

IZ-Arbeitsbericht Nr. 19

**Virtuelle Fachbibliothek Sozialwissenschaften
- Projektskizze -**

IZ Sozialwissenschaften: M. Kluck, J. Krause, M.N.O. Müller
in Kooperation mit AG Elfikom, TU Darmstadt:
R. Schmiede, H. Wenzel, S. Winkler, W. Meier

Februar 2000



InformationsZentrum
Sozialwissenschaften

Lennéstraße 30
D-53113 Bonn
Tel.: 0228/2281-0
Fax.: 0228/2281-120
email: krause@bonn.iz-soz.de
mr@bonn.iz-soz.de
schmiede@ifs.tu-darmstadt.de
meier@ifs.tu-darmstadt.de
Internet: <http://www.social-science-geis.de>

Dieser Arbeitsbericht basiert auf dem DFG-Projektantrag für die Virtuelle Fachbibliothek Sozialwissenschaften. Stand Juni 1999.

ISSN: 1431-6943
Herausgeber: Informationszentrum Sozialwissenschaften der Arbeitsgemeinschaft 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 Gottfried Wilhelm Leibniz (WGL).

Inhalt

| | |
|---|----|
| Über das Projekt | 4 |
| 1 Ausgangslage und Vorarbeiten | 6 |
| 1.1 Ausgangslage | 6 |
| 1.1.1 Generelle Aspekte und BMBF Vorhaben | 6 |
| 1.1.2 Das Modell polyzentrischer Informationsversorgung | 9 |
| 1.1.3 Lösungsansatz Schalenmodell | 14 |
| 1.1.4 Nutzung bestehender Vorarbeiten und Dokumentquellen | 20 |
| 1.2 Stand von Wissenschaft und Technik | 21 |
| 1.3 Ausgangspunkte der Realisierung des Projekts | 24 |
| 1.3.1 Der Darmstädter Virtuelle Gesamtkatalog (DVK) | 24 |
| 1.3.2 TOSCANA – Erkundungssystem zur Literatursuche | 26 |
| 1.3.3 Die sozialwissenschaftlichen Datenbanken und informationswissenschaftliche Entwicklungen am IZ Sozialwissenschaften | 27 |
| 1.3.4 Das SSG Sozialwissenschaften der Universitäts- und Stadtbibliothek Köln | 32 |
| 1.3.5 Die Bibliothek der Friedrich-Ebert-Stiftung | 33 |
| 1.3.6 Parallele Arbeiten sozialwissenschaftlicher Verlage | 35 |
| 2 Ziele des Projekts | 35 |
| 2.1 Integration verschiedenartiger Informationssysteme (Kataloge, Datenbanken) | 36 |
| 2.2 Integration und Transfer von verschiedenen Inhaltserschließungsverfahren | 37 |
| 2.3 Integrierte Bestellverfahren (für Volltexte, Papierkopien, Ausleihe) | 38 |
| 2.4 Verkürzung der Informationskette und Nutzung der elektronischen Volltexte | 38 |
| 2.5 Anbindung von Clearinghouse-Funktionen und Internetquellen (Planung 3. Jahr) | 39 |
| 2.6 Integration englischsprachiger Dokumente (Planung 3.Jahr) | 40 |
| Literaturverzeichnis | 41 |

Über das Projekt

Das Projekt Virtuelle Fachbibliothek Sozialwissenschaften (ViBSoz) zielt auf die integrierte Bereitstellung sozialwissenschaftlicher Literaturinformationen aus verteilten, verschieden strukturierten und inhaltlich unterschiedlich erschlossenen Datenbeständen, die sich in miteinander nicht verbundenen, heterogenen Organisationsstrukturen und Zugänglichkeitskontexten befinden (Institutsbibliotheken, Sondersammelgebiete der Universitätsbibliotheken (SSG), wissenschaftliche Spezialbibliotheken, Referenzdatenbanken, digitale Volltexte).

Im Projektkontext sollen also in erster Linie die Probleme des inhaltlichen Zugriffs auf verteilte Dokumentenbestände gelöst werden, die hinsichtlich ihrer Erschließungsart und -intensität (RSWK, Thesaurus Sozialwissenschaften, verschiedene Klassifikationen, freie Schlagwörter, strukturierte Begriffsmengen usw.) wie auch in Bezug auf Inhalt und Struktur (reine Sachtitel, Kurzreferate, Volltexte) höchst unterschiedlich sind. Dies ist für den Benutzer ein schwerwiegendes Problem, weil bei der integrierten Suche Konsistenzbrüche entstehen. Es müssen daher intelligente Transferkomponenten zwischen den verschiedenen Formen der Inhaltserschließung hinzukommen. Bei den Sondersammelgebieten ist die Problematik der heterogenen Inhaltserschließung generell besonders kritisch, weil die Grenzen des Gebiets nicht für alle Benutzerwünsche gleichermaßen sinnvoll gezogen werden können. Es gibt immer Überschneidungsbereiche, die das SSG wegen seiner Spezialisierung nicht allein bedienen kann. Der Einstieg über eine Universitätsbibliothek mit einem speziellen SSG wird somit besonders auf eine vom Benutzer möglichst problemlos handhabbare Verbindung zu parallelen, sich überschneidenden und ergänzenden Beständen an anderen Orten angewiesen sein. Hinzu kommt als immer wieder gefordertes Desiderat die exemplarische Einbindung von Mechanismen zur direkten Bestellung und die Ausgabe des Volltextes im Anschluß an eine Suche. Der Benutzer soll in *einem* Suchvorgang möglichst hochqualifizierte Suchinstrumente nutzen und ohne Umwege auf die gefundenen Dokumente zugreifen können. Im Förderzeitraum wird ein – auch räumlich zwischen den Standorten der beteiligten Institutionen – verteiltes System zur integrierten Recherche in den zur Verfügung stehenden Dokumentenbeständen entwickelt und über das Internet (WWW) zugänglich gemacht. Diese Lösung steht dann der Integration weiterer Datenbestände

bzw. Datenquellen (Bibliothekskataloge, Datenbanken, Internetressourcen usw.) - auch fremdsprachlichen - offen.

Das Projekt ist ein Gemeinschaftsprojekt des InformationsZentrum Sozialwissenschaften (IZ), Bonn und des Institut für Soziologie der Technische Universität Darmstadt (TUD). Kooperationspartner sind das Sondersammelgebiet Sozialwissenschaften der Universitäts- und Stadtbibliothek Köln, die Bibliothek der Friedrich-Ebert-Stiftung, Bonn, das Zentrum für Interdisziplinäre Technikforschung der TU Darmstadt, sowie der Westdeutscher Verlag, Wiesbaden und der Leske + Budrich Verlag, Leverkusen-Opladen.

Die Laufzeit des Projekts ist zunächst Mai 1999 bis April 2001. Bei der Antragstellung wurde ein weiteres Jahr geplant, dass der Integration von Internetquellen und Englisch sprachiger Literatur dienen soll.

Träger des Projekts ist die Deutsche Forschungsgemeinschaft (DFG), in dessen Förderungsbereich 'Elektronische Publikationen im Literatur- und Informationsangebot wissenschaftlicher Bibliotheken' es angesiedelt ist. Es zielt auf die Realisierung einer Forderung, die das DFG-Memorandum „Weiterentwicklung der überregionalen Literaturversorgung“ als zukünftige Aufgabe der Sammelschwerpunktbibliotheken nennt: „... daß sie stärker als bisher den wissenschaftlichen Nutzern ... elektronisch verfügbare Informationsquellen vermitteln. Insofern müssen entsprechende Aktivitäten über die jeweils eigenen Sammelschwerpunktbestände hinausgehen und externe Daten- und Informationsressourcen des jeweiligen Faches einbeziehen. So können Spezialbibliotheken, Fachinformationsanbieter ... oft eine intensivere Erschließung fachbezogener Informationen ... leisten ... Mit Hilfe moderner Technologie kann der Zugriff auf solche Datenbanken erleichtert werden.“¹

Das Gesamtprojekt nimmt somit erstmals im sozialwissenschaftlichen Bereich die virtuelle Integration verschiedener Dokumententypen und -inhalte einschließlich des Bestandes des Sondersammelgebietes Sozialwissenschaften in Angriff und zielt damit inhaltlich wie informationstechnisch auf einen neuen Weg zum Umgang mit Informationen für Anbieter und Informationssuchende.

¹ DFG 1998, Kap. 2.1.3

1 Ausgangslage und Vorarbeiten

1.1 Ausgangslage

1.1.1 Generelle Aspekte und BMBF Vorhaben

Daß die „Verfügbarkeit von hochaktuellen und qualitativ hochwertigen wissenschaftlichen und technischen Informationen für Nutzer in Industrie, Wissenschaft und Staat“ (BMBF 1996) der Schlüssel für effizientes Forschungs- und Wirtschaftshandeln ist, kann heute als allgemeiner Grundkonsens zwischen Politik, Forschung und Wirtschaft angesehen werden. Diesen Grundkonsens belegt das Programm der Bundesregierung „Information als Rohstoff für Innovationen (INFO 2000)“.

Das BMBF faßt die Kernaussagen des Programms wie folgt zusammen:

„Die optimale Erschließung und Bereitstellung einmal erarbeiteten Wissens hilft, die Qualität und Effizienz des Forschungs- und Entwicklungsprozesses zu steigern sowie die Verwertung der Forschungs- und Entwicklungsergebnisse für Innovationen in Wissenschaft und Wirtschaft zu stimulieren. ... Die Entwicklungen zur globalen Informationsgesellschaft und Informationsstruktur bewirken grundlegende Veränderungen in Wissenschaft und Technik. Der schnellen und umfassenden Verfügbarkeit von aktuellen Informationen kommt hierbei eine Schlüsselrolle zu. Die hohe Qualität von Forschung und Entwicklung in Deutschland kann nur gesichert werden, wenn Forscher und Entwickler schnell und umfassend Zugang zum weltweit vorhandenen Wissen in ihrem Fachgebiet haben.“²

Traditionelle Produzenten und Mittler für die Zurverfügungstellung von wissenschaftlicher Information sind die Wissenschaftler, die Verlage, die Bibliotheken und die Informationsservicestellen wie das IZ Sozialwissenschaften der GESIS. Ihre Aufgaben blieben im Kern bis vor einigen Jahren unverändert. Was sich heute dramatisch ändert und deutliche Weiterentwicklungen und Anpassungen verlangt, sind die Rahmenbedingungen. Die derzeitige Situation ist durch einen tiefgehenden Wandel in der gesamten zugrundeliegenden Informationstechnologie geprägt, der nicht nur technologische Anpassungen erzwingt, sondern auch neue Formen des Informationsservice eröffnet und andere als die bisher gewohnten

² BMBF 1996, S. 1f.

Formen des Zusammenspiels von Wissenschaftlern, Verlagen, Bibliotheken und Informationsservicestellen nahelegt.

(BMBF 96) gibt eine gute allgemeine Analyse dieser Entwicklungen und belegt sie im Anhang mit den entsprechenden Statistiken, die die Wichtigkeit und Schnelligkeit der bevorstehenden technologischen Umwälzungen aufzeigen. Global geht es vor allem um die folgenden in (BMBF 96) näher erläuterten und begründeten Bereiche:

- Stichwort globale digitale Bibliothek

Wissenschaftler sollen von ihrem Computer aus einen optimalen Zugang zu den weltweit vorhandenen elektronischen und multimedialen Volltext-, Literaturhinweis-, Fakten- und Softwareinformationen haben.³ Auf technischer Seite setzt dies u.a. im Netz zugängliche verteilte Datenbanken voraus, auf konzeptueller Seite die Integration verschiedener Informationsgehalte und -strukturen.

- Stichwort Elektronisches Publizieren

Die weltweiten Netze erlauben es, Publikationen dezentral elektronisch für die weltweite Nutzung bereitzustellen. Dies führt zu einer neuen Arbeitsteilung zwischen Autoren, Verlagen, Buchhandel, Bibliotheken (einschließlich ihrer SSG) und Informationsservicestellen.

- Stichwort Multimedia

Es steht für neue Möglichkeiten der Dokumentengenerierung, -darstellung und -bearbeitung (Integration und Kombination von Grafiken, Bild, Ton etc. und Text).

Diese Grundgedanken teilt das DFG-Memorandum „Weiterentwicklung der überregionalen Literaturversorgung“. Auch hier setzt man auf gemeinsame Anstrengungen aller Beteiligten und auf eine Konzentration der Kräfte:

„Dabei ist der Bedarf sowohl für die fachspezifische als auch für die fachübergreifende Information zu berücksichtigen. Deshalb müssen in

³ Dazu Atkins 1997: „Digital Libraries research ... should provide users with toolkits to overcome the risk of an information surfeit, allowing people to navigate, to make sense, and to use productively increasingly rich and heterogeneous data sources.“ (S. 3)

Deutschland kooperative Verfahren für die Versorgung der Wissenschaft ausgebaut und für die Bildung Virtueller Fachbibliotheken eingesetzt werden, die von den Sondersammelgebietsbibliotheken und Zentralen Fachbibliotheken in Zusammenarbeit mit anderen Einrichtungen initiiert und aufgebaut werden sollten.“⁴

(BMBF 96) nennt auch die in den obigen Bereichen derzeit vom BMBF geförderten und geplanten Projekte, bei denen besonders die Vorhaben der Initiative GLOBAL INFO von Interesse sind.⁵ Die BMBF Projekte haben ihren Schwerpunkt im naturwissenschaftlichen Bereich (z.B. Mathematik, Physik). Die DFG fördert entsprechende Vorhaben vor allem im „Digital Library Program“.⁶ Ein Schwerpunkt bildet der Aufbau von virtuellen Fachbibliotheken für verschiedene Fächer. Ihnen ist die „Virtuelle Fachbibliothek Sozialwissenschaften“ zuzuordnen. Unseres Wissens nach behandelt keine der bisherigen virtuellen Fachbibliotheken systematisch die Heterogenität der Inhaltserschließung über die technologische Integration hinaus, so daß die Erfahrungen aus dem Projekts prototypischen Charakter für die anderen Projekte haben können.

Es wäre ein Irrtum zu glauben, die sozialwissenschaftlichen Fächer bräuchten nur auf den Abschluß der naturwissenschaftlich orientierten Projekte (bzw. auf die aus anderen geisteswissenschaftlichen Fächern der DFG-Projekte) zu warten und sie dann unverändert zu implementieren. Gerade für den sozialwissenschaftlichen Informationsbereich gibt es heute keine Patentrezepte mehr. Die Besonderheiten der sozialwissenschaftlichen Dokumentgrundlage und die der sozialwissenschaftlichen Informationsbedürfnisse müssen bekannt sein und die Grundlage jeder informationstechnologischen Lösung in diesem Bereich bilden. Ein Beispiel hierfür ist die Auseinandersetzung um die „richtige“ Art der Inhaltserschließung von Textdokumenten. Die informationswissenschaftliche Forschung diskutiert sie heute wie vor zwanzig Jahren kontrovers. Das Spektrum der in Anwendung und Forschung vertretenen Modelle reicht von der intellektuellen Indexierung über die automatische Freitexterschließung

⁴ DFG 1998, Kap. 2.3

⁵ siehe <<http://www.global-info.org>>

⁶ siehe Bunzel 1998

(evtl. mit computerlinguistischen Komponenten) bis hin zu quantitativ-statistischen Verfahren und neuronalen Netzen. Sicher scheint heute nur, daß Sinn und Zweck eines gewählten Inhaltserschließungsverfahrens inhaltlich und sprachlich stark vom Anwendungsfeld abhängig sind. Was sich in einem Fachgebiet bewährt, kann in einem anderen zu negativen Ergebnissen führen. So zeigten z.B. Signifikanztests der TREC-Evaluationsstudien zu verschiedenen Verfahren des Information Retrievals, daß trotz einer stetigen Verbesserung der generellen Leistungsfähigkeit der Systeme in den letzten Jahren die Performanz stärker mit den verschiedenen Topics als mit den getesteten Systemen korreliert.⁷

1.1.2 Das Modell polyzentrischer Informationsversorgung

Bei einer näheren Analyse der obengenannten Integrationsprobleme, die im beantragten Projekt am Anwendungsfall der Integration von SSG-Beständen, mehrerer Bibliothekskataloge und Literaturdatenbanken wie SOLIS prototypisch bearbeitet werden sollen, zeigt es sich, daß enge technologische Einzelkonzepte, auch wenn sie fraglos notwendig sind, bei den sich derzeit vollziehenden informationstechnologischen und gesellschaftlich-wirtschaftlichen Entwicklungen allein nicht ausreichen.⁸ Sie sind durch neue konzeptuelle Überlegungen zur Behandlung der Konsistenzbrüche zwischen den verschiedenen Arten der Inhaltserschließung zu ergänzen. Erst die Verbindung beider Aspekte führt zu akzeptablen Lösungen.

1.1.2.1 Allgemeiner Rahmen und technologische Integration

Als zukünftig entscheidendes Kriterium für eine integrative Neustrukturierung des Zusammenspiels von Wissenschaftlern, Fachgesellschaften, Verlagen, Bibliotheken mit ihren SSG und den Informationsservicestellen wird ein deregulativer Ansatz angesehen, der auf

⁷ vgl. Womser-Hacker 1996

⁸ Weitergehende Überlegungen dieser Art werden in Krause 1996 ausführlich dargestellt. Es wurde versucht, nicht nur eine informationswissenschaftlich akzeptable Lösungsstrategie zu finden, sondern die politische und gesellschaftliche Durchsetzbarkeit und die wirtschaftlichen Rahmenbedingungen gleichermaßen zu berücksichtigen.

die systemische Vernetzung von Informationsinhalten und -strukturen hinausläuft.

Ausgangspunkt der Überlegungen für eine Neuordnung des Informationswesens in den Sozialwissenschaften ist die Vernetzung der bereits bestehenden umfangreichen heterogenen Datenbestände, die in sehr unterschiedlicher Weise erschlossen und auf Informationssysteme ungleicher Art verteilt vorliegen. Diese Grundsituation entspricht nicht mehr den heutigen technologischen Möglichkeiten. Unterschiedliche Dokumententypen, Datenstrukturen und Datenbanken sind deshalb zur gemeinsamen Nutzung in einer für Nutzer wie Anbieter transparenten Weise zusammenzuführen. Dies erfolgt sinnvollerweise durch möglichst standardisierte Verfahren für die Datenintegration in verteilten Systemen. Als Standardprotokoll wird hierfür nach gegenwärtigem Stand Z.39.50, evtl. erweitert um andere Abfragesprachenelemente für Datenobjekte, angestrebt. Technisch ist eine solche Zusammenführung als polyzentrisches verteiltes System zu realisieren (cf. Abb. 1). Der Nutzerzugang sollte im Internet innerhalb des WWW erfolgen.

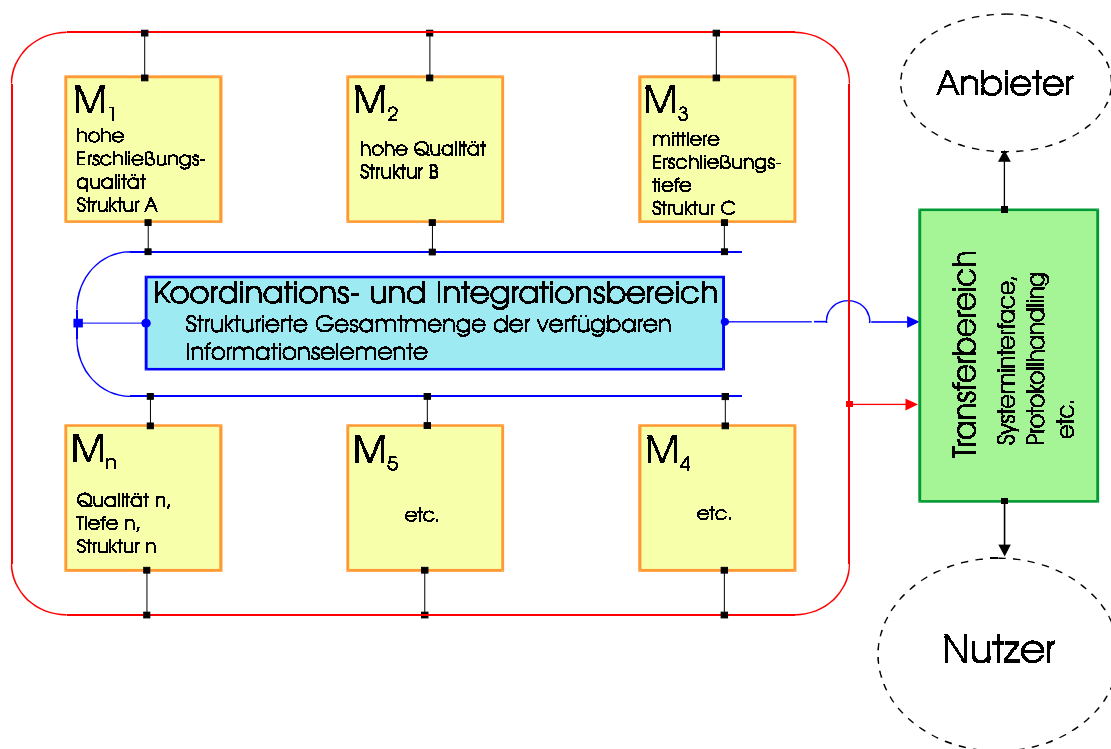


Abbildung 1: Modell polyzentrischer Informationsmengen

1.1.2.2 Inhaltliche Integration

Unter inhaltlichen Gesichtspunkten ist für eine solche polyzentrische Struktur verteilter Informationsbestände charakteristisch, daß es keine eindeutigen Gewichtungskriterien für die Auswahl von Informationsbeständen und deren Bewertung gibt. Was für den einen Nutzer am Rande des Interessensfeldes liegen mag, steht für den anderen Nutzer oder auch eine Teil- oder Nachbardisziplin im Zentrum der Aufmerksamkeit. Wissenschaftliche Fachliteratur einerseits und graue oder populäre Medien aus den umfangreichen praxisbezogenen Anwendungsbereichen und Bezügen von Wissenschaft andererseits können nicht von vornherein in ein hierarchisches Verhältnis der Über- und Unterordnung gesetzt werden; beide Zugänge haben ihre jeweilige Berechtigung, aber ihre differierende Prioritätensetzung. Entsprechendes gilt für die verschiedenen Formen der Inhaltserschließung, derer sich u. a. die Bibliotheken und die Informationsservicestellen bedienen. Auch unter dem Blickwinkel des Benutzerzugangs aus dem Bestand eines SSG heraus kommt man zur gleichen Schlußfolgerung: Klare, für alle Benutzer und Informationsbedürfnisse angebbaren Grenzziehungen gibt es nicht. Es muß daher ein Verfahren gefunden werden, das dem Nutzer eine flexible Verwendung von Schlagwort- und Begriffskombinationen (navigating and browsing) erlaubt und mit der sich zwangsläufig ergebenden Heterogenität auf der Seite der Inhaltserschließung intelligent umgeht.

Die Grundproblematik soll im folgenden an einigen Beispielen verdeutlicht werden.

Gerade bei Nutzern der in SSGen erschlossenen Bestände werden die oben thematisierten Schwierigkeiten besonders deutlich. So erfaßt das SSG Sozialwissenschaften der USB Köln jährlich etwa 3500 sozialwissenschaftliche Monographien und über 1300 laufende Zeitschriften nach den Erschließungsregeln der USB Köln, in deren Katalog der Bestand auch nachgewiesen wird. Ein Teil dieser Literatur ist auch in anderen Bibliothekskatalogen wie dem Darmstädter Virtuellen Gesamtkatalog oder der Bibliothek der Friedrich-Ebert-Stiftung erfaßt, jeweils jedoch in einem anderen Dokumentenkontext. An diesen zusätzlichen Quellen ist der Benutzer interessiert, wenn er in einem Überschneidungsbereich arbeitet, der den Nachweis relevanter Dokumente aus dem Darmstädter Katalog erwarten läßt, oder auch nur, wenn er lokal auf die Darmstädter Bibliotheken einfacher zugreifen kann. Dort werden jedoch eine andere Art der Klassifikation verwendet und andere Schlagwörter vergeben. Will der Benutzer aus dem

Kontext der Bibliotheksrecherche der USB Köln heraus auch sozialwissenschaftliche Zeitschriftenaufsätze oder Beiträge in Sammelwerken nachgewiesen bekommen, muß er hierzu zur Datenbank SOLIS des IZ wechseln, die auch die Monographien des SSG enthält, beides jedoch mit der IZ-eigenen Klassifikation und einer Schlagwortvergabe auf der Grundlage des IZ-Thesaurus. In Teilen ist zudem der Abdeckungsgrad anders. „Arbeitsmarkt und Berufsforschung“ findet sich nur beim IZ und der Friedrich-Ebert-Stiftung, „Unternehmensführung“ beim IZ unter dem Aspekt Betriebssoziologie, an der USB Köln beim SSG Betriebswirtschaft mit wiederum eigener Klassifikation und Schlagwortvergabe. Die graue Literatur, die das IZ erfaßt, ist in ihrer Gesamtheit in der USB Köln nur über eine spezielle Liste zugänglich, jedoch nicht in den OPAC des SSG integriert.

Trotz dieser heterogenen Ausgangslage soll der Benutzer des SSG der USB Köln z.B. nicht gezwungen werden, sich zuerst in das Erschließungssystem der Universität Köln einarbeiten zu müssen, um dann bei einer Erweiterung seiner Suchintention auf unselbständige Publikationen ein zweites System der Inhaltserschließung lernen und in geeignete Suchstrategien umsetzen zu müssen, und eventuell wiederum ein anderes, wenn er weitere Bibliothekskataloge ergänzend durchsuchen möchte. Tut er dies jedoch nicht und benutzt einfach die Suchterme aus der Recherche im SSG weiter, verliert er relevante Dokumente und kann keine vernünftige Suchstrategie mehr aufbauen. Ein vom Benutzer gewähltes Schlagwort A kann in den verschiedenen Dokumentenbeständen die unterschiedlichsten Bedeutungen annehmen. Auch im engen Bereich der Fachinformation kann ein Term A, der aus einem hochrelevanten, mit viel Aufwand qualitativ hochwertig ermittelten Dokumentenbestand stammt, nicht mit dem Term A gleichgesetzt werden, den z. B. eine automatische Indexierung auf der Basis von Titeln aus einem Randgebiet liefert. Deshalb genügt eine rein technologische Verknüpfung verschiedener Dokumentenbestände und die formale Integration unter einer Benutzungsoberfläche allein nicht. Sie führt zu einer Fülle von irrelevanten Treffern. Deshalb ist die Unterstützung der Benutzerrecherche unter Nutzung zusätzlicher inhaltserschließender Elemente, die sich aus der Integration der verschiedenen Bestände ergeben, ein vorrangiges Ziel des Projektes.⁹ Hinzu kommt, daß die Bestände der Grauen Literatur, die in den Sozialwissenschaften für den wissenschaftlichen

⁹ vgl. DFG 1998: Memorandum zur Weiterentwicklung der überregionalen Literaturversorgung, insbesondere 2. u. 3.

Diskurs eine beachtliche Rolle spielen,¹⁰ erstmals vollständig nachgewiesen werden können. Insgesamt ergibt sich eine bessere Erschließung der Bestände des SSG und eine Ergänzung um weitere Dokumentarten und Informationsressourcen.

Beim gegenwärtigen Stand der beteiligten Kataloge bzw. Datenbanken ergibt z.B. eine Recherche im DVK (Darmstädter Virtueller Gesamtkatalog) nach dem Autor „Wolfgang Littek“ in sechs Bereichsbibliotheken (darunter Soziologie) und der LHB Darmstadt mehrere Treffer. Darunter wiederum befindet sich auch der Titel „Einführung in die Arbeits- und Industriesoziologie“, dessen Herausgeber Wolfgang Littek ist. Dieser Titel kommt in den Bereichsbibliotheken Politik, Soziologie, Berufspädagogik und Arbeitswissenschaft sowie in der LHB vor. Mit einer autorenbezogenen Suche läßt sich dieser Titel auch in den Datenbeständen der USB Köln, der Bibliothek der FES und des IZ (SOLIS) auffinden. Führt man nun im DVK eine Suche mit dem Titelstichwort „Arbeitssoziologie“ (wobei gleichzeitig auch das entsprechende Schlagwort gesucht wird), so erhält man zwar (andere) Treffer in verschiedenen angeschlossenen Bibliotheken, aber nur in der Bereichsbibliothek Politik das obengenannte Werk. Warum? Nur in der LHB und in der Bereichsbibliothek Politik werden Schlagwörter verwendet, aber bei der LHB ist dieses Schlagwort offensichtlich nicht vergeben worden. In der Bereichsbibliothek Soziologie usw. werden gar keine Schlagwörter vergeben und im Titel der Monographie ist „Arbeitssoziologie“ nur verkürzt wiedergegeben (als Arbeits[- und Industrie]soziologie), so daß das Buch nicht gefunden werden kann. Auch bei der Bibliothek der FES wird der Titel nicht mit der Schlagwortsuche gefunden, obwohl generell Schlagwörter vergeben werden. In dem OPAC der USB kann der Titel nicht gefunden werden, weil die entsprechenden Bestände noch nicht digitalisiert sind.¹¹ In der Datenbank SOLIS des IZ wird der Titel selbstverständlich mit der Schlagwortsuche gefunden. Hier ist außerdem noch ein Kurzreferat vorhanden.

¹⁰ Vgl. Artus 1992

¹¹ Ein entsprechendes Projekt der USB Köln läuft gegenwärtig, so daß diese Lücke innerhalb der Projektlaufzeit geschlossen sein wird.

Dieses Beispiel zeigt die bestehenden Probleme sehr deutlich: nicht immer werden Schlagwörter vergeben, soweit Schlagwörter vergeben werden, bestehen Inkonsistenzen aufgrund der Vielzahl der beteiligten Institutionen und Personen und der verschiedenen Thesauri. Andererseits bestehen durch das Projekt große Chancen zu einer Verbesserung der Situation zu kommen, da aufgrund der hohen Überschneidung zumindest im Bereich der Monographien unter äußerst günstigen Bedingungen Transferkomponenten entwickelt werden können, die die bestehen Mängel weitgehend ausgleichen können und zwischen verschiedenen Thesauri und Klassifikationssystem vermitteln können, ohne das der Benutzer sich noch im Detail um die einzelnen Systeme kümmern muß.

1.1.3 Lösungsansatz Schalenmodell

Im oben dargestellten Schaubild (Abb. 1) ist das Modell polyzentrischer Informationsmengen skizziert, welches im Projekt als Grundlage für den Aufbau einer polyzentrischen Serverstruktur dienen soll. Generell läßt dieses Modell verschiedene Niveaus der Inhaltserschließung zu, die in einem gemeinsamen virtuellen Informationssystem im Sinne eines Schalenmodells aufeinander bezogen werden.

1.1.3.1 Beispielhafter Modellaufbau und allgemeine Anforderungen

Bei sozialwissenschaftlichen Literaturinformationen wäre z.B. folgender Aufbau denkbar:

- Eine Menge M1 enthält den Kern der als relevant eingeschätzten Literatur. Für die Sozialwissenschaften bilden z.B. der Bestand des SSG und der des IZ den Datenbestand mit gesicherter hoher Relevanz. Idealerweise wird er möglichst tief und qualitativ hochwertig erschlossen. Es kann jedoch auch vorkommen, daß hochrelevante Literatur inhaltlich nicht sehr tiefgehend erschlossen vorliegt. Auf diesen Kern kann weder aus fachwissenschaftlich-dokumentarischen noch aus organisatorischen Gründen verzichtet werden. Nur bei ausreichend hoher Relevanz und angemessenem Umfang dieses Kernbereiches lassen sich weitere Bereiche darum herum auf Dauer konstituieren.
- Die folgenden Mengen, die jeweils andere Informationsmengen repräsentieren, lockern die Relevanzbedingungen und möglichst parallel dazu die Anforderungen an die Qualität der Inhaltserschließung. Für

SOLIS wäre Schale M2 z.B. die Menge an Dokumenten, die zwar nach dem IZ-Thesaurus verschlagwortet, aber ohne Kurzreferat (Abstract) angeboten werden.

- Die Menge M3 könnte die Ansetzung bestehender digitaler Bibliothekskataloge enthalten. Neben den gebundenen Deskriptoren (Beispiele: Autor, Körperschaft) steht für die Inhaltserschließung nur der Titel zur Verfügung, der automatisch indexiert wird; eventuell auch unter Nutzung von Schlagwörtern nach einem anderen Bezugssystem als z.B. dem in Menge 1 angewendeten oder von Schlagwörtern der Deutschen Bibliothek (RSWK) oder von Klassifikationen aus anderen Quellen (DDB, DDC, UDC, Library of Congress, RVK o.a.).
- Eine Menge M4 könnte aus Internetquellen bestehen, die sehr unterschiedliche Qualität haben und deren gegebenenfalls vorhandenen Metadaten sich für die inhaltliche Erschließung nutzen lassen.
- Eine Menge M5 könnte die Informationen aus Inhaltsverzeichnissen von Zeitschriften enthalten (Current Contents), durch die nur der Titel suchbar gemacht wird und noch nicht der direkte Zugriff auf die Volltexte möglich wird.

Wie viele Mengen angesetzt werden und welche Merkmale sie definieren, richtet sich nach den Gegebenheiten eines Fachgebiets und den sich beteiligenden Anbietern. Deshalb bildet die empirisch basierte Ansetzung der Mengen des Schalenmodells und die Festlegung ihrer Eigenschaften einen wichtigen ersten Schritt beim Aufbau einer Virtuellen Bibliothek Sozialwissenschaften. Entscheidend hierbei ist, daß eine Zuordnung von erschlossenen Daten nach den Prinzipien der Relevanz und Erschließungstiefe bzw. -variation erfolgt und daß ein konsistent und tief erschlossenen Informationsbestand im Kernbereich angeboten wird.

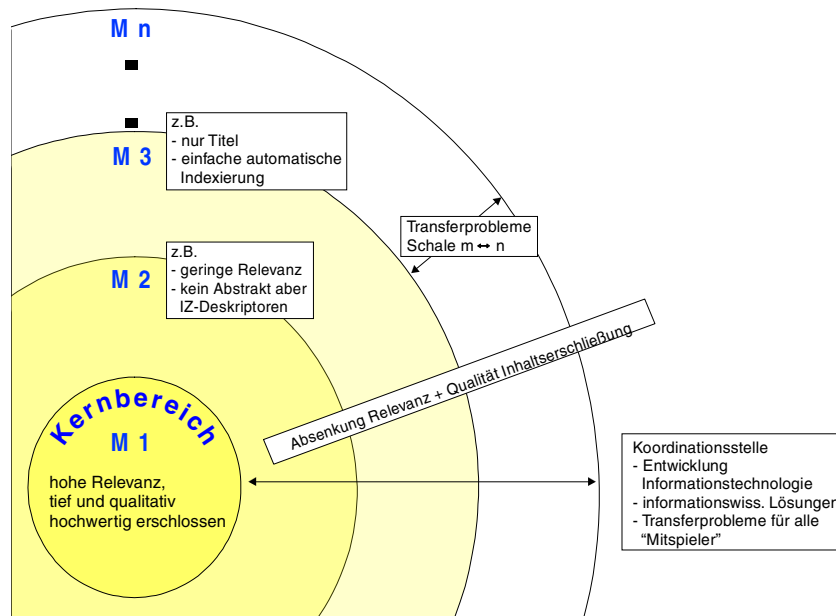


Abbildung 2: Modell verschiedener Erschließungsebenen (Schalenmodell)

An diesem System können sich prinzipiell alle Mitspieler der Sozialwissenschaften beteiligen, unabhängig davon, ob sie viel oder wenig erschließen, ob sie regelmäßig liefern können oder nur schwerpunktmäßig und projektbezogen arbeiten. Damit müßte die Menge der insgesamt 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 noch im veralteten zentralistischen Ansatz zum Ausschluß der Mitspieler. Gegenüber dem Idealmodell einer durchgehend konsistenten Erschließung sinkt auch unter besten Bedingungen die Datenkonsistenz und damit die Recherchequalität. Die so erreichbare Datenfülle gleicht allerdings die Konsistenzbrüche aus, wenn folgende Voraussetzungen gegeben sind:

- Die Koordinationsstelle kümmert sich um offene Systemstrukturen und einen effizienten Informationsfluß. Sie löst, soweit erforderlich, Probleme der Informationstechnik für alle am Verbund beteiligten Partner und befaßt sich mit den neueren informationswissenschaftlichen Erkenntnissen. Dazu gehört die Pflege der Netzstruktur, die fachspezifische Auswahl und Anpassung geeigneter Software für die

Inhaltserschließung und Recherche genauso wie die Entwicklung geeigneter Verfahren zur Erstellung von elektronischen Dokumenten und Printprodukten.

- Bei der Recherche müssen die unterschiedlichen Erschließungstiefen und -strukturen voneinander 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 aufgezwungen werden. Umgekehrt muß aber auch die primär suchfeldbezogene Recherche ohne Rücksichtnahme auf differierende Erschließungstiefen möglich sein; auch eine Orientierung an dieser Erschließungsstruktur darf dem Nutzer nicht aufgezwungen werden.

1.1.3.2 Transferproblematik

Das Schalenmodell ordnet in einem ersten Schritt die verschiedenen Datentypen des Anwendungsfeldes Sozialwissenschaften - wie oben exemplarisch aufgezeigt - nach den unterschiedlichen Niveaus der Datenrelevanz und Inhaltserschließung. Die strukturierte Gesamtmenge der Informationselemente, die damit auf M1-Mn zur Verfügung stehen, enthält insbesondere vielfältige Begriffsstrukturen, die integriert und koordiniert werden müssen: Thesauri, Klassifikationen, eventuell auch thematische Begriffsfelder¹² und Abfragestrukturen begrifflicher Datensysteme usw.

In der Terminologie des Schalenmodells heißt dies, daß das Problem des Übergangs von einer Struktur zur anderen zu lösen ist; es sind die Transfermodule zwischen je zwei benachbarten Schalen zu entwickeln. Deskriptoren einer Teilmenge (Beispiel: nur automatische Titelindeizierung) können nicht gleichwertig mit denen einer anderen Teilmenge (Beispiel: mit Thesaurusbegriffen tiefer erschlossene Bestände) verbunden werden. Ob es gelingt, vernünftige Transferstrategien zu finden und programmtechnisch umzusetzen, entscheidet wesentlich mit über die Effizienz des Modells.¹³

¹² Die Möglichkeit zur Zusammenstellung eigener Begriffsfelder mit ihren jeweiligen inhaltlichen Kontexten ist zu prüfen. Hierbei dürfte der Verwendung von reaktiven wie proaktiven Agenten eine nicht unerhebliche Bedeutung im Kontext einer intentionalen Systemkonzeption im Sinne Dennetts zukommen. Vgl. Shoam 1993 und Dennett 1971.

¹³ vgl. aus Konzeptpapier von Krause/Schwänzl 1998 zu Global Info Slot 3.

Generell gibt es drei Verfahrensweisen, die in Bezug auf ihre Wirksamkeit im Projekt überprüft und exemplarisch implementiert werden sollen.

- Crosskonkordanzen zu Klassifikationen und Thesauri

Die verschiedenen Begriffssysteme werden im Anwendungskontext analysiert und der Versuch gemacht, ihre Begrifflichkeit intellektuell aufeinander zu beziehen. Im Projektkontext konzentrieren wir uns dabei vor allem auf die bibliothekarischen Verfahren der Universitätsbibliothek Köln, die im SSG eingesetzt werden, und auf die von SOLIS. Crosskonkordanzen decken den statisch bleibenden Teil der Transferproblematik ab, der auf der Basis bereits entwickelter Begriffssysteme aufbaut und diese Vorarbeiten nutzt.

Bei der Recherche bieten solche Verzeichnisse die Möglichkeit, Terme des einen Begriffssystems auf der Schale *Mx* auf die eines anderen auf Schale *My* auszuweiten, im einfachsten Fall im Sinne einer Synonymie- oder Ähnlichkeitsrelation, aber auch als deduktive Regelbeziehung.

Im Rahmen von GLOBAL INFO wird ein Arbeitspaket der Universitätsbibliothek Regensburg vorgeschlagen, das neben den in Mathematik und Physik verbreiteten Klassifikationen MSC und PACS auch die Allgemeinklassifikationen DDC und RVK einbezieht. Bei den Thesauri sollen die Schlagwortnormdatei (SWD), der IZ-Thesaurus, der Thesaurus Bildungsforschung und der Thesaurus Pädagogik des DIPF verknüpft werden.

Dies Projekte werden genau beobachtet und die Arbeiten gegenseitig abgestimmt, um Doppelarbeiten zu vermeiden.

- Quantitativ-statistische Ansätze

Das Transferproblem läßt sich allgemein als Vagheitsproblem modellieren. Hierfür sind im Information Retrieval verschiedene Problemlösungsstrategien vorgeschlagen worden (probabilistische Verfahren, Fuzzy Ansätze und neuronale Netze)¹⁴, die sich auf die Transferproblematik anwenden lassen. Verfahren dieser Art benötigen Trainingsdaten, bei denen einzelne Dokumente nach zwei

¹⁴ vgl. Womser-Hacker 1996 und Mandl 1998

Begriffsschemata erschlossen oder bei denen zwei unterschiedliche und unterschiedlich erschlossene Dokumente in Beziehung gesetzt sind. Für das multilinguale IR kann dies z.B. der gleiche Text in zwei Sprachen sein. Die Ausgangssituation für solche Verfahren ist bei der virtuellen Fachbibliothek Sozialwissenschaften besonders günstig, da alle Dokumente des SSG gleichzeitig in SOLIS erfaßt wurden.

Welches Verfahren sich als erfolgversprechend erweist, läßt sich nur empirisch auf der Basis sozialwissenschaftlicher Daten klären, wobei die Evaluation der Text Retrieval Conference (TREC)¹⁵ den allgemeinen Hintergrund bilden werden.

- Qualitativ-deduktive Verfahren

Empirische Untersuchungen am Textmaterial der virtuellen Fachbibliothek Sozialwissenschaften dürften deduktive Zusammenhänge offen legen, die mit Techniken aus dem Bereich der Expertensysteme zu behandeln sind. Deduktive Komponenten finden sich beim Intelligenten Information Retrieval¹⁶, bei intelligenten Recherchesystemen und im Bereich der Expertensysteme, eingeschränkt auch bei OSIRIS¹⁷.

Generell wissen wir noch recht wenig über den richtigen Aufbau eines Schalenmodells und seiner Transferkomponenten. Sicher ist nur, daß sich die Lösungen in verschiedenen Anwendungsfeldern und abhängig von den zu integrierenden Dokumententypen deutlich unterscheiden werden. Deshalb müssen die möglichen Transferkomponenten für die virtuelle Fachbibliothek Sozialwissenschaften systematisch am Textmaterial analysiert und exemplarisch in prototypische Lösungen umgesetzt werden.

¹⁵ Die Text Retrieval Conference (TRE) ist eine Initiative des National Institute for Standards and Technology (NIST) in Githersburg (MD, USA). Siehe Aufruf zu TREC-6: Call for Participation Text Retrieval Conference January 1997-November 1997 <www-nlpir.nist.gov/trec/call97.2> [eingesehen: 30.05.1997]. WWW-Homepage von TREC <<http://trec.nist.gov>>

¹⁶ vgl. Belkin 1996, Ingwersen 1996

¹⁷ vgl. Zillmann 1997

1.1.4 Nutzung bestehender Vorarbeiten und Dokumentquellen

Die Virtuelle Bibliothek Sozialwissenschaften nutzt soweit wie möglich bereits bestehende Dokumentensammlungen mit ihren Inhaltserschließungen und Inhaltserschließungswerkzeugen (Klassifikationen und Thesauri). Generell wird Standardsoftware eingesetzt (Datenbank, Indexierung, Bestellservice), die nur dann durch eigene Komponenten ergänzt wird, wenn dies die Problemstellung des Projekts unvermeidlich macht.

Es ist vorgesehen folgende Datenbestände zu integrieren:

- Bestand des SSG Sozialwissenschaften, der nach den Regeln der Universitätsbibliothek Köln erfaßt ist (Monographien),
- Bestand SOLIS des IZ Bonn, der zusätzlich Zeitschriftenaufsätze und Beiträge in Sammelwerken enthält,
- Bestand Bibliothek Friedrich-Ebert-Stiftung, der zusätzlich Volltexte enthält,
- Bestand des Darmstädter Virtuellen Gesamtkatalogs (DVK, etwa 60 Bibliotheken, nur Monographien).

Im dritten Jahr sollen weitere Quellen wie Internetdokumente und Daten des Clearinghouse Sozialwissenschaften des IZ sowie englischsprachige Dokumentquellen hinzukommen.

Für den programmtechnischen Aufbau eines polyzentrischen Systems heterogener Informationsmengen kann auf die Erfahrungen mit dem DVK zurückgegriffen werden. Die Allegro-Datenbanken des DVK eignen sich jedoch nicht zur direkten programmtechnischen Übernahme. Geplant ist, die Virtuelle Fachbibliothek Sozialwissenschaft als verteilte ORACLE-Datenbank neu zu implementieren, was die Performanz und den Funktionsumfang erweitert (cf. Abschnitt 2.3). Alle Dokumentenbestände sind somit nach Oracle zu transformieren und das Gesamtsystem neu aufzubauen. Auf dieser Basis sind die inhaltlichen Transferkomponenten zu entwickeln, die bisher auch für den DVK noch nicht vorliegen, sowie die weiteren Leistungsmerkmale wie Bestellservice und Volltextausgabe zu integrieren.

1.2 Stand von Wissenschaft und Technik

Neben den in Abschnitt 2.1 genannten Problemstellungen sind vor allem die folgenden Aktivitäten relevant.

DBV-OSI, Z39.50

DBV-OSI¹⁸ ist ein gemeinsam von DFG und BMBF gefördertes Digital Library Projekt, das Bibliotheksverbünde (DBI, SWB, BVB, BRZN) und Fachinformationsdatenbanken (FIZ, DIMDI) in einem virtuellen Netzwerk auf der Basis des SR/Z39.50 - Standards integriert. Die in diesem Projekt eingesetzte Z39.50 Serviceprovider-Software wurde von Crossnet Systems Limited entwickelt und ist als Freeware öffentlich zugänglich. Unter <http://z3950gw.dbf.ddb.de/> ist der Zugang zu dem überregionalen Verbundsystem über einen WWW-Gateway möglich. In der zweiten Projektphase (DBV-OSI II) wurde ein Dokumentenliefersystem entwickelt, das auch von den Lieferbibliotheken der SUBITO-Initiative eingesetzt wird. Im Frühjahr 1995 genehmigte die DFG die Weiterführung des Projektes „Praktischer Einsatz einer SR-Testumgebung“, die Interoperabilitätstests zwischen verteilten Z39.50-Anwendungen ermöglicht.

Allegro Z39.50-Gateway¹⁹

Seit Mitte 1996 läuft an der Universität Braunschweig ein DFG-Projekt zur Schnittstellen-Erweiterung des Bibliotheksystems Allegro. Insbesondere konzentrierten sich die Arbeiten auf einen WWW-Gateway für Allegro-Datenbanksysteme (avanti) auf der Basis von C++ Klassenbibliotheken und Perl-Scripten sowie eine Z39.50 Schnittstelle. Plattformunabhängige Programme auf der Basis von Java mit Client-Funktionen zum Zugriff auf Allegro-Datenbanken sind Projekte der nahen Zukunft. Verwendete Metadatenstandards sind MAB, MARC oder ISBD-Aufbereitungen. Zu Fragen der Broker-Funktionalität, Billing-Algorithmen oder Navigation in Volltexten bietet dieses Projekt jedoch keine Perspektive.

¹⁸ vgl. http://www.dbi-berlin.de/projekte/d_lib/einzproj/dbvosi/dbvosi01.htm >

¹⁹ vgl. http://www.dbi-berlin.de/projekte/d_lib/einzproj/z3950/z3950.htm>

MeDoc und GLOBAL INFO

MeDoc²⁰ (Multimediale elektronische Dokumente) war ein vom BMFT bis Ende 1997 gefördertes Schwerpunktprojekt der Informatik, das im Rahmen von Global Info unter dem Titel Interdoc weiter ausgewertet wird. Mit MeDoc wurde ein Prototyp eines Informationsdienstes (elektronische Bibliothek) mit heterogenen verteilten Datenbeständen im Internet aufgebaut und erprobt. In der Regel wurden Volltexte (wissenschaftliche Zeitschriften und Bücher) aber auch Nachweisdatenbanken und Preprint-Server (NCSTRL) in den MeDoc Dienst eingebunden. Der Informationsvermittlungsdienst (Broker) von MeDoc verfügt über Metainformation von den ihm bekannten Informationsquellen (Providern) und schlägt einem Nutzer aufgrund dieser Metainformation für jede Anfrage dynamisch eine Auswahl einschlägiger Informationsbestände vor. Der Informationsvermittlungsdienst fragt parallel jeden dieser Bestände in dessen jeweiliger Anfragesprache ab, bündelt die Ergebnisse und präsentiert sie dem Nutzer. Die Integration zusätzlicher Datenbanken ist prinzipiell möglich, allerdings sind Fragen der Anpassung / Übersetzung von Metadaten, Abfragesprachen und Rankingalgorithmen von den Entwicklern in einem aufwendigen Prozeß von Fall zu Fall neu zu klären.

Neben der Auswertung der Ergebnisse aus MeDoc ist die zukünftige Entwicklung von GLOBAL für die Virtuelle Fachbibliothek Sozialwissenschaften von Interesse.²¹

Stanford Digital Libraries Project (Stanford Infobus²², StARTS²³)

Das Stanford University Digital Libraries Project ist einer der 6 Teilnehmer an der Digital Library Initiative der USA und konzentriert sich auf Fragen, die die Interoperabilität bei verteilten, heterogenen Informationsdiensten betreffen. Der in diesem Projekt entwickelte Infobus sieht eine Erweiterung des Internet Protokolls (TCP/IP) um eine Reihe von höherwertigen

²⁰ vgl. <<http://medoc.informatik.tu-muenchen.de>>

²¹ vgl. Abschnitt 2.1. Der gegenwärtige Stand der Vorschläge findet sich unter <<http://www.global.info.org>>.

²² vgl. Roscheisen et.al. 1997

²³ vgl. <<http://www-diglib.stanford.edu/diglib/WP/PUBLIC/DOC104.html>>

Protokollen (Service-Layers) vor, die die Interoperabilität zwischen den verschiedensten Informationsanbietern bzgl. Metadaten, Anfragesprachen, Abrechnungsmodellen und Urheberrechten sicherstellen sollen. Insbesondere der StARtS-Protokollvorschlag behandelt Probleme des Trading/Brokering, der Anfrageübersetzung und des Rankings bzw. Sortings der zusammengeführten Ergebnismengen. Wichtige Paradigmen dieses Projekts sind Erweiterbarkeit, Ausfalltoleranz und die Vermeidung unnötiger Komplexität. Die Verwendbarkeit verbreiteter Metadatenstandards (USMARC, Dublin Core), Retrieval-Systeme (FULCRUM, WAIS, Harvest), Architekturen für objektorientierte verteilte Systeme (CORBA/IDL, DCOM), Protokolle (TCP/IP, HTTP, Z39.50, telnet) und Austauschformate (Harvest/SOIF) ist sichergestellt bzw. Bestandteil erster Prototypen.

SSG-Informationssysteme im Internet

Das Projekt²⁴ sammelt Internetressourcen zunächst in einigen für das Projekt modellhaft ausgewählten Sondersammelgebieten (Reine Mathematik; Astronomie; Geowissenschaften; Geographie und thematische Karten; Sprache, Literatur, Geschichte und Verfassung sowie Bibliotheks- und Informationswesen des anglo-amerikanischen Raumes). Im Mittelpunkt stehen dabei die qualitative Bewertung der elektronischen Fachinformationen, die Bereitstellung von Zugriffen auf elektronische Dokumente durch das Sammeln von Adressen und deren intellektuelle Erschließung über eine HTML-Datei, die lokale Speicherung, Erschließung und die dauerhafte Archivierung (Langzeitsicherung) dieser Dokumente. Bei den für das Projekt ausgewählten Göttinger SSG-Fächern sollen Hypertext-Informationsseiten erstellt werden, die wichtige Daten und Hinweise zum jeweiligen Sondersammelgebiet enthalten. Beabsichtigt ist zudem eine Vernetzung mit dem Sondersammelgebiets-WWW-Server an der Staats- und Universitätsbibliothek Hamburg, über den Überblicksinformationen zu allen SSG-Bibliotheken und -Fachgebieten über das Internet zugänglich gemacht werden sollen.

²⁴ vgl. <http://www.dbi-berlin.de/projekte/d_lib/einzproj/ssg_is/ssg_is.htm>

1.3 Ausgangspunkte der Realisierung des Projekts

1.3.1 Der Darmstädter Virtuelle Gesamtkatalog (DVK)

In den Jahren 1995 und 1996 wurde an der Technischen Universität Darmstadt von einer Arbeitsgruppe am Institut für Soziologie der „Darmstädter Virtuelle Gesamtkatalog - DVK“, ein WWW-basierter OPAC für die wissenschaftlichen Bibliotheken an der TU Darmstadt, entwickelt (vgl. Schmiede/Wenzel 1996, 1997). In dem System werden gegenwärtig die Bestände aller dezentralen Bibliotheken der TU Darmstadt, die im Verbundsystem Hebis unter Pica elektronisch nachgewiesenen Bestände der LHB Darmstadt sowie einiger anderer Einrichtungen (u.a. öffentlicher Bibliotheken) in der Region Darmstadt nachgewiesen. Insgesamt handelt es sich um gut 60 Bibliotheken und Bestände von rund 600.000 Titelnachweisen. Es ist möglich, kombinierte Suchen in verschiedenen Datenbeständen gleichzeitig durchzuführen, sei es die Suche über alle Bibliotheken eines oder mehrerer Fachbereiche oder die simultane Abfrage aller Datenbestände (virtueller Gesamtkatalog). Im Durchschnitt wird der DVK mit etwa 4.000 Anfragen pro Arbeitstag genutzt.

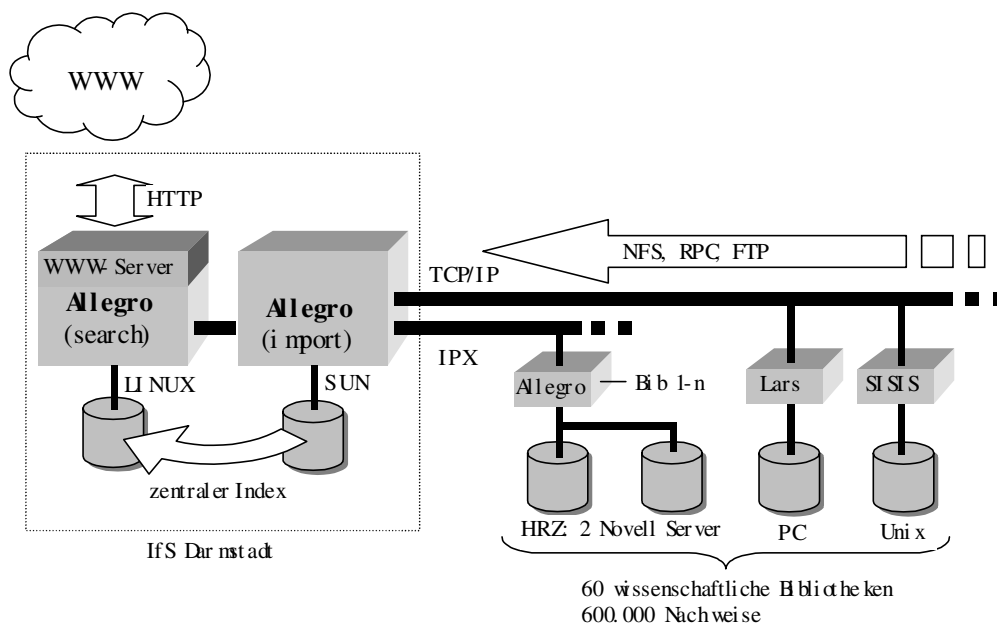


Abbildung 3: Systemarchitektur des Darmstädter Virtuellen Gesamtkataloges (DVK)²⁵

²⁵ vgl. Schmiede/Wenzel 1996, 1997

Kern des DVK ist das Bibliothekssystem Allegro. Es wird auf zwei Rechnern verteilt betrieben, die im Zusammenspiel die Gesamtfunktionalität des Informationsdienstes gewährleisten. Der Zugriff auf die Bestände der verschiedenen beteiligten, die Datenkonversion, die automatische Generierung der Indextabellen sowie die Aufbereitung der Nutzungsstatistik erfolgen auf dem Integrations Rechner. Dieser übergibt die Ergebnisse dann an den WWW-Server, wo sie für einen definierten Zeitraum temporär zwischengespeichert und zur Abfrage genutzt werden. Diese Konfiguration erhöht nicht nur die Performanz des Gesamtsystems, sie reduziert darüber hinaus vor allem seine Störanfälligkeit bei gelegentlichen Zugriffsproblemen zu den Daten der beteiligten Bibliotheken bzw. bei sonstigen Netzstörungen, da das WWW-Angebot bis auf weiteres zunächst unbehelligt weiterlaufen kann, während der Arbeitsserver an der Aufbereitung der benötigten Informationen bzw. an der Wiederherstellung der Datenkonsistenz arbeitet.

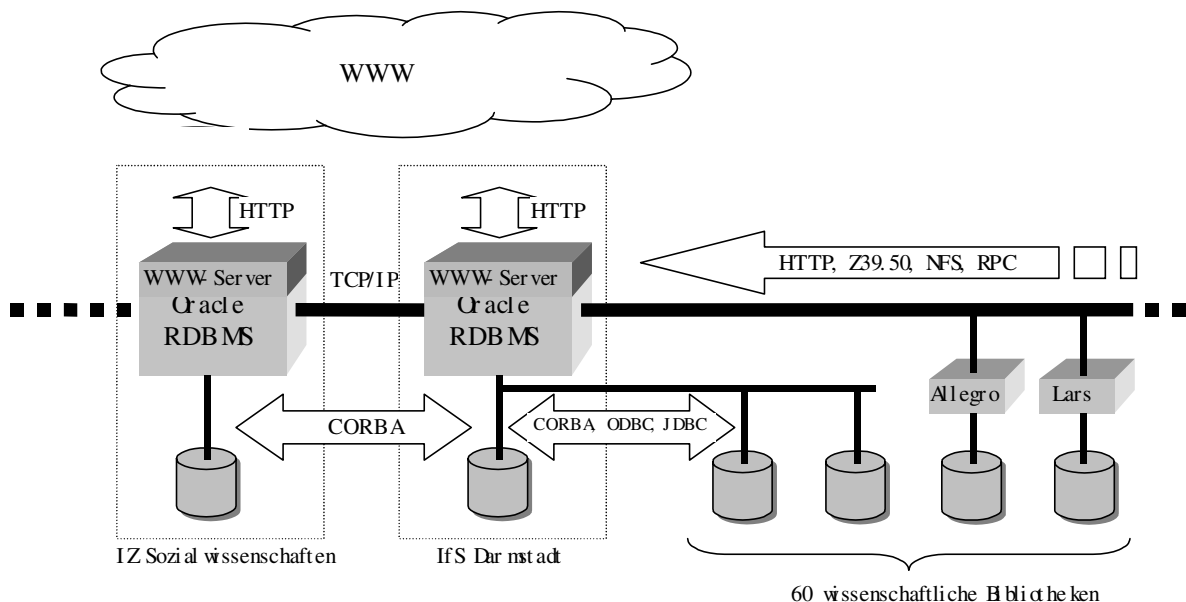


Abbildung 4: Mögliche Systemarchitektur der Virtuellen Fachbibliothek Sozialwissenschaften

Für den Ausbau des Virtuellen Gesamtkatalogs, insbesondere was die Erweiterungen im Rahmen der Virtuellen Fachbibliothek Sozialwissenschaften angeht, muß das bestehende System auf eine neue, zukunftssichere Basis gestellt werden. Die Architektur muß vor allem auch in Richtung auf ein gutes Zusammenspiel mit dem IZ Sozialwissenschaften ausgerichtet sein und eine gute Skalierbarkeit besitzen. Ein wichtiger Schritt in diese Richtung wäre die Verwendung eines leistungsfähigen

Datenbanksystems, das in seiner Performanz wie in seinem Funktionsumfang deutlich über die bislang verwendete Allegro-Datenbank hinausgeht und insbesondere SQL-Abfragen erlaubt. Aufgrund von Studien, die am IZ Sozialwissenschaften zum Vergleich relationaler Datenbanksysteme angestellt wurden, sowie der positiven Evaluation im Rahmen von MeDoc haben wir uns für den Marktführer Oracle entschieden (vgl. Mutschke 1995, Papenfuß 1997).

Darüber hinaus soll eine Schnittstelle nach Z39.50 implementiert werden, um die Kommunikation mit anderen Informationsdiensten im internationalen Forschungs- und Kooperationsraum zu verbessern. Dies ist besonders auch im Hinblick auf die absehbaren Entwicklungstendenzen dokumentenbezogener Informationssysteme notwendig, welche zukünftig sich nicht mehr bloß mit dem Nachweis bibliographischer Daten werden begnügen können, sondern zunehmend vor die Anforderung gestellt werden, die Bereitstellung und sachliche Erschließung multimedialer Dokumente zu realisieren. Die Entwicklung einer entsprechenden Konzeption sowie der Ausbau des bestehenden Systems zu einem universalen Informationserschließungs- und Retrievalsystem sind Bestandteil des vorgeschlagenen Projekts; die Basis dafür ist mit dem bestehenden Informationssystem gelegt.

1.3.2 TOSCANA – Erkundungssystem zur Literatursuche

In den vergangenen Jahren wurde von der Arbeitsgruppe Begriffliche Wissensverarbeitung an der TU Darmstadt ein Modell der begrifflichen Erschließung erarbeitet und in Form des Softwaresystems TOSCANA funktionsfähig implementiert. 1996 wurde es mit einem eigens inhaltlich erschlossenen Literaturbestand verbunden: Seit Januar 1997 steht den Nutzern der Bibliothek des Darmstädter „Zentrums für Interdisziplinäre Technikforschung“ (ZIT) ein TOSCANA-Erkundungssystem für die Literatursuche zur Verfügung.

Die inhaltliche Information über die Bücher der ZIT-Bibliothek ist in einem großen Datensatz gespeichert, in dem 1556 Bücher mit 377 Schlagwörtern über ca. 50.000 Zuweisungen verbunden sind. Um auf diese Informationen zugreifen zu können, wurde der Datensatz auf einem DOS-Rechner unter MS-Windows 3.1 mit dem relationalen Datenbankmanagementsystem MS-Access zugänglich gemacht. Auch die 137 begrifflichen Suchstrukturen mit ihren Liniendiagrammen wurden über ANACONDA-Dateien funktionsgerecht eingebunden. Zu einem *begrifflichen Erkundungssystem*

werden diese Dateien durch das sie verbindende Programm TOSCANA, mit dem die begrifflichen Strukturen aufgerufen, kombiniert und zweckbezogen dargestellt werden können. TOSCANA läuft ebenfalls unter MS-Windows 3.1, ist objektorientiert in C++ programmiert und benutzt die Structured Query Language (SQL) zur Abfrage von Datenbanken. Mit dem Programm TOSCANA können allgemein gespeicherte Daten begrifflich strukturiert und anhand von begrifflichen Darstellungen und Navigationen untersucht und erkundet werden.

Das an der TU Darmstadt entwickelte Programmsystem TOSCANA wird seit 1994 von der Firma „NaviCon – Gesellschaft für begriffliche Wissensverarbeitung mbH“ auch im kommerziellen Bereich zur begrifflichen Analyse und Erkundung von Daten eingesetzt.

1.3.3 Die sozialwissenschaftlichen Datenbanken und informationswissenschaftliche Entwicklungen am IZ Sozialwissenschaften

1.3.3.1 FORIS und SOLIS

Für die *Scientific Community* und die Akteure in der Politik, Verwaltung und den Medien bieten die Datenbanken zu sozialwissenschaftlichen Forschungsarbeiten und Veröffentlichungen ein reichhaltiges Informationsreservoir.²⁶

FORIS - die Datenbank zu sozialwissenschaftlichen Forschungsarbeiten

Die Informationen über die Forschungsaktivitäten basieren im wesentlichen auf den Erhebungen, die das IZ jährlich in der **Bundesrepublik Deutschland**, SOWIS (Sozialwissenschaftliche Informationsstelle der Universitätsbibliothek der Wirtschaftsuniversität Wien) in **Österreich** und SIDOS (Schweizerischer Informations- und Daten-Archivdienst für die Sozialwissenschaften) in der **Schweiz** durchführen. Die **postalischen Befragungen** werden in der Regel im 4. Quartal gestartet und die Ergebnisse in den Folgemonaten in die Datenbank eingearbeitet. 1996 erbrachten sie rund 80% der insgesamt 6.317 Forschungsnachweise. 20% resultierten aus

²⁶ Der folgende Text stützt sich weitgehend auf Nase 1996; einige Angaben wurden aktualisiert.

Kooperationen mit dem Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit, mit der Bundesforschungsanstalt für Landeskunde und Raumordnung, der Zentralstelle für Agrardokumentation und -information, der IWK-Redaktion der Historischen Kommission zu Berlin, der Kommission für die Erforschung des sozialen und politischen Wandels in den neuen Bundesländern und **ergänzenden Auswertungen von Forschungs- und Jahresberichten** zentraler Fördereinrichtungen und Stiftungen. Solche flankierenden Maßnahmen sind bei Erhebungen erforderlich, wenn man ein möglichst vollständiges Bild der Forschungslandschaft zeichnen will. Nur so kann das IZ beispielsweise sicherstellen, daß alle von der DFG geförderten sozialwissenschaftlichen Forschungsarbeiten in FORIS erfaßt sind. Im Jahre 1995 konnten auch erstmals Projektinformationen aus dem Internet übernommen werden; inzwischen sind dies mehrere Hundert. Seit 1996 steht ein **Internet-Formular** zur Eingabe von Forschungsinformationen bereit. Seit 1997 wird diese Möglichkeit der direkten Meldung eines Forschungsprojektes durch die Bereitstellung der drei neuesten Jahrgänge aus **FORIS im Internet** (<http://www.bonn.iz-soz.de/information/databases/foris/foris-search.htm>) unterstützt.

SOLIS - die Datenbank zu sozialwissenschaftlicher Literatur

1996 wurden in SOLIS 12.243 Literaturnachweise neu aufgenommen. 60% davon hat das IZ selbst erstellt, 40% wurden von Partnereinrichtungen, mit denen das IZ seit vielen Jahren zusammenarbeitet, geliefert.

Der **Gesamtbestand** von SOLIS umfaßte 1997 **über 220.000 Dokumente**.

In SOLIS wird „selbständige“ und „unselbständige“ Literatur erfaßt; d.h. **Monographien, Sammelwerke**, aber auch **Beiträge** in den Sammelwerken und **Aufsätze** in wissenschaftlichen Zeitschriften. Dies ist - zusammen mit der inhaltlichen Erschließung mittels Abstracts - ein entscheidender Unterschied zu den Verfahren der Titelaufnahme in Bibliotheken. Außerdem wird in SOLIS auch **Graue Literatur** aufgenommen, z.B. die Reihen des WZB und anderer Institute oder Berichte, die dem IZ im Kontext der Forschungserhebung zugeschickt werden. Diese Graue Literatur wird anschließend über den Leihverkehr der **Universitäts- und Stadtbibliothek Köln**, die das **Sondersammelgebiet Sozialwissenschaften** betreut, der Öffentlichkeit zur Verfügung gestellt. Die Verteilung der Literaturarten bleibt über die Jahre relativ konstant. Der Anteil der Grauen Literatur liegt seit

1991 etwas über dem Durchschnittswert des SOLIS-Bestandes, da im Rahmen der vom Wissenschaftsrat empfohlenen Bestandssicherung bisher nicht veröffentlichte Forschungsberichte der ehemaligen DDR aus Archiven beschafft und ausgewertet wurden.

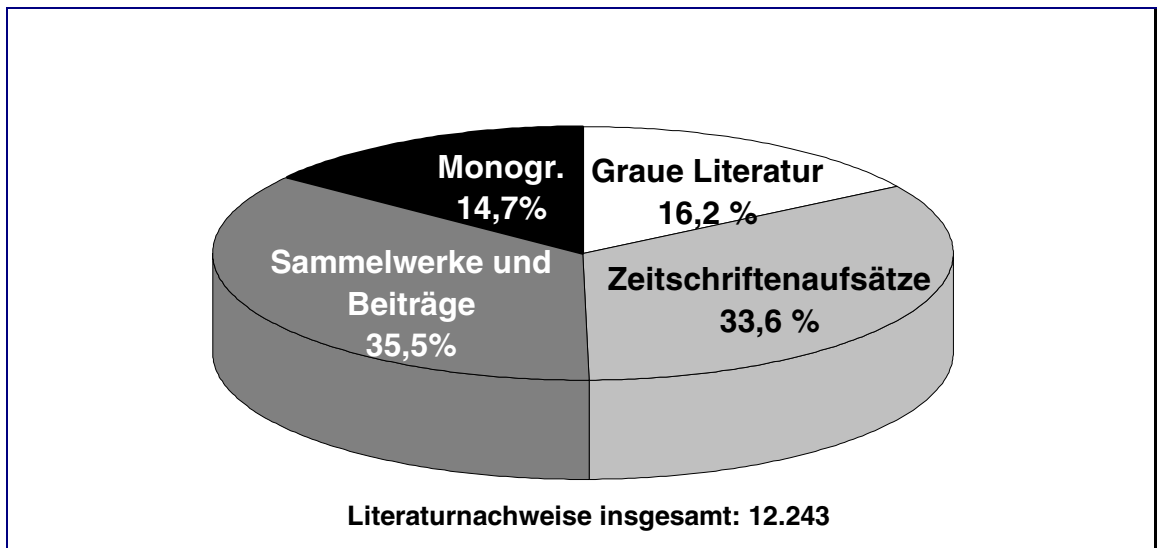


Abbildung 5: Anteile der Literaturarten für den Zugang 1996 zur Datenbank SOLIS

1.3.3.2 Testdatenbank zur automatischen Indexierung (Projekt GIRT)

Das Projekt **German Indexing and Retrieval Testdatabase (GIRT)** soll den Rahmen für einen aussagefähigen Vergleich moderner, intelligenter Indexierungs- und Retrievalsysteme schaffen, auf dessen Basis die Leistungsfähigkeit dieser Systeme gegenüber herkömmlichen Standardsystemen beurteilt werden kann.²⁷

Aus bibliothekarischer Sicht wird die Situation folgendermaßen eingeschätzt: „Es ist keine Frage, daß maschinelle Indexierung kein vollständiger Ersatz für intellektuelle Erschließungstätigkeit, insbesondere nicht der Vokabularpflege, sein kann. Maschinell Indexierung kann jedoch eine wesentliche Ergänzung für die Erschließung bibliographischer Daten und den

²⁷ vgl. zu solchen Vergleichen Womser-Hacker 1996, S. 19.

Aufbau sachlicher Abfragekomponenten bedeuten. (...) Verfahren maschineller Indexierung können nicht alle Probleme lösen, die aus den einschlägigen Benutzerstudien als Unzulänglichkeiten des sachliche Retrievals berichtet werden. Sie können jedoch einen wichtigen Beitrag zur Verbesserung der Datenbasis schaffen. Dies kann Raum schaffen, sich bei den intellektuellen Tätigkeiten auf Bereiche zu beschränken, die entweder durch maschinelle Verfahren nicht abgedeckt werden können oder wesentliche Voraussetzung für hochwertige maschinelle Indexierung bedeuten.“²⁸

Bei GIRT werden Datenbestände, die aus dem Bereich der Fachinformation stammen und deutschsprachig sind²⁹, für den Test intelligenter Indexierungs- und Retrievalsysteme zur Verfügung gestellt. Dabei handelt es sich um Auszüge aus den Datenbanken FORIS (Forschungsinformationssystem Sozialwissenschaften) und SOLIS (Sozialwissenschaftliches Literaturinformationssystem) des IZ, die in einer integrierten Testdatenbank namens SOLFOR zur Verfügung gestellt werden³⁰. Insgesamt enthält die Testdatenbank ca. 13.000 Dokumente. Damit kann hinsichtlich der Mehrfachindexierung von Dokumenten, die

- einerseits durch automatische Indexierung, die von den Testsystemen vorgenommen wird,
- andererseits durch vorab intellektuell (in diesem Fall vom IZ) durchgeführte Indexierung geschieht,

²⁸ Gödert/Lepsky 1997, S. 14

²⁹ Zu den besonderen Probleme der deutschen Sprache in diesem Zusammenhang siehe Womser-Hacker 1996, S. 104.

³⁰ Thematisch stammen die Dokumente aus den für den gedruckten sozialwissenschaftlichen Fachinformationsdienst (soFid) mit den Themen „Industrie- und Betriebssoziologie“, „Frauenforschung“ und „Migration und ethnische Minderheiten“ ausgewählten Dokumentmengen der Erscheinungsjahrgänge 1990 bis 1996 und aus den Heften der Zeitschriften „Kölner Zeitschrift für Soziologie und Sozialpsychologie“ sowie „Soziale Welt“ dieser Jahrgänge.

getestet werden, ob eine solche Mehrfachindexierung bei bestimmten Testsystemen zu einer nennenswerten Steigerung der Leistungsfähigkeit führt³¹.

Vor den eigentlichen Tests wurde vorab ein Pretest durchgeführt, der vornehmlich der Erprobung der Versuchsanordnung und der Meß- und Analyseverfahren diente³². Ferner werden vom IZ und der Universität Konstanz weitere Tests mit verschiedenen Indexierungs- und Retrievalsystemen durchgeführt.

Die besondere Bedeutung von Fachsprachen stellt entsprechende Anforderungen an die Indexierungs- und Retrievalsysteme. Neuere Forschungen zeigen die besondere Schwierigkeit, Fachbegriffe der Soziologie von allgemeinsprachlichen Begriff zu differenzieren, nochmals deutlich auf: „its [sociology] words are common words that are in general use such as community and immigrant“³³. Es gibt neben der hohen Überschneidung der Fachbegriffe der Sozialwissenschaften mit Wörtern der Allgemeinsprache in vielen Fällen eine deutliche Abweichung der Bedeutung der Fachbegriffe von der allgemeinsprachlichen Verwendung und starke Bedeutungsdifferenzen der Fachbegriffe selbst aufgrund inhaltlicher Konnotationen wie „Schulen“, Theorien, politische Implikationen, ethische Grundüberzeugungen usw., die eine Verwendung automatisch generierter Schlagwörter für die Indexierung und die Recherche erschweren.

³¹ Entsprechende Hinweise auf eine Verbesserung der Leistung durch Einbeziehung der intellektuellen Indexierung finden sich bei TREC-4: „.. the manually indexed fields were included in the test documents, though not in the learning documents. As can be seen, there is a 5% to 6% improvement ..., which is reasonably substantial.“ (Buckley/Singhal/Mitra/Salton 1996, S.10).

³² siehe dazu Frisch/Kluck 1997; Workshop: Testverfahren für intelligente Indexierungs- und Retrievalsysteme anhand deutschsprachiger sozialwissenschaftlicher Fachinformation (GIRT), September 1997, Bonn; Kluck 1998

³³ Haas 1997, S. 78; siehe dort auch auf S. 74: „T tests between discipline pairs showed that physics, electrical engineering, and biology had significantly more domain terms in sequences than history, psychology, and sociology (...) the domains with more term sequences are those which may be considered the hard sciences, while those with more isolated domain terms tend to be the social sciences and humanities.“

1.3.3.3 Projekt ELVIRA II

Erste Erfahrungen mit der Text-Fakten-Integration, die einen Aspekt der Behandlung der Heterogenität darstellt, konnten im Projekt ELVIRA (Elektronisches Verbands-, Informations-, Retrieval- und Analysesystem) gesammelt werden. Die Inhaltserschließung der Text- und Faktendaten in den von ELVIRA abgedeckten heterogenen Bereichen unterscheiden sich stark und es wird jeweils verschiedenes Vokabular genutzt. Zwischen den verschiedenen Datentypen müssen geeignete Übergänge durch Transformationen geschaffen werden. Die Transformationen können nicht allein auf der Basis deduktiver Verfahren implementiert werden. Zusätzlich müssen vage Verfahren (z.B. neuronale Netze), wie sie im Information Retrieval bereits erfolgreich eingesetzt werden, für diese Aufgabe adaptiert werden.

1.3.4 Das SSG Sozialwissenschaften der Universitäts- und Stadtbibliothek Köln

Im Sondersammelgebiet Sozialwissenschaften wird die einschlägige monographische sozialwissenschaftliche Literatur gesammelt und für die Ausleihe zur Verfügung gestellt. Der Bestand sozialwissenschaftlicher Literatur umfaßt gegenwärtig über 1.300 laufend gehaltene Zeitschriften und einen jährlichen Zugang von über 3.500 sozialwissenschaftlichen Monographien. Die Integration der Inhaltserschließung des IZ in einen virtuellen Gesamtkatalog sowohl für die Graue Literatur als auch für die sozialwissenschaftlichen Monographien würde eine deutliche Verbesserung der Recherchemöglichkeiten bieten und auch die Nutzung der Bestände erhöhen. Seit 1985 besteht die Möglichkeit, im Wege der Direktbestellung Aufsatzkopien zu bestellen (Literaturdienst); das Bestellverfahren wird auch im WWW angeboten. Daneben gibt es auch die Verfahren des document ordering und des document delivery im Rahmen von JASON und SUBITO, an denen die USB teilnimmt. Alle Bestellverfahren sind auch über den SSG-Server WEBIS zugänglich. Eine Teilnahme am Datenbankservice des Deutschen Bibliotheksinstituts „DBI-LINK“, der ebenfalls ein document order System umfaßt wird angestrebt. Ebenfalls über WEBIS können Rechercheaufträge für einige DFG-finanzierte CD-ROM-Datenbanken erteilt werden, die aus lizenzrechtlichen Gründen nicht zur Nutzung im Internet bereitgestellt werden können.

Die Zukunftsaufgabe der Bereitstellung von Zugriffsmöglichkeiten auf externe Informationsressourcen soll durch die vorgesehene Integration von Datenbankinformationen, lokalen OPACs, elektronischen Volltexten und Internetressourcen in ein einheitliches Angebot angegangen werden. Dabei sollen unter dem Gesichtspunkt der Profilierung des Sammelschwerpunkts Sozialwissenschaften sowohl die Vollständigkeit als auch die Flexibilität des Angebots verbessert werden. Über die bisherigen Bestandsinformationen hinaus, die eine RSWK-Verschlagwortung beinhalten, sollen durch die Koppelung mit anders bzw. tiefer erschlossenen und mit Kurzreferaten oder Volltexten versehenen Nachweisen die Nutzerrecherchen wesentlich unterstützt werden. Gleichzeitig soll die Verknüpfung der Nachweisfunktion mit den Bestell- und Lieferkomponenten realisiert werden.

1.3.5 Die Bibliothek der Friedrich-Ebert-Stiftung

Die Bibliothek der Friedrich-Ebert-Stiftung (FES) sammelt schwerpunktmäßig Literatur (Primär- und Sekundärliteratur) zu den Gebieten Arbeiterbewegung, Entwicklungspolitik, Sozialgeschichte, Parteien und Gewerkschaften und hat einen Bestand von über 500.000 Bänden. Damit zählt sie zu den größten historisch-sozialwissenschaftlichen Spezialbibliotheken Deutschlands. Neben monographischer Literatur werden im engsten Hauptsammelgebiet Aufsätze dokumentiert. Monographien und Aufsätze werden mit Hilfe eines Vokabulars erschlossen, das sich an das Vokabular des Deutschen Bundestages anlehnt (System Gülich). Der ungewöhnliche Bestandszuwachs der letzten Jahre resultiert durch die Übernahme großer internationaler und nationaler Gewerkschaftsbibliotheken. Zu nennen sind in diesem Zusammenhang die Bibliothek des Internationalen Metallgewerkschaftsbundes, des Deutschen Gewerkschaftsbundes und der Industriegewerkschaft Medien. Seit 1994 bietet die Bibliothek ihre vollständigen bestände im WWW an (LINUX-Version mit der Bibliothekssoftware ALLEGRO). Neben dem konventionelle Fernleihverkehr hat die Bibliothek seit Januar 1998 ein kostenpflichtiges Direktbestellsystem im Internet eingeführt (mit Rechnungslegung). Benutzer können nun wahlweise Papierkopien, Faxkopien oder faksimilierte Artikel (als Image) bestellen. Die Bestellung faksimilierter Artikel (als Image) erfolgt über den FTP-Server der GMD in St. Augustin. Der vollständige Bestand der Bibliothek kann auch über den von der DFG aufgebauten Sondersammelgebietsserver WEBIS angesteuert werden. Die Benutzer werden über WEBIS ebenfalls auf das Direktbestellsystem verwiesen.

Die FES gibt jährlich weltweit ca. 800 Monographien heraus. Sie zählt damit zu den größten sozialwissenschaftlichen „Verlegern“. Die Bibliothek der FES hat testweise damit begonnen, Publikationen der FES, die in digitaler Form vorliegen, als Volltexte im WWW anzubieten. Als Formate können wahlweise HTML oder PDF gewählt werden. Die Metadaten der Volltexte werde in einer Allegro-Datenbank mit Hinweisen auf die Fundstelle des FES-Servers verwaltet. Geplant ist die Digitalisierung wichtiger älterer sozialwissenschaftlicher FES-Monographien, die als Beitrag zur digitalen Bibliothek Sozialwissenschaften bereitgestellt werden sollen. Darüber hinaus ist es notwendig maschinenlesbare FES-Texte aus dem Ausland in internetgängige Formate umzuwandeln. Zur Inhaltserschließung setzt die Bibliothek der FES den Thesaurus Sozialwissenschaften des IZ ein, der vor allem für die Erschließung mit englischen Deskriptoren wichtig ist.

Einen weiteren Dienst im Rahmen ihrer internationalen Zusammenarbeit bietet die Bibliothek mit dem IALHI-Net an. In der IALHI (International Association of Labour History Institutions) sind knapp 100 bedeutende Forschungseinrichtungen, Bibliotheken und Archive versammelt. Die Einrichtung wurde 1970 mit Hilfe der FES gegründet. Die größten Institute bauen dezentral einen Current-contents-Dienst der Periodika zur Geschichte der Arbeiterbewegung und Sozialgeschichte auf, wobei in jedem Land eine Einrichtung die Federführung hat. Dieses Projekt kann als wichtiges internationales arbeitsteiliges Projekt angesehen werden. Es ist auch geplant die Server internationaler Gewerkschaftsorganisationen mit ihren vielfältigen Publikationen und Informationsmaterialien zu spiegeln und in Deutschland besser zugänglich zu machen.

Die Bibliothek der FES hat bereits eine Vereinbarung mit dem Göttinger Gemeinsamen Bibliotheksverbund (GBV) abgeschlossen. Die Bibliothek der FES bemüht sich so weit wie möglich, die aktuellen Veröffentlichungen des FES bereitzustellen und über die Datenbanken des GBV nachzuweisen. Die Dokumente werden inhaltlich mit dem deutschen und englischen Vokabular des Thesaurus Sozialwissenschaften des IZ erschlossen. Die Bibliothek der FES bemüht sich ebenfalls, Kurzreferate (Abstracts) zu diesen Titeln in deutscher und englischer Sprache von den Autoren und Herausgebern zu erhalten. Bei Dokumenten, die im HTML-Format vorliegen, werden diese Informationen (Verschlagwortung und Kurzreferate) entsprechend den Spezifikationen des Dublin Core Metadata Element Set in den Dokumenten verankert. Zusätzlich liefert die Bibliothek der FES Katalogisate der entsprechenden Titel im MAB- bzw. PICA-Format. Der GBV wird

demnächst vollständig die an die ZDB gemeldeten Periodikatitel der Bibliothek der FES in den Online-Dokumentenlieferdienst (GBV-direkt) integrieren.

1.3.6 Parallele Arbeiten sozialwissenschaftlicher Verlage

Der Westdeutsche Verlag mit der Zeitschrift „Kölner Zeitschrift für Soziologie und Sozialpsychologie“ sowie der Verlag Leske + Budrich mit der Zeitschrift „Soziologie“ sind als Mitglieder eines Konsortiums sozialwissenschaftlicher Verlage an einem Vorprojekt innerhalb des BMBF-Förderprogramms „Global Info“ beteiligt, das unter anderem die Vorbereitung der Beteiligung an einem fachübergreifenden Projekt der Bereitstellung von Zeitschrifteninhalten (Volltexten) zusammen mit den entsprechenden Metadatenätzen im WWW zum Inhalt hat. Die Bereitstellung von Volltexten ist daher in der Laufzeit des beantragten Projekts zu erwarten. Beide Verlage und Redaktionen haben ein aktives Interesse daran, ihr Angebot in eine virtuelle sozialwissenschaftliche Bibliothek einzubinden, und zu diesem Zweck eng mit dem Projektteam zu kooperieren

2 Ziele des Projekts

In dem hier skizzierten Projekt sollen die zentralen Aspekte eines neu zu definierenden Zusammenspiels von Wissenschaftlern, Fachgesellschaften, Verlagen, Bibliotheken, SSG und Informationsservicestellen prototypisch für die Sozialwissenschaften untersucht und in einem neuen verteilten System realisiert werden. Ausgangspunkt ist dabei die spezielle Problematik der Recherche im SSG. Bei SSGen ist die Problematik der heterogenen und mehrfachen Inhaltserschließung generell besonders kritisch, weil die Grenzen des Gebiets nicht für alle Benutzerwünsche gleichermaßen sinnvoll gezogen werden können. Es gibt immer Überschneidungsbereiche, die das SSG wegen seiner Spezialisierung nicht allein bedienen kann. Ein Einstieg über eine Universitätsbibliothek mit einem speziellen SSG wird somit besonders auf eine vom Benutzer möglichst problemlos handhabbare Verbindung zu parallelen, sich überschneidenden und ergänzenden Beständen an anderen Orten angewiesen sein.

Bei der Systemrealisierung soll so weit wie möglich auf Standardsoftware, Vorarbeiten der BMBF-Projekte aus dem naturwissenschaftlichen Bereich und bestehende Lösungen der Projektpartner zurückgegriffen werden.³⁴ Gleichzeitig sollen parallele Entwicklungen in der Bundesrepublik Deutschland wie andere DFG-Projekte im Bibliotheksbereich (z.B. MILOS, KASCADE und die in Abschnitt 2.2 genannten), das Förderprogramm „Global Info“ des BMBF³⁵, die europäischen Entwicklungen (DESIRE, Telematik für Bibliotheken) und internationale Aktivitäten (vgl. Digital Library Initiative in Abschnitt 2.2, Dublin Core) auf geeignete Lösungen hin untersucht werden, die sich im Projekt implementieren lassen. Die prinzipielle Offenheit des realisierten Systems für die sich abzeichnenden Entwicklungen muß gewährleistet sein.

Ziel des Projekts ist es, ein verteiltes System zur integrierten Informationserschließung aufzubauen und der Öffentlichkeit durch ein funktionsfähiges System über das Internet (WWW) zugänglich zu machen. Dieses System arbeitet nach dem Modell polyzentrischer Informationsmengen auf mehreren Servern, die bei den beteiligten Institutionen eingerichtet werden und über eine zu standardisierende Schnittstelle über das Z39.50-Protokoll miteinander kommunizieren.

Gegenstandsbereiche des Projektes Virtuelle Bibliothek Sozialwissenschaften sollen also sein:

2.1 Integration verschiedenartiger Informationssysteme (Kataloge, Datenbanken)

Bibliothekskataloge werden in verstärktem Maße im Internet zur Verfügung gestellt. Sie decken jedoch jeweils nur einen Teilbereich der einschlägigen Literaturbestände ab. Z. B. enthalten einige Kataloge der Universitätsbibliotheken nicht die Publikationen von Fachbereichsbibliotheken; die Kataloge von Spezialbibliotheken liegen getrennt von beiden vor. Ziel muß es deshalb sein, die verschiedenen elektronisch zugänglichen Bibliothekskataloge zu einem virtuellen Gesamtkatalog zusammenzufassen

³⁴ Grundsatz ist es, nur solche Elemente selbst zu entwickeln, die nicht durch kommerzielle Standardlösungen oder Ergebnisse vergleichbarer Projekte realisiert werden können.

³⁵ siehe <www.global-info.org>

und über ein integratives Verfahren anzubieten. Neben technischen Aspekten ist hier auch das Problem einer eventuell unterschiedlichen Sacherschließung der Bestände zu lösen, wozu vor allem die Begriffsstrukturen im Koordinations- und Integrationsbereich herangezogen werden sollen. Ausgangspunkt der Integrationsbemühungen ist das SSG der USB Köln.

2.2 Integration und Transfer von verschiedenen Inhaltserschließungsverfahren

Bibliothekskataloge erschließen in der Regel Monographien und beschränken sich auf die formale Erfassung der bibliographischen Angaben. Der Inhalt des jeweiligen Dokuments ist durch den Titel und eventuell über eine Sachgebietsklassifikation und in der Regel eine einfache Verschlagwortung repräsentiert, die jedoch zwischen Bibliothekskatalogen verschiedener Standorte unterschiedlich sein kann. Dies gilt auch für die Bestände des SSG Sozialwissenschaften. Informationszentren wie das IZ und Spezialbibliotheken wie die FES erschließen auch unselbständige Literatur und erweitern den Titel durch ein Abstract bzw. den Volltext. Die Inhaltserschließung erfolgt entweder intellektuell und thesaurusgebunden oder über automatische Indexierungsverfahren. Die sozialwissenschaftlichen Zeitschriften und Verlage wiederum stellen i.d.R., wenn überhaupt, nur Volltexte bereit. Aus Benutzersicht ergänzen sich die drei Informationsquellen ganz natürlich. Sie sollen deshalb für den Benutzer über ein integriertes Verfahren zugänglich sein. Die Klärung der hier entstehenden technischen, organisatorischen und konzeptuellen Probleme ist ein wesentlicher Bestandteil des Projekts. Auf der konzeptuellen Ebene muß z.B. eine Integration der verschiedenen Inhaltserschließungsansätze gefunden werden. Deskriptoren aus Bibliothekssystemen können nicht einfach mit den tiefer und differenzierter erschlossenen einer Informationsservicestelle wie dem IZ ohne zwischengeschaltete Transformationsstufe verbunden und semantisch gleichgesetzt werden; sie müssen gleichwohl systemverträglich in das Gesamtangebot integriert werden. Noch differenzierter wird die Erschließungsvielfalt, wenn das textbasierte Angebot im Internet - über die wiederum auf anderen Grundlagen arbeitenden Suchmaschinen - ebenfalls in eine integrative Suche einbezogen werden soll. Die unterschiedlichen Verfahren, Tiefen und Sichtweisen der Erschließung gilt es ebenso einzufangen wie unterschiedliche Bewertungen und Kommentierungen von Autoren, Verlagen, Projekten, Benutzern, Spezial- und

Universalbibliotheken. Kurzum, inhaltsbezogene Metadaten verschiedenster Provenienz sind zu integrieren.³⁶ Die Erfahrungen aus dem Projekt „Katalogerweiterung durch Scanning und automatische Dokumenterschließung“ (KASCADE) können im Laufe des Projektes ausgewertet werden. Dabei kann auch auf Vorarbeiten und Erfahrungen der TUD im Rahmen des TOSCANA-Projekts zurückgegriffen werden.

2.3 Integrierte Bestellverfahren (für Volltexte, Papierkopien, Ausleihe)

Im Zuge der Etablierung einer Virtuellen Bibliothek Sozialwissenschaften wird eine Online-ordering-Komponente (für die Ausleihe von Monographien, den Bezug von digitalen oder gedruckten Aufsatzkopien, den Zugang zu Volltexten) von dem Sondersammelgebiet Sozialwissenschaften der Universitäts- und Stadtbibliothek Köln (USB) und der Bibliothek der FES als wichtig erachtet. Die technischen und organisatorischen Voraussetzungen zