6 J. Appl. Phys., Lancaster 41 (1970) 1, s. 204-214, 10 Abb., 20 Lit.,

Bde 01

8 D221: Reg.-Nr. M 0556/08

9 E

Bild 1 Struktur des Dokumentennachweises im R300-Projekt des Kombinats Schaltelektronik (s. [Gü71])

Jede Kategorie beginnt mit ihrem Namen, dargestellt durch eine der Ziffern von 0 bis 9, gefolgt von einem Leerzeichen. Folgezeilen in einer Kategorie werden an zwei führenden Leerzeichen erkannt. Für jeden Einzelwert, der recherchierbar sein soll, ist ein Trennzeichenformat festgelegt: Das Endkennzeichen für einen solchen Einzelwert (Kategorie1:Notation, Kategorie2: Deskriptor(gruppe)) ist das Semikolon. In den anderen Kategorien kann nur rein sequentiell im gesamten Eintrag gesucht werden.

In den Kategorien 1 bis 8 sind Folgezeilen zugelassen. Die maximale externe Zeilenlänge beträgt 60 Zeichen. Damit wird die im Format A4 mögliche Zeilenlänge eingehalten, die für die Kontrollmitschrift beim Ablochen der Eingabedaten und für die rechnerunterstützt erzeugten Nutzer-Informationsmittel von Bedeutung ist.

Kategorie 0 - externe Dokumentennummer = Nachweisnummer. Diese Kategorie ist als einzige auf genau 8 Zeichen formatiert. Die laufende Nummer wird durch Kodezeichen für die Dokumentenart und den Vertraulichkeitsgrad ergänzt.

Kategorie 1 - Notationen. Nach dem gleichen Prinzip wie bei der UDK ist bei der Einsatzvorbereitung für das R300-Projekt eine für den konkreten Anwenderbereich spezifische Klassifikation zu erarbeiten. Sie heißt im R300-Projekt **Systematik der Notationen**. Diese Klassifikation muss vor dem Beginn der Nachweisespeicherung eingegeben werden. Jede Notation beginnt mit einem Buchstaben, gefolgt von Ziffern oder Ziffern-paaren, die jeweils vom Vorgänger durch einen Punkt zu trennen sind. Im Dokumenten-nachweis können mehrere Notationen vergeben werden, wenn das Dokument mehr als ein Fachgebiet betrifft. Bei der Eingabe des Nachweises werden die Werte in Kategorie 1 mit den Notationen in der Systematik verglichen. Jeder Fehler wird protokolliert, und der Nachweis wird abgewiesen. Die fehlerhaften Nachweise erscheinen mit allen er-kannten Fehlern im Eingabeprotokoll. Sie können extern korrigiert und dann erneut eingegeben werden.

Kategorie 2 - Deskriptoren. Der **Thesaurus**, die Liste der zum Indexieren zugelasse-nen Deskriptoren, ist ebenfalls im Rahmen der Einsatzvorbereitung zu erarbeiten und anhand praktischer Erfahrungen weiter zu entwickeln [Lö72]. Da die Deskriptoren zur Kennzeichnung der wichtigen Inhaltskomponenten des Primärdokuments bestimmt sind, sollten in den Thesaurus keine Bezeichnungen der Fach- oder Wissensgebiete aufge-nommen werden. Die Gebiete werden durch die Notationen kodiert! Eine wichtige Funk-tion eines Thesaurus in diesem Sinne ist die Zusammenführung synonyme bzw. quasi-synonymer Bezeichnungen, die als Nichtdeskriptoren im Thesaurus verzeichnet sind, auf eine aus dieser Menge - den vereinbarten Deskriptor. Die Verwendung einer nicht als Deskriptor definierten Bezeichnung führte im R300-Projekt ebenfalls zum Abweisen des Nachweises.

In vielen Fällen ist es sinnvoll, nicht zu viele **Komposita** als Deskriptoren zu definieren, zumal die Komposita in Zusammenschreibung ohnehin eine Eigenheit der deutschen Sprache darstellen. Deshalb bietet fast jeder Thesaurus die Möglichkeit, bis zu drei Wör-ter zu verbinden, was im R300-Projekt durch ein verbindendes Komma kodiert wird. Die Reihenfolge *Grundwort, Bestimmungswort1, Bestimmungswort2* sollte als allgemeine Regel im Thesaurus verwendet werden.

Am Nachweis Nr. 01407 (Bild 2) erkennt man gut, dass am Anfang der Arbeit mit dem R300-Projekt im Arbeitsthesaurus die eben diskutierte normierte Aufgliederung der Komposita noch nicht berücksichtigt war. Das war Gegenstand einer grundsätzlichen Überarbeitung des Thesaurus, die bald folgte ([Lö72], s. a. [Lö75a]).

Kategorie 4 - Verfasser. Hier werden die Verfasser in der Reihenfolge *Nachname, Vorname*; angegeben. Die im Bibliothekswesen geltenden Richtlinien sind bei der Erfassung der Verfasseramen verbindlich. Unstimmigkeiten bei der Schreibweise von Verfassern treten unabhängig von diesen Richtlinien dann auf, wenn unterschiedliche Schriftsysteme Anwendung finden (s. dazu z. B. [Lö71] für die unterschiedliche resultie-rende Schreibweise japanischer Autorennamen).

Kategorie 5 - Titel des Primärdokuments. Beim Kodieren der Angaben für diese Kategorie sind ebenfalls die Festlegungen des Bibliothekswesens einzuhalten. Ein aus einer Fremdsprache übersetzter Titel ist in Klammern anzugeben.

Kategorie 6 - Quelle des Primärdokuments. Die Quellenangaben folgen gleichermaßen den im Bibliothekswesen geltenden Richtlinien.

Kategorie 8 - Standort des Primärdokuments. Hier wird der Standort des Dokuments in der für die Nutzer zuständigen wissenschaftlich-technischen Bibliothek, gekennzeichnet durch ihr Sigel, angegeben, bei selbständigen Dokumenten ergänzt durch seine Signatur.

Kategorie 9 - Sprache. Für die Kennzeichnung der Sprache des Primärdokuments sind eigene Kurzzeichen festzulegen. Die Angabe mehrerer Sprachen ist zulässig.

2.3 Aus dem Speicher abgeleitete Informationsmittel: Bibliographie, SIV und Recherche

Die konventionelle Informationsversorgung der Nutzer einer Informationsstelle erfolgt über Breiteninformationsmittel wie Neuerwerbungslisten und die speziellen anforderungsbezogenen Recherchen. Diese beiden Leistungen werden auch vom R300-Projekt erbracht: Bibliographie und Recherche. Zusätzlich besteht die Möglichkeit, im Rahmen der *Selektiven Informationsverbreitung* anhand eines gespeicherten Nutzerprofils periodisch im gleichen Rhythmus wie die Bibliographie, also nach jeder Neueinspeicherung von Nachweisen oder in anderem Abstand, die Nutzer mit genau für sie selektierten Informationen zu versorgen. Ein Nutzerprofil in diesem Sinn ist eine relativ stabile und dauerhaft gespeicherte Recherchefrage zur ständigen Abarbeitung. Tabelle 2 zeigt die Informationsmittel in der Übersicht.

In einer Speicheraanfrage (Recherchefrage oder Nutzerprofil) lassen sich die Suchelemente beliebig durch Konjunktion, Disjunktion und Negation (Vorsicht! S. Abschnitt 1.3) kombinieren. Die sprachliche Formulierung einer solchen Aufgabe wird leicht ungenau: das 'Und' kann als *'sowohl A als auch B'* (konjunktiv), aber auch als *'A und auch B'* (disjunktiv) gedeutet werden. Formuliert man das sprachlich genau, dann wird die Anfrage sprachlich kompliziert - und außerdem: Wie spricht man die Klammern? Die Darstellungsmöglichkeit als Schaltnetzwerk bietet den Ausweg: Die Anfrage wird exakt und übersichtlich (s. dazu Bild 3 und Bild 4).

Tabelle 2 Informationsmittel im R300-Projekt

Art \ Häufigkeit	ohne Anfrage: Breiteninformations- mittel	auf Anfrage: nutzerbezogene Informationsmittel	Anfrage formuliert als
periodisch	Bibliographie	selektive Informations- verbreitung	Nutzerprofil (ge- speichert und leicht zu editieren)
einmalig	nicht relevant	Recherche	Recherchefrage

Der fachliche Teil einer einfachen Anfrage in einem Nutzerprofil oder in einer Recherchefrage hat im R300-Projekt z. B. folgende Gestalt:

f (1:(F.4.3. ;* H.2.1.;)! + 2:(DDR + Informatik) * (Rechentechnik + Information/Dokumentation)!)?

Durch die Nutzung einer Deskriptorgruppe *Informatik*, *DDR* wird die Anfrage schärfer, weil die Nachweise, in denen die Deskriptoren *Informatik* und *DDR* zwar beide aber einzeln auftreten, nicht angezeigt werden. Da die Konjunktion kommutativ ist, spielt im Falle der Einzeldeskriptoren, der offeneren Anfrage, ihre Reihenfolge oder ob sie in der Kategorie 2 überhaupt benachbart sind, keine Rolle für die Relevanz des Nachweises.

Mit diesem Beispiel wird gezeigt, auf welcher guten theoretischen Grundlage bereits Anfang der 1970-er Jahre im Gebiet Information und Dokumentation, also bei der Arbeit mit unformatierten Daten gearbeitet wurde. Die rechnergestützte Arbeit im Fachgebiet zwang uns dazu, sie theoretisch zu durchdringen, um alle Möglichkeiten, die die neue Technik eröffnete, auch nutzen zu können.

2.4 Zusammenarbeit mit den Nutzern

Das Ziel der gesamten Arbeit mit der Speicherung von Sekundärinformationen besteht in der Befriedigung des Nutzerbedarfs an Informationen für die Forschungsarbeit bzw. die Informationstätigkeit der Nutzerbereiche im weitesten Sinn. Bis Ende der 1960-er Jahre bestand die Recherche in den vorhandenen Speichern im Suchen relevanter Nachweise in gedruckten Bibliographien oder Referatezeitschriften wie z. B. Technisches Zentralblatt oder Реферативный Журнал sowie in der Suche relevanter Karteikarten in diesem Referatespeicher. Das Problem komplizierter Anfragen stellte sich nicht ausdrücklich. Einen Fortschritt brachte die Anwendung des Sichtlochkartenverfahrens mit der Möglichkeit, nach mehreren (und auch verknüpften) Deskriptoren als Merkmal für die Relevanz eines Dokuments zu suchen. Damit war die Frage nach den Kombinationsmöglichkeiten für die Deskriptoren entstanden - und damit nach der Anwendung der Grundoperationen der Aussagenlogik in einer Recherchefrage. Das blieb anfangs noch unerschwellig, doch mit der Nutzung von Rechnerprogrammen für diese Aufgaben wuchs die Forderung nach bestmöglich strukturierten Recherchefragen.

Es gilt noch heute: Der Nutzerbedarf wird erst dann klar formuliert, wenn der Nutzer gezwungen wird, ihn zu formalisieren. Nun ist es wohl kaum nutzerfreundlich, vom Fachmann eines beliebigen Gebiets zu fordern, seine Anfrage in der sehr formalen Schreibweise eines Programmsystems zu formulieren. Eine heute ganz normale Zwischenstufe besteht in der Verwendung einer aufgabengerechten, aber noch nicht programmbezogenen Sprache. Für die Formulierung einer Recherchefrage bietet sich dafür bis heute ein Schaltnetzwerk an - wie oben gezeigt (Bild 3 und Bild 4). Die klar erkennbare logische Struktur der Anfrage hat auch den großen Vorteil, dass sich Fachleute jedweden Gebiets sehr schnell in diese formale Beschreibungsart hineindenken können. Damit stehen sie als gleichberechtigte aktive Partner bereit, um ihren Informationsbedarf möglichst genau zu formulieren, denn eine verbale Anfrage muss normalerweise mehrfach durchdacht werden, ehe man dafür die richtige Recherchefrage herausgearbeitet hat. Das ist bei komplexen Anfragen ein Prozess mit mehreren Zyklen. In der Praxis haben wir bis zu vier Varianten erarbeitet, bis die Anfrage dem Nutzerbedarf wirklich genau entsprach. Am Beispiel des Schaltnetzwerks konnten dem Nutzer auch die logischen Grundoperationen anschaulich erläutert werden - bis hin zur Gefährlichkeit der Negation eines Suchmerkmals in einer Anfrage.

