

IZ-Arbeitsbericht Nr. 6

**Informationserschließung und -bereitstellung
zwischen Deregulation, Kommerzialisierung
und weltweiter Vernetzung
- Schalenmodell -**

Jürgen Krause

September 1996



InformationsZentrum
Sozialwissenschaften

Lennéstraße 30
D-53113 Bonn
Tel.: 0228/2281-145
Fax.: 0228/2281-120
email: krause@bonn.iz-soz.de
Internet: <http://www.social-science-geis.de>

ISSN: 1431-6943

Herausgeber: Informationszentrum Sozialwissenschaften der Arbeits-
gemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI)

Druck u. Vertrieb: Informationszentrum Sozialwissenschaften, Bonn
Printed in Germany

Das IZ ist Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS),
einer Einrichtung der Wissenschaftsgemeinschaft Blaue Liste (WBL)

Inhalt

1 Einleitung	4
2 Deregulation durch Kooperation und Internet	8
2.1 Stellung der Kooperationspartner	9
2.2 Internet	10
2.3 Meinungen über die Datenkonsistenz intellektueller Erschließungsverfahren	10
3 Informationstechnologischer und informationswissenschaftlicher Wandel	12
4 Beispiel: Integriertes Informationssystem	14
5 Schalenmodell und Deregulation als Lösungsansatz	16
6 Literatur	23

1 Einleitung

Informationsservicestellen und Informationszentren mit ihren zentralen Diensten der Inhaltserschließung und -bereitstellung von Fachliteratur, der Forschungsdokumentationen und Faktendatenbanken als Basis einer umfassenderen Informationsversorgung haben sich in den letzten zwanzig Jahren als wesentlicher Bestandteil der Informationsbeschaffung in Wissenschaft und Wirtschaft durchgesetzt.

Der Zuwachs an weltweit aufgelegten Datenbanken ist enorm. Im Programm der Bundesregierung 1996-2000 „Wissenschaftlich-technische Information für das 21. Jahrhundert“ vom Dezember 1995 (S. 12) wird angegeben, daß sich der weltweite Bestand an Online-Datenbanken von 1990 - 1994 um 2000 auf 6.220 erhöht hat. Hinzu kommen 1994 ca. 6.250 CD-ROM Titel. Der Bericht des Bayerischen Staatsministeriums für Unterricht, Kultus, Wissenschaft und Kunst „Wissenschaftliche Information im elektronischen Zeitalter“ (Abschnitt 1.1) vom Juli 1995 geht von 8.800 Fachdatenbanken aus (Erhöhung in den letzten fünf Jahren um 30 %).

Wenn auch die klassischen Informationserschließungsinstanzen wie Zeitschriften, Bücher, Konferenzen und zwischenmenschliche Netzwerke nichts an Anziehungskraft verloren haben und weiter wachsen (z.B. Erhöhung der Anzahl der wissenschaftlichen Zeitschriften in den Jahren 1991 - 1995 um 26 % auf 147.000 Titel), gehört heute die Datenbankrecherche zur Normalität, wie früher nur die Nutzung der Büchereien. Neue Medien wie die CD-ROM und die wachsende Teilnahme an Netzwerken wie dem Internet verstärken den Trend zur Normalität beim Gebrauch elektronisch erschlossener Informationen.

Bei aller Einzelkritik an den Verfahren der Inhaltserschließung, die den heutigen Systemen zugrunde liegen (cf. Abschnitt 1.3), wäre diese rasche Verbreitung der Datenbanken nicht möglich gewesen, wenn die verwendeten Inhaltserschließungsverfahren zur Indexierung der Textdokumente sich nicht als ein im Kern gangbarer und leistungsfähiger Weg erwiesen hätten.

Literaturdatenbanken werden heute entweder intellektuell oder automatisch indexiert. Hersh 1995:76 geht sogar davon aus, daß „most modern commercial databases“ intellektuell indexiert werden.

Die Diskussion um Vor- und Nachteile beider Grundtypen der Inhaltserschließung ist so alt wie die ersten praktisch eingesetzten Retrievalsysteme.

- a) Intellektuelle Indexierungsverfahren verkörpern am klarsten die Grundidee einer möglichst weitgehenden Regulation, die Konsistenz erreichen soll.

So erschließt z.B. das Informationszentrum Sozialwissenschaften (IZ, Bonn, cf. Krause/Zimmer 1996 als Überblick) seine Literatur- und Forschungsprojektdatenbanken intellektuell auf der Basis eines Thesaurus, der partiell hierarchisch organisiert ist. Alle vergebenen Begriffe (Ausnahme: Zusatzfeld für „freie Begriffe“) entnimmt der Indexierer einem ständig zu pflegenden Thesaurus. Die hier enthaltenen Terme bilden ein geschlossenes semantisches System. Ihr Vorteil ist, daß die Indexierungstiefe bei der Thesauruserstellung kontrolliert werden kann und eine semantische Vereinheitlichung bereits auf der Ebene der Indexierung erzwungen wird. Die Begriffe „verlieren“ allerdings ihre umgangssprachliche Semantik und sind quasi formalsprachlich in dem beschränkten, vorgegebenen Vokabular zu interpretieren. Diese Eigenschaft kontrollierter, intellektueller Indexierung wird am deutlichsten, wenn ein Indexierer einen Wunschbegriff im Thesaurus nicht findet.

Die Frage, ob eine intellektuelle Indexierung auf der Basis kontrollierter Thesauri einer automatischen qualitativ überlegen sei, führt zu den unterschiedlichsten Einschätzungen, die sich in der Regel auf Evaluationsergebnisse zu verschiedenen Studien aus dem englischsprachigen Raum stützen (cf. als Überblick Hersh 1995: chapter 7).

Blair/Maron 1985 schlossen aus ihren Evaluationen zu juristischen Dokumenten, die mit STAIRS erschlossen wurden, daß frühere Evaluationsergebnisse, die für die intellektuelle Indexierung zu nachteiligen Ergebnissen gekommen waren, sich nicht auf reale, große Textbestände übertragen ließen. Sie konstatierten immer dann Vorteile für das intellektuelle Verfahren, wenn ein hoher recall erwünscht sei:

„fulltext system [=automatische Indexierung] means the additional cost of inputting and verifying 20 times the amount of information that a manually indexed system [= intellektuelle Indexierung] would deal with. This difference alone would more than compensate for the added time needed for manual indexing and vocabulary construction.“

Salton 1986 wendet sich gegen diese Schlußfolgerung. Er kommt nach einer Sichtung verschiedener weiterer Studien und einer Analyse der Argumentation von Blair/Maron 1985 zu dem Schluß:

„...these conclusions may be more sentiment than fact. ... the evidence from several retrieval evaluations conducted with very large collections does not support the notion of output overload ... comparisons between manual and automatic indexing systems on large document collections indicate that the automatic-text-based systems are at least competitive with, or even superior to the systems based on intellectual indexing.“ (Salton 1986:650)

D. R. Swanson, den Salton 1986:656 mit seiner Arbeit von 1960 noch zitiert als „probably the earliest result showing the superiority of automatic text searching“, kommt in Swanson 1988:95 zu dem Schluß:

„Machines cannot recognize meaning and so cannot duplicate what human judgement in principle can bring to the process of indexing and classifying documents ... Consistently effective fully automatic indexing and retrieval is not possible.“

Dahlberg 1996:82 formuliert die Überzeugung der Überlegenheit einer intellektuell durch Fachwissenschaftler vorgenommenen Begriffswahl überspitzt, aber gerade dadurch in der Richtung klar verständlich:

„... vermeinte man seit mehr als 30 Jahren nun schon, mit dem Computer das menschliche Denken müssen ersetzen zu können. Gott sei Dank geht das nicht -... bei geschriebenen Texten, in denen die Wortwahl eines Autors mit ihren vielfachen Möglichkeiten eine unendlich große Vielfalt erlaubt, Gedanken in Worte zu kleiden, handelt es sich um eine unvorhersehbare Situation, die auch als indeterminierter Prozeß bezeichnet werden kann, der sich bekanntlich jeglicher befriedigenden Programmierung entzieht.“

Auch bei Vertretern statistisch quantitativer Verfahren wird die Güte einer intellektuell basierten Indexierung anerkannt. So stellt z.B. Womser-Hacker 1996: 7 fest:

„Die ideale Form der Anfrageerweiterung besteht in der Benutzung eines Thesaurus, um die Terme der Anfrage durch ihre Synonyme, Oberbegriffe und sonstige geeignete Terme zu ergänzen. Da nur selten ein manuell erstellter und gepflegter Thesaurus zur Verfügung steht, ist man auf automatische Techniken ausgewichen.“

Es ist wichtig zu erkennen, daß die fachlich fundierten Kritikpunkte gegen die intellektuelle Indexierung in der Regel nicht das dahinter stehende

informationswissenschaftliche Prinzip in Frage stellen. Es geht entweder um Erweiterungen und Zusätze (Behandlung vager Informationen, Ranking u.a.) oder um Alternativen, die wegen der Kosten der intellektuellen Verfahren vorgeschlagen werden; sie scheinen bei einer steten Ausdehnung des Umfangs der zu erschließenden Dokumente als nicht mehr tragbar.

- b) Aus Kostengründen wurde die intellektuelle Aufbereitung in kommerziellen Systemen in vielen Fällen durch automatische Freitextverfahren abgelöst, bei denen Thesauri und eventuell computerlinguistische Methoden (z.B. Grundformenrückführung) nur noch in Teilbereichen eine Bereinigung der sprachlichen Vielfalt der Ausgangstexte vornehmen.

Die DATEV-Datenbanken sind ein Beispiel für die „Reinform“ traditioneller automatischer Freitextsysteme (cf. DATEV 1994). Mit Ausnahme einiger aspektgebundener Deskriptoren regelt nur eine Stopwortliste die Auswahl. Ein weiteres Beispiel sind die JURIS-Datenbanken (Basis GOLEM/PASSAT, cf. Möller 1993). Hier ist interessant, daß nach über 15jähriger praktischer Erfahrung vehement die bei Entwicklungsbeginn abgelehnte intellektuelle Erschließung wieder als Heilmittel für die empirisch beobachteten schlechten Retrievalleistungen von JURIS angesehen wird (cf. Wolf 1992; Möller 1993).

Somit ist die Frage, welche der beiden Methoden heute vorzuziehen ist, durchaus offen. Sie wäre durch vergleichende Evaluationen für den jeweiligen Anwendungsbereich zu klären und ist zudem stark mit der Verbreitung von Autorenabstracts in einem Fachgebiet und der rechtlichen Ausgangssituation bei der Verwertung von Volltexten verknüpft (cf. Abschnitt 1.3). Die gleiche Unsicherheit herrscht in bezug auf alternative Ansätze automatischer Indexierung, die die heutigen Standardverfahren nicht ergänzen, sondern vollständig ersetzen würden, wie z.B. statistisch-quantitative Verfahren. Schon Fuhr/Müller 1987 stellten fest, daß der einen solch tiefgreifenden Wechsel rechtfertigende deutliche Qualitätssprung bisher nicht nachgewiesen werden konnte.

Womser-Hacker 1996:1, die einen modifizierten statistisch-quantitativen Ansatz der Inhaltserschließung vertritt, kommt zu dem Schluß:

„Daß aber Online-Hosts wie STN-International, DATASTAR oder Dialog Information Services, deren Datenbanken aus mehreren Millionen Dokumenten bestehen und täglich wachsen, die Philosophie ihrer Retrievalsprache und ihres Systemdesigns komplett umstoßen, ist nahezu unvorstellbar, zumal ... vielfältige Konsequenzen mit einem derartigen Wechsel verbunden sind. Kosten/Nutzen-Faktoren, die hier

wohl die entscheidende Rolle spielen dürften, weisen nicht ausschließlich, d.h. für alle vorliegenden Konstellationen, in die positive Richtung.“

Wenn im folgenden dennoch für ein alternatives Modell für Informationseinrichtungen plädiert wird, ist das kein Widerspruch zur obigen Argumentation. Per se ist das bisherige Modell der Inhaltserschließung durchaus ein gangbarer Weg. Es gibt zudem eine Reihe von Vorschlägen erfolgversprechender Zusatzkomponenten, die Effizienzsteigerungen durch kontinuierliche Verbesserungen erwarten lassen (Visualisierung der Oberflächen, intelligente Unterstützungskomponenten, extended Boole'sche Algebra, cf. Krause 1996a und 1996b).

Was sich geändert hat, sind die Rahmenbedingungen. Der technologische, wirtschaftliche, politische und gesellschaftliche Wandel der letzten Jahre brachte Strömungen und Meinungen hervor, die zum Modell der Informationsservicestellen - wie es sich in den letzten 20 Jahren mit seinen Dienstleistungen konsolidiert hat - in einigen Punkten in Widerspruch gerät. Die Irritationen reichen bis in die Wahl der Verfahren zur Inhaltserschließung hinunter, da Entscheidungen in diesem Bereich auch organisatorische und wirtschaftliche Rahmenbedingungen tangieren.

2 Deregulation durch Kooperation und Internet

IuD-Stellen, wie sie in den 70er Jahren gegründet wurden, waren sowohl technisch als in ihrer Philosophie der Informationserschließung weitgehend zentralistisch und monolithisch. Ein zentral aufgestellter Großrechner verwaltete die Daten. Die Klientel wurde über Terminals oder offline über Anfragen an die eine Zentralstelle bedient.

Dem entsprach die theoretische Grundlage der Inhaltserschließung. Nach einem normierten, intellektuell kontrolliertem Verfahren, das die Zentralstelle entwickelte und durchsetzte, erfolgt eine einheitliche Erfassung der Literatur. In diesem Denken kommt der Datenkonsistenz die höchste Priorität zu.

Auch die Auswahl der Dokumente folgt diesem Prinzip. Zwar unterstützt von Fachwissenschaftlern, aber letztlich wieder mit Stichentscheid der zentralen Organisation, legt die Zentralstelle z.B. die Liste der auszuwertenden Zeitschriften fest.

2.1 Stellung der Kooperationspartner

In diesem Denken bedeutet die Ausdehnung der Dokumentationstätigkeit auf ein neues Gebiet, daß die Zentralstelle um die entsprechenden Fachwissenschaftler aufgestockt werden sollte. Kooperative Strukturen, die ergänzend zur Zentralstelle Dokumente in anderen Organisationen erfassen, sich aber an die Normen der Zentralstelle halten und deren Daten in eine Gesamtdatenbestand integriert werden, sind die ersten Aufweichungen des zentralistischen Grundansatzes. Idealerweise gehen sie in nationale und internationale Kooperationen über, deren gleichberechtigte Partner sich an gemeinsam ausgehandelte Übereinkünfte und Normierungen halten.

Diese Art der Deregulation schiebt heute eine Fülle ungelöster Probleme vor sich her:

- Kleinere Kooperationspartner sehen sich emotional immer mehr als reine „Zulieferer“ der Zentralstelle, ohne direkten Einfluß. Da ihnen der technologische Fortschritt - im Gegensatz zur Epoche der Großrechner - durchaus erlaubt, einfache Spezialdatenbanken z.B. auf CD-ROM selbst zu vertreiben, verstärken sich die Auflösungserscheinungen.
- Kleinere Kooperationspartner stehen wie die großen Informationsservicestellen unter dem Druck, ihre Einnahmen zu erhöhen und eigenständiges Profil zu zeigen. Vordergründig scheint dies mit Alleingängen schneller erreichbar.
- Auf den allgemeinen Kostendruck reagieren viele Institutionen mit der Auflösung ihrer Informationsabteilung. Für die großen Informationsservicestellen entsteht dadurch eine Lücke bei der Dokumentenerfassung, die sich nicht mehr schließen läßt.

Die Summierung der psychologischen, politischen und wirtschaftlichen Einflußfaktoren erhöht derzeit den Aufwand, Kooperationen funktionsfähig zu erhalten, deutlich, ohne gleichzeitig bei Beibehaltung der herrschenden IuD-Philosophie längerfristig die Strukturen vernünftig stabilisieren, geschweige denn ausweiten zu können. Die Datenkonsistenz und die lückenlose Erfassung der relevanten Dokumente ohne Zeitverzug sicherzustellen, wird im gegenwärtigen Umfeld immer schwieriger. Die Gefährdung beider Kriterien, Vollständigkeit und Aktualität, trifft jedoch den Kern der oben geschilderten IuD-Philosophie.

2.2 Internet

Der breite Zugang zu den Netzwerken wirkt einer zentralistischen Doktrin der Informationserschließung per se entgegen. Überall auf der Welt können Gruppen auftreten, die zu Spezialgebieten Informationen gesammelt haben. Der Benutzer wird auf sie zugreifen wollen, gleich nach welchen Verfahren sie erschlossen oder in welchem System sie angeboten werden. Das obige Kooperationsmodell würde verlangen, daß die zuständige Informationsservicestelle mit diesen Anbietern Kontakt aufnimmt und sie überzeugt, bestimmte Normen der Inhaltserschließung einzuhalten. Das mag im Einzelfall funktionieren, jedoch nie als generelle Strategie. Es wird immer eine Fülle von Angeboten geben, die sich vorgegebenen Leitvorstellungen nicht unterordnen lassen. Früher lehnten die zentralen Informationsservicestellen Dokumente ab, die nicht bestimmte Regeln der Erschließung einhielten, wodurch der Benutzer (idealerweise) immer einem homogenisierten Datenbestand gegenüberstand. Darauf ist die gesamte IuD-Methodik, einschließlich der Verwaltungsstruktur der Zentren ausgerichtet. Ob man dies für richtig oder falsch hält, diese Ausgangssituation ist in einem System weltweiter Vernetzung nicht mehr gegeben. Das Postulat der Datenkonsistenz als wesentlicher Eckpfeiler heutigen IuD-Handelns erweist sich schon aus diesem Grund als Illusion.

Auf diese Veränderung muß die heutige IuD-Landschaft reagieren. Es sind inhaltliche und auch organisatorische Konzepte zu entwickeln, die mit statt gegen diese Deregulation arbeiten.

Die informationswissenschaftliche Auseinandersetzung begegnet dieser Situation bisher durch Zusatzkonzepte wie den Hypertextsystemen. Sie ergänzen die bisherigen Suchstrategien, was richtig und unverzichtbar ist. Sie betreffen jedoch nicht den Kern der Deskriptorensuche, die die Hauptlast der Rechercheleistung trägt.

2.3 Meinungen über die Datenkonsistenz intellektueller Erschließungsverfahren

Als generelle Kritikpunkte werden genannt:

- Die Kosten der intellektuellen Aufbereitung sind zu hoch.

Häufig überschätzen Außenstehende die tatsächlich anfallenden Kosten. Im Schnitt liegen sie für die intellektuelle Deskriptorenvergabe bei 22,00 DM pro Dokument. Das Einsparungspotential ist somit bei einem Systemwechsel

hin zur automatischen Indexierung begrenzt, da auch sie nicht zum Nulltarif zu bekommen ist. Mehr ließe sich bei der Abstrakterstellung sparen (etwa 35,00 DM pro Dokument). Solange es sich nicht durchsetzen läßt, daß zumindest Autorenzusammenfassungen üblich sind, kann dieser Posten aber nicht eingespart werden. Auch eine automatische Indexierung braucht zumindest diese Textgrundlage. Der Anteil von Zeitschriftenaufsätzen mit Autorenreferaten liegt aber z.B. bei den Daten der sozialwissenschaftlichen Literaturdatenbank SOLIS nur bei 20 %.

Die Alternative Volltextindexierung kommt wegen der Rechtslage derzeit in vielen Bereichen nicht in Frage. Die Verlage erlauben die Wiedergabe vollständiger Zeitschriftenaufsätze und Bücher nicht, weil sie um den Absatz ihrer Printprodukte fürchten (wohl zu recht).

Mittelfristig gesehen könnte sich bei beiden Alternativen rasch etwas ändern. Es sollte auch versucht werden, diese Veränderungen zu induzieren.

- Die hohe Qualität und vor allem die Datenkonsistenz der intellektuellen inhaltlichen Erschließung durch Fachwissenschaftler an Informationszentren wird bestritten:
 - Jeder Fachwissenschaftler indexiere nach eigenen Kriterien und Wissenshintergrund. Die Konsistenz werde postuliert, aber in der Praxis nicht erreicht.
 - Fachwissenschaftler an Informationsservicestellen verlieren die Verbindung zu neueren Entwicklungen. Zumindest indexieren sie im Geist früherer Ansichten, was die Thesaurusvorgabe noch unterstützt. Deshalb sollte eine zentrale Informationsstelle zumindest weitgehend die fachwissenschaftliche Kompetenz universitärer Forschungsstellen nutzen, um einer wachsenden Fachferne entgegenzuwirken.
- Benutzer solcher Systeme sehen in der Regel nicht in der Schlagwortliste nach, sondern formulieren ihre Anfrage direkt. Wird der Thesaurus nicht ständig gepflegt und um neue (Mode-)begriffe eines Fachgebiets erweitert, ergeben sich zu große Diskrepanzen zwischen den vom Benutzer gewählten Begriffen und den Thesaurusstrukturen.

Interessant ist, daß all diese Argumente nicht prinzipieller, sondern wirtschaftlicher Natur sind. Bei einem höheren Geldeinsatz und veränderten Organisationsformen ließen sich die angegebenen Nachteile durchaus beseiti-

gen, ohne das zugrunde liegende Postulat der intellektuellen Indexierung aufzugeben.

Wie in der Einleitung angeführt, muß die Frage nach der effizientesten Art der Inhaltserschließung heute als offen angesehen werden. Die Frage läßt sich nur durch anwendungsbezogene, vergleichende Evaluationen der verschiedenen vorgeschlagenen Verfahren klären. Ob den Informationseinrichtungen, die heute intellektuell indexieren, noch die Zeit zu solchen Klärungen gelassen wird, ist allerdings fraglich. Die Meinung der politischen Entscheidungsinstanzen präferiert deutlich die Gegenargumente (cf. S. 30-32 aus dem Programm der Bundesregierung 1996-2000 „Information als Rohstoff für Innovation“ vom Sept. 1996).

Nicht unerheblich für die Ablehnung der intellektuellen Indexierung dürfte zudem die globale Dominanz der Deregulationstheorie in allen Bereichen wirtschaftlichen und politischen Handelns sein, die sich heute bei politischen Entscheidungen aller Art in den Vordergrund schiebt.

Intellektuelle Erschließung ist zweifellos die von allen Ansätzen weitestgehende Regulation, deren Regeln wiederum eine zentrale Instanz vorgibt. An sie ist die intellektuelle Indexierung in bezug auf die Durchsetzungsfähigkeit gebunden. Deshalb gilt hier das gleiche wie bei dem generellen Internet-Beispiel. Alle „Mitspieler“ eines Anwendungsgebiets müßten von der Richtigkeit dieses Vorgehens überzeugt werden. Weltweit gesehen, eine Illusion.

3 Informationstechnologischer und informationswissenschaftlicher Wandel

Die derzeitige Situation kommerzieller und öffentlich geförderter Informationszentren ist durch einen tiefgehenden Wandel in der gesamten zugrunde liegenden Informationstechnologie geprägt (Beispiele: Internet, Multimedia, verteilte Datenbanken, integrative Recherche verschiedener Datentypen wie Fakten und Literatur, Print on demand).

Informationserschließung und -bereitstellung kann heute nicht mehr nur in der Anwendung standardisierter Literaturdatenbanken bestehen oder sich auf Buchpublikationen und das Knüpfen fachwissenschaftlicher Kommunikationsnetze auf der Basis von Zeitschriften, Tagungen und Telefongesprächen beschränken - so wichtig all diese konventionellen Methoden der Informationsgewinnung auch heute noch sind.

Akzeptiert man diese Überlegungen wird ein Strukturproblem sichtbar. Gründungen der 70er Jahre sahen den personellen Hauptbedarf bei den Fachwissenschaftlern. Dokumentare verstanden sich vor allem als Fachleute für Inhaltserschließung nach dem Muster der intellektuellen Verschlagwortung. Die Vorstellung war, daß ein kleines Rechenzentrum mit dem für die Anwendungsindustrie üblichen 5 % Personalanteil Fremdsoftware (vor allem das Datenbanksystem) einkauft und installiert, auf die Fachwissenschaftler und Dokumentare ohne vertiefte EDV-Kenntnisse zugreifen. Die Personalzusammensetzung spiegelt noch heute diese Ausgangssituation wider.

Diese Struktur ist den heutigen Anforderungen zweifach nicht gewachsen

- Die technologischen Probleme lassen sich nicht mehr nebenbei erledigen. Dies scheint ein Widerspruch, da z.B. die Installation und der Betrieb eines Datenbanksystems heute fraglos wesentlich einfacher ist als vor 20 Jahren. Heute gibt es jedoch eine Fülle neuer Möglichkeiten mit spezifischen Problemstellungen (das Internet ist nur ein Beispiel), die für hochwertige Informationsdienstleistungen zu nutzen sind. Ohne weitergehende Informatikkenntnisse bei 10 - 20 % der Belegschaft, zusätzlich zum heutigen Rechenzentrumspersonal, bleibt jede Beteiligung an den neuen Informationsmedien - sei sie auch noch so gut gemeint und fachwissenschaftlich wertvoll - zwangsläufig dilettantisch.
- Der Informationswissenschaft ist es nicht gelungen, sich mit ihrem Know how in den kommerziellen und öffentlich geförderten Informationsservicestellen zu verankern. Vergleicht man die heute in Informationsservicestellen eingesetzten Verfahren, z.B. die der Inhaltserschließung, mit den Themen der informationswissenschaftlichen Literatur, wird die Kluft zwischen Forschungsansätzen und in der Praxis eingesetzten Verfahren deutlich. Ein Beispiel sind die statistisch-quantitativen Recherchealgorithmen, die erst langsam Eingang in die Praxis finden. Die überwiegende Mehrzahl kommerziell angebotener Datenbanken verwenden die Deskriptorensuche auf der Basis der Boole'schen Algebra. Ein anderes Beispiel sind Erkenntnisse der Softwareergonomie, die direkt die Benutzerakzeptanz beeinflussen.

Generell gilt deshalb: Moderne Informationserschließung und -bereitstellung muß heute gleichgewichtig auf drei Säulen ruhen, auf der der Informationstechnologie, der fachwissenschaftlichen und der informationswissenschaftlichen.

4 Beispiel: Integriertes Informationssystem

An dem Beispiel der Verbindung von textueller und Faktenrecherche über unsicheres Wissen in einem intelligenten Informationssystem soll die notwendige Verknüpfung zwischen fachwissenschaftlichen, informationstechnologischen und informationswissenschaftlichen Know how verdeutlicht werden.

Information Retrieval Systeme zur Suche in unstrukturierten Dokumenten (vor allem in Texten) und Datenbanksysteme für Faktenabfragen werden derzeit als zwei unterschiedliche Forschungsrichtungen gehandhabt, obwohl die Parallelen zwischen beiden Konzepten offensichtlich sind und ein Austausch für beide Seiten fruchtbar wäre. Dem entspricht auch die Nutzung kommerzieller Produkte in der Industrie.

Diese gegenseitige Abgegrenztheit zeigt sich in der Softwareindustrie in verschiedenen, voneinander getrennten Produkten und in der Forschung z.B. in der Durchführung eigener Konferenzen oder auch in der Existenz unterschiedlicher Fachorganisationen (z.B. verschiedene Fachgruppierungen innerhalb der Gesellschaft für Informatik).

In vielen Anwendungsbereichen lassen sich die realen Gegebenheiten jedoch nicht analog zu den oben beschriebenen Konzepten trennen. Zum einen liegt das Datenmaterial meist in heterogener Form vor und sollte nicht künstlich auf getrennte Systeme verteilt werden, sondern in einem homogenen Informations- und Archivsystem auffindbar und dokumentierbar sein.

Ein integriertes System muß dabei über folgende Grundelemente verfügen:

- a) Verschieden mediale Wissensseinheiten mit geeigneten, kompatiblen Repräsentationsformalissen, vor allem unstrukturiertes Text- und Faktenwissen in Tabellen (dies ist ein Bereich, der für bestimmte Medien weitgehend unerforscht ist, z.B. die inhaltliche Repräsentation von Bildern, Videos, Tonträgern etc.)
- b) objektorientierte Kategorisierungsmechanismen zur Differenzierung verschiedener Typen innerhalb eines Mediums/einer Dokumentationseinheit (z.B. bei Texten innerhalb der Büroumgebung: Publikation, Brief, e-mail-Mitteilung, Umlauf, Kooperationsabsprache etc.) im Sinne einer Objekthierarchie, die neben den eigentlichen Inhalten als Suchinhalt bzw. Filter dienen kann (z.B. Suche nach allen Briefen mit gelbem Logo in der linken Ecke)

- c) einheitliche Retrievalsprache - was im weitesten Sinn zu interpretieren ist, und z.B. auch eine intelligente, objektorientierte, graphisch-direktmanipulative Benutzungsoberfläche darstellen kann (z.B. Suche nach Werkstoffen, die bei einer hohen Temperatur die Elastizität nicht verlieren, die im firmeneigenen Prüfbericht positiv beurteilt werden und eine bestimmte Preisvorstellung nicht überschreiten).

Dabei könnte die Systemintelligenz soweit gehen, daß das System die Entscheidung, in welchem Teil des Datenbestands die Antwort zu finden ist, übernimmt (Durchlässigkeit auf der Repräsentationsebene)

- d) vage Anfrageattribute wie *groß, klein, gut, bedingt* etc., die mittels geeigneter Modellierungsansätze (z.B. über die sog. membership function der fuzzy logic theory) repräsentiert werden müssen, wobei hier insbesondere die Kontextproblematik eine Rolle spielt (d.h. vage Attribute können nur innerhalb eines Kontexts adäquat interpretiert werden)
- e) Systemkomponenten, die Unsicherheiten auf seiten des Wissens handhaben und u.U. ausgleichen können (z.B. Ähnlichkeitsmodule zur Überbrückung von Datenlücken).

Die Benutzer vieler Informationsservicestellen haben die oben beschriebenen Probleme und bräuchten für eine qualitativ hochstehende und effiziente Informationsbeschaffungen integrierte Informationssysteme. M. E. ist es Aufgabe der Informationsstellen, ihren Bedarf zu erfüllen. Wer sollte es sonst tun? Das gleiche gilt für die neuen Probleme, die sich durch die rasche Entwicklung des Internet ergeben und die sich ebenfalls nicht auf die Kommunikationsfunktion (besseres e-mail) und das Browsen mit Hypertextlinks reduzieren lassen.

5 Schalenmodell und Deregulation als Lösungsansatz

Überlegungen zu einer Umgestaltung müssen die Argumente der verschiedenen Ebenen berücksichtigen, die in den Kap.1 - 3 genannt wurden. Es genügt z.B. nicht, eine informationswissenschaftlich akzeptable Lösungsstrategie zu haben, wenn sie nicht politisch und gesellschaftlich durchsetzbar ist und die wirtschaftlichen Rahmenbedingungen außer Acht läßt. Als Faktoren sind zu berücksichtigen:

- Überzeugungen im politischen, fachwissenschaftlichen und gesamtgesellschaftlichen Umfeld:
 - Vorherrschen von Deregulationsüberzeugungen (Rückzug des Staates aus dem Informationsmarkt)
 - Zweifel am Konzept einer rein intellektuellen Inhaltserschließung in einer zentral verwalteten Gesamtstruktur (Kostengründe, wachsende Fachfremdheit, nicht erreichbare Datenkonsistenz)
 - Wachsende Schwierigkeiten des Erhalts und des Aufbaus neuer Kooperationsbeziehungen im Rahmen einer zentral gelenkten Informationsstruktur
- Veränderungen durch die neueren technologischen und informationswissenschaftlichen Entwicklungen (genereller Entwicklungsrückstau, neue Möglichkeiten von Informationserschließung und -bereitstellung, Veränderung bestehender Strukturen durch die Erweiterung auf weltweite Informationssuche)
- Überführungsmöglichkeiten für die bisher bestehenden v. a. personellen Strukturen (Weiterqualifizierung, Integration neuen Know hows)
- Steigerung der Wirtschaftlichkeit (Kostenreduktion pro Informationseinheit)

Zentral für viele der obigen Problembereiche scheint die Deregulationsthese. Es soll versucht werden, sie einem Lösungsansatz zugrunde zu legen, der Deregulation adäquat, d. h. mehrschichtig, operationalisiert. Deregulation einer bisher vor allem zentralistisch tradierten Informationserschließung und -bereitstellung ist sowohl aus informationstechnologischen, aus fachwissenschaftlichen, informationswissenschaftlichen und politisch-wirtschaftlichen Gegebenheiten und geltenden Überzeugungen meines Erachtens zu empfehlen. Deregulation allein durch strikte Kommerzialisierung greift jedoch zu kurz. Sie erreicht das Gegenteil von dem, was als Ziel vorgegeben wird. Deregulation muß so interpretiert werden, daß sinnvolle neue Strukturen entstehen, die auch in Zukunft Informationsdienstleistungen auf hohem qualitativem Niveau ermöglichen.

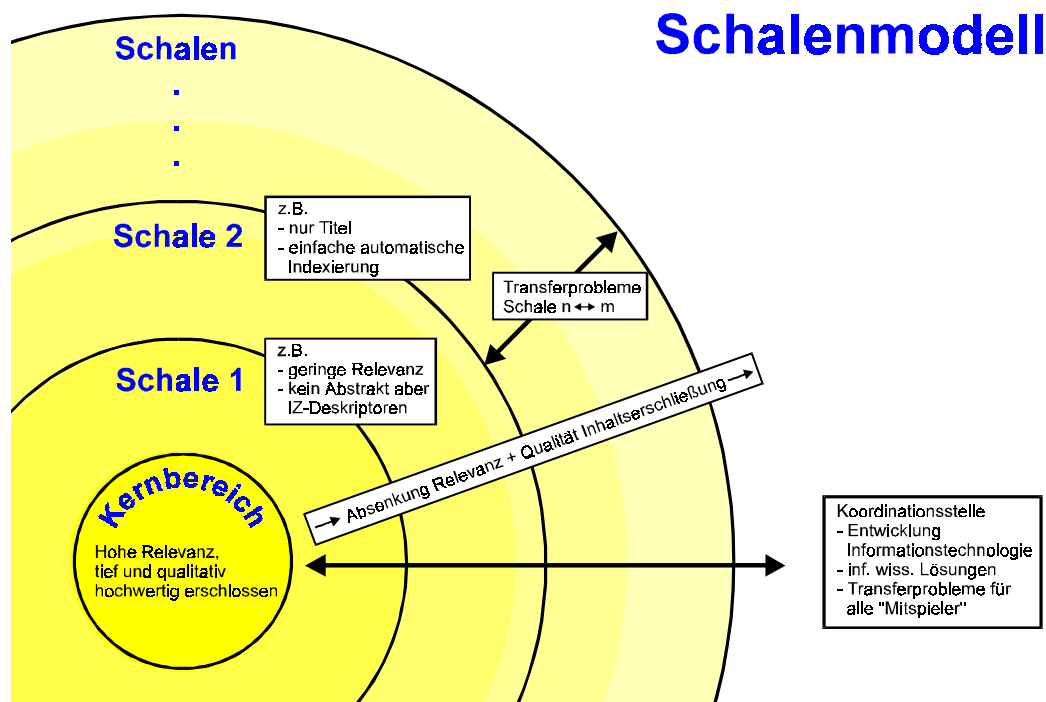
Deregulation ohne Koordinationsinstanzen führt zu anarchischen Strukturen. Je stärker auf allen Ebenen dereguliert wird, um von zentralistischen Strukturen wegzukommen (personell, informationstechnologisch oder beim informationswissenschaftlichen Modell der Informationserschließung und -bereit-

stellung), um so wichtiger werden Kontrollinstanzen. Die zerfallenden Einheiten und neu hinzukommende Gruppen müssen flexibel aufeinander bezogen werden, ohne daß der Herrschafts- und Normierungsanspruch der überwundenen Modellvorstellungen unter einer neuen Terminologie wieder auflebt.

Ein Schalenmodell der Informationserschließung scheint diesen Grundüberlegungen und vielen der genannten Probleme gerecht zu werden. Es soll am Beispiel des Datenbankaufbaus und der Inhaltserschließung von Literaturdatenbanken diskutiert werden.

Generell läßt das Schalenmodell verschiedene Niveaus der Datenrelevanz und Inhaltserschließung zu, die in einem gemeinsamen Informationssystem aufeinander bezogen werden. Normierung und qualitative Anforderungen werden nicht zentralistisch durchgesetzt, sondern koordiniert und verwaltet. Dies führt zu der Möglichkeit, Gruppen und Datenbestände zu integrieren, die in einem zentralistischen Modell abgelehnt würden.

Die Aufgabe der Informationsservicestelle verschiebt sich von der Rolle als Durchsetzungsinstanz und Anbieter einer einheitlichen Datenbank zum Koordinator verschiedener nicht-hierarchisch miteinander verbundener Gruppen, der die informationswissenschaftlichen und informationstechnologischen Hürden der Informationserschließung und -bereitstellung für alle Partner löst und für eine organisatorische Koordination sorgt.



Bezogen auf die Literaturdatenbank SOLIS des IZ ergäbe sich z.B. folgender Aufbau:

- Die innerste Schale enthält den Kern der relevanten Literatur. Er wird möglichst tief und qualitativ hochwertig erschlossen. Die Qualitätskontrolle liegt in der Hand der koordinierenden Informationsservicestelle.

Auf diese Schale kann weder aus fachwissenschaftlich-dokumentarischen noch aus organisatorischen Gründen verzichtet werden. Erst die dadurch entstehende Datenkonsistenz eines Kernbereichs schafft den Anreiz für weitere Partner, die folgenden Schalen zu bedienen.

- Die zweite Schale lockert die Relevanzbedingungen und parallel dazu die Anforderungen an die Qualität der Inhaltserschließung. Für den Bereich der Kommunikationswissenschaften wäre Schale 2 z.B. die Menge an Dokumenten, die zwar nach dem IZ-Thesaurus verschlagwortet, aber ohne Abstract angeboten werden.
- Schale 3 könnte alle Dokumente enthalten, deren Relevanz gegenüber den Schalen 1 und 2 niedriger ist und die nach anderen Normen erschlossen sind (z.B. anderer Thesaurus).

- Schale 4 enthielte die Ansetzung der Bibliotheken. Neben den gebundenen Deskriptoren (Beispiel Autor) steht für die Inhaltserschließung nur der Titel zur Verfügung, der automatisch indexiert wird.

Wieviele Schalen angesetzt werden und welche Merkmale sie definieren, richtet sich nach den Gegebenheiten eines Fachgebiets und den sich beteiligenden Gruppen. Wichtig ist nur, daß eine Zuordnung von erschlossenen Daten nach den Prinzipien der Relevanz und Erschließungstiefe bzw. -variation erfolgt und daß die Koordinationsstelle einen konsistent und tief erschlossenen Kernbereich vorgibt.

Am Schalenmodell können sich prinzipiell alle Mitspieler der Fachwissenschaften beteiligen, ob sie viel oder wenig erschließen, ob sie regelmäßig liefern können oder nur schwerpunktmäßig und projektbezogen arbeiten. Damit müßte die Menge der erschließbaren Literatur steigen, ohne daß der Etat der bestehenden Informationsservicestellen wächst. Selbstverständlich wird die Koordinationsstelle versuchen, die Partner auf eine möglichst einheitliche und qualitativ hochwertige Erschließung einzuschwören. Bei Interessens-, Leistungs- oder Meinungsdivergenzen führt dies jedoch nicht mehr wie im zentralistischen Modell zum Ausschluß der Mitspieler.

Gegenüber dem Idealmodell einer durchgehend konsistenten Erschließung sinkt auch unter besten Bedingungen die Datenkonsistenz und damit die Recherchequalität. Gegenüber einem realistischen Szenario gleicht die so erreichbare Datenfülle die Konsistenzbrüche aus, wenn folgende Voraussetzungen gegeben sind:

- Die Koordinationsstelle löst die Probleme der Informationstechnik für alle am Verbund beteiligten Partner. Dazu gehört die Pflege der Netzstruktur, die Auswahl und eventuell Entwicklung geeigneter Software für die Inhaltserschließung und Recherche genauso wie die Entwicklung geeigneter Verfahren zur Erstellung elektronischer und Printprodukte.

Kleinere fachwissenschaftliche Partner sind in der Regel gezwungen, informationstechnologisch suboptimal zu arbeiten, weil das entsprechende Know how fehlt. Dieses Know how - einschließlich der maschinellen Ausstattung - ist so kostenintensiv, daß es nicht mehrfach in Kleingruppen vorgehalten werden sollte.

- Die Koordinationsstelle befaßt sich mit den neueren informationswissenschaftlichen Erkenntnissen. Zu klären wäre z.B. die Leistungsfähigkeit statistisch-quantitativer Methoden im Information Retrieval oder die Effizienz intelligenter Komponenten. Auf dieser Basis stellt sie die für die

Fachwissenschaften geeigneten Inhaltserschließungs- und Recherchemethoden zur Verfügung. Gleichzeitig erschließt sie neue Möglichkeiten der Informationserschließung und -bereitstellung, die durch die informationstechnische Entwicklungen eröffnet werden.

- Bei der Recherche muß jede Schale von den entfernter liegenden getrennt werden können. Benutzern, die auf eine qualitativ hochwertige Inhaltserschließung Wert legen - und dafür einen kleineren Datenbestand in Kauf nehmen - darf eine Vermischung mit den entfernteren Schalen nicht aufgezungen werden.

Gleichzeitig ist das Problem des Übergangs von einer Schale zur anderen zu lösen. Deskriptoren einer äußeren Schale (Beispiel: nur automatische Titelindeizierung) können nicht gleichwertig mit denen der inneren Schale verbunden werden. Ob es gelingt, hier vernünftige Transferstrategien zu finden und programmtechnisch umzusetzen, entscheidet wesentlich mit über die Effizienz des Schalenmodells. Informationswissenschaftlich betritt man hier Neuland.

Das Schalenmodell verspricht zudem einen interessanten Effekt auf die Bemühungen, bei der Inhaltserschließung Normen zur Erhöhung der Datenkonsistenz durchzusetzen. Nicht von ungefähr scheitern Normierungsbestrebungen in individualistisch verfaßten Fachwissenschaften (Beispiel Sozialwissenschaften) oft oder kommen nur mit quälender Langsamkeit zu als unverbindlich angesehenen Regelungen. Das Schalenmodell erlaubt ein kontinuierliches Aufeinanderzubewegen verschiedener Ansichten und Standards. Entscheiden sich z.B. die meisten Gruppen für ein bestimmtes Erschließungsmodell, kann die Koordinationsstelle versuchen, diesen Trend in der Fachwissenschaft durch organisierte Gesprächsrunden zu verstärken bzw. unter informationswissenschaftlichen Aspekten zu modifizieren.

Es ist auch nicht zu befürchten, daß sich durch die Freigabe Gruppen vermehrt in weiter außen liegende Schalen einordnen. Ich würde vielmehr mit dem Ehrgeiz der Fachwissenschaftler rechnen, nicht hinter die Leistungsfähigkeit der Partnergruppen zurückzufallen.

Die stärksten Veränderungen bewirkt das Schalenmodell beim Selbstverständnis, der Personalstruktur und den Leistungsanforderungen heutiger Informationsservicestellen:

- Die zentrale Informationsservicestelle für ein Fachgebiet wird von einer zentrale Richtlinien durchsetzenden Instanz, die entscheidet, was relevant ist

und wie Inhaltserschließung und Recherche zu normieren sind, zu einer Koordinationsstelle in einem partnerschaftlichen Netz zwar kleinerer und größerer, aber immer gleichberechtigter Partner.

Von kleineren Partnern nimmt sie nicht nur erfaßte und inhaltlich erschlossene Dokumente entgegen, sondern sie gibt etwas zurück. Sie unterstützt den Partner in allen informationstechnischen und informationswissenschaftlichen Fragen und stellt konkret Software zur Verfügung. Das Know how, das sonst nur rudimentär selbst aufgebaut werden kann, setzt die Partner in die Lage, in der immer komplexer werdenden Welt der Informationstechnologie Schritt zu halten.

Partnerbeziehungen lassen sich in diesem System flexibel gestalten. Eine Regel muß nicht für alle als verbindlich angesetzt werden.

- Die Koordinationsstelle muß strukturelle Veränderungen vornehmen. Von der fast ausschließlich fachwissenschaftlich und dokumentarisch fundierten Kompetenzstruktur muß ein Teil auf die informationstechnologische und informationswissenschaftliche Kompetenz umgelenkt werden.

Gleichzeitig ist das allgemeine EDV-Niveau der Mitarbeiter anzuheben. Alle Arbeitsprozesse sind daraufhin zu überprüfen, ob ein verstärkter Rechner- und Softwareeinsatz Freiräume für neu hinzukommenden Aufgaben schafft.

Nicht vergessen werden darf, daß auch bei einer Fortführung des bisherigen Strukturmodells informationstechnologische Kompetenz weit über das bisher vorhandene Ausmaß hinaus bereitgestellt werden müßte. Der rasche Wechsel der technologischen Rahmenbedingungen in den letzten Jahren, der heute ganz andere Möglichkeiten der Informationserschließung und -bereitstellung eröffnet, läßt sich nicht mehr allein mit Rechenzentren bewältigen, die beim Personal den EDV-Ansätzen industrieller Betriebe oder der öffentlicher Verwaltungen folgen. Sie sind durch eine informationstechnologisch und informationswissenschaftlich ausgerichtete Forschungs- und Entwicklungsabteilung zu ergänzen.

Im begrenzten Umfang lassen sich die Strukturänderungen über Weiterqualifizierungsmaßnahmen erreichen. Die hierfür anzusetzende Qualifizierungszeit liegt nicht unter 12 Personenmonaten.

Weiterqualifizierung vor Ort läßt sich nur erfolgreich gestalten, wenn zumindest für eine Übergangszeit von zwei bis drei Jahren Sachkompetenz von außen (z.B. über Sonder- oder Drittmittel) gewonnen werden kann. Von

einer „Selbsteilungskraft“ bei den bisher üblichen Kompetenzstrukturen auszugehen, ist illusorisch.

Ist eine Verstärkung der informationstechnologischen Kompetenz nicht möglich, muß jede Umstrukturierungsanstrengung scheitern.

Gelingt die durch die neuen informationstechnologischen Rahmenbedingungen erzwungene Umstrukturierung, dürfte dies - als Nebeneffekt - neue Möglichkeiten eröffnen, durch Industriekooperationen einen höheren Anteil der Eigenfinanzierung zu erreichen. Die informationstechnischen und informationswissenschaftlichen Aspekte der Informationserschließung und -bereitstellung (z.B. intelligente Informationssysteme, Internet-researchsoftware, softwareergonomisch gestaltete graphische Benutzungsoberflächen) versprechen bei der Softwareerstellung höhere Synergieeffekte mit industriellen Bedürfnissen als die fachwissenschaftlich orientierten Informationsdienstleistungen.

Der vorgeschlagene Strukturwechsel auf der Basis des Schalenmodells läßt sich metaphorisch formuliert als Übergang von einer Planwirtschaft mit zentralistischer Neigung in eine Form der freien informationellen Marktwirtschaft beschreiben. Gesteuerte Deregulation mit verbindenden Koordinationsstrukturen scheint ein Rahmenkonzept, das den Übergang der Informationsservicestellen in die informationstechnologisch radikal veränderte Gegenwart und Zukunft bewältigen könnte. Die Möglichkeiten für solch ein Konzept bieten erst die neuen technologischen Rahmenbedingungen.

Zumindest ist meines Erachtens derzeit keine Alternative sichtbar, die eine effiziente Informationserschließung und -bereitstellung im Bereich der wissenschaftlichen Information im heutigen wirtschaftlichen und gesellschaftlichen Umfeld mit seinen rigiden Rahmenbedingungen sicherstellen könnte. Eine Linie, die sich auf Deregulation durch reine Kommerzialisierung in allen Bereichen beschränkt, mag in wenigen industriellen Bereichen wie in der Pharmazie oder der Werkstoffwissenschaft möglich sein. Für den Großteil der bisherigen IuD-Strukturen würde diese Politik das Ende einer umfassenden Informationserschließung und -bereitstellung bedeuten. Die Folgen wären erhöhte Doppelarbeit und sinkende Arbeitsqualität. An der Schlichtheit dieser Erkenntnis, die seit den 70er Jahren als Konsens der modernen Industriegesellschaft gelten konnte und ganz selbstverständlich auch bei Effizienz- und Wettbewerbsüberlegungen der Hochschulen zugrunde gelegt wurde, kommt man auch bei den gegenwärtigen gesellschaftlichen Rahmenbedingungen nicht vorbei.

6 Literatur

- Bericht der Sachverständigenkommission „Elektronische Fachinformation (EFI)“ an den Hochschulen in Bayern (1995): Wissenschaftliche Information im elektronischen Zeitalter. Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst, RB-Nr. 05/95/14. Juli 1995. München.
- BMBF (1996): Information als Rohstoff für Innovation. Programm der Bundesregierung 1996-2000. BMBF Bonn.
- Blair, D. C.; Maron, M. E. (1985): An evaluation of retrieval effectiveness for a full-text document-retrieval system. *Commun. ACM* 28.3, S. 289-299.
- Dahlberg; I. (1996): Replik zur Kritik des Hauptartikels „Ingetraut Dahlberg: Zur „Begriffskultur“ in den Sozialwissenschaften: Lassen sich ihre Probleme lösen?“. *EuS* 7. Heft 1, S. 82.
- DATEV 1994: Datenbanken Anwenderhandbuch, Nürnberg.
- Fuhr, N.; Müller, P. (1987): Probabilistic Search Term Weighting - some Negative Results. In: *Proceedings of the Tenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, S. 13-18.
- Hersh, W. R. (1995): *Information Retrieval: A Health Care Perspective*, Springer.
- Krause, J. (1996a): Principles of Content Analysis for Information Retrieval Systems. An Overview. In: Zuell, C.; Harkness, J.; Hoffmeyer-Zlotnik, J. (Eds.): *Text Analysis and Computer*. ZUMA-NACHRICHTEN Spezial, S. 77-104.
- Krause, J. (1996b): Visualisierung und graphische Benutzungsoberflächen. Arbeitsbericht Informationszentrum Sozialwissenschaften (IZ), Bonn und Universität Koblenz-Landau, Institut für Informatik, Koblenz.
- Krause, J.; Zimmer, M. (Hrsg.) (1996): *Informationsservice des IZ Sozialwissenschaften. Datenbankentwicklung und -nutzung, Netzwerke, Wissenschaftsforschung*. Bonn.
- Möller, Tong (1993): *Juris für Juristen (Law for Jurists)*. Dissertation - Universität des Saarlandes.
- Programm der Bundesregierung 1996-2000: Wissenschaftlich-technische Information für das 21. Jahrhundert. Stand: 01.12.1995, Bonn.
- Salton, G. (1986): Another look at automatic text-retrieval systems. *Comm. Assoc. Comp. Mach.*, 29, S. 648-656.
- Swanson, D.R. (1960): Searching natural language text by computer. *Science* 132, 3434, S. 1099-1104.
- Swanson, D.R. (1988): Historical note: Information retrieval and the future of an illusion. *J. Am. Soc. Info. Sci.*, 39, S. 92-98.

Wolf, Gerhard (1992): JURIS - Ein denkbarer einfacher Zugang zu allen Informationen, die Sie brauchen. jur-pc. 4. S. 1524-1810.

Womser-Hacker, C. (1996): Erweiterung Boole'scher Verfahren durch Gewichtungsansätze. Arbeitsbericht Informationswissenschaft Universität Regensburg.