Die genannten Probleme waren neben der theoretischen Bedeutung auch praktisch relevant, wenn man bedenkt, dass Rechercheläufe in einem R300-Nachweisespeicher mehrere Stunden dauern konnten - und die dabei entstehenden Kosten, die der Nutzerbereich tragen musste, sollten sich in einem bestmöglichen Ergebnis niederschlagen!

3 Unformatierte Daten in der Ausbildung

An den Sektionen Informationsverarbeitung der TU Dresden und der IH Dresden wurde ab 1974 in der Spezialisierungsphase je ein Praktikum unter Nutzung des R300-Projekts [Ro69] durchgeführt ([Lö75b]).

In diesen Praktika wurden die Studenten mit den Besonderheiten der Verarbeitung wissenschaftlich-technischer Informationen bekannt gemacht. Dazu gehörte die aktive Arbeit mit einem Arbeitsthesaurus einschließlich der Festlegung neuer Deskriptoren und die Anwendung von Grundsätzen einer systemspezifischen Klassifikation. Außerdem waren aktuelle Primärdokumente auszuwerten und die dafür geeigneten Sekundärdokumente (Nachweise) zu erarbeiten und einzuspeichern. Schließlich mussten aus dem Gesamtspeicher aufgabengerechte Informationsmittel abgeleitet werden.

Die Studenten lernten auch die mit diesen Verfahren verbundenen technologischen Fragen der Verarbeitung dieser unformatierten Daten kennen, so z. B. die Notwendigkeit, die ständig anwachsende Zahl der Nutzerdatensätze extern zu speichern und sicher zu verwalten. Zu diesem Zeitpunkt wurde bereits erkannt, dass die bisher als qualitativ unterschiedlich behandelten Verfahren zur Speicherung und Weiterverarbeitung von Datensätzen mit formatierten und unformatierten Informationen in absehbarer Zeit zusammengeführt werden müssen. Die sich entwickelnde Datenbanktechnik sollte das in Zukunft ermöglichen. Das Praktikum wurde nach der Praxiseinführung des AIDOS (s. a. [Nau08]) auch in der Ausbildung auf dieses System umgestellt.

Literaturverzeichnis

- [De05] Dewey-Dezimalklassifikation DDC22 Band 3 Haupttafeln 600-999. KG Saur München (2005), 1076 S.
- [Gü71] Güldner, R.; Löhr, H.: Aufgaben und Arbeitsweise der Informationsstelle eines Kombinars. Gleitender Übergang von der Steilkartei über die Sichtlochkartei zur Datenverarbeitung. Informatik, Berlin 18(1971) 6, S. 46-49, 4 Abb., 1 Tab.
- [Gü72] Güldner, R.; Löhr, H.: Stand der EDV-Anwendung in der Zentralen Informationsstelle des VEB Kombinat Schaltelektronik. Informatik, Berlin 19(1972) 2, S. 35-37, 2 Abb., 1 Tab., 3 Lit.
- [Lö71] Löhr, H.: Transliteration japanischer Wörter über die kyrillische in die lateinische Schrift. Informatik, Berlin 18(1971) 5, S. 48-52, 1 Abb., 3 Tab., 5 Lit.
- [Lö72] Löhr, H.: Anpassung eines Arbeitsthesaurus an übergeordnete Thesauren. Informatik, Berlin 19(1972) 6, S. 30-34, 1 Abb., 1 Tab., 4 Lit.
- [Lö75a] Löhr, H.: Analyse der Struktur eines Begriffs - ein Beitrag zur Festlegung von Deskriptoren. Berlin 22(1975) 5, S. 28-33, 3 Abb., 4 Tab., 8 Lit.
- [Lö75b] Löhr, H.: Erste Erfahrungen bei der Ausbildung von Hochschulingenieuren im Fachgebiet wtI. Informatik, Berlin 22(1975) 4, S. 36-40, 1 Abb., 2 Tab., 5 Lit.
- [Nau08] Naumann, P.: Das Robotron Informationsrecherchesystem (IRS) AIDOS: Software-Technologie von gestern? In diesem Band S.
- [Ro69] Typenfeinprojekt "Einsatz des Robotron 300 für die Information und Dokumentation", Dresden: VEB Kombinat Robotron, Dresden (1969)
- [Rö74] Löhr, H.; Römer, R.: Überführung des R300-Typenprojekts in den Routinebetrieb. Informatik, Berlin 21(1974) 3, S. 30-34, 5 Abb., 4 Tab., 5 Lit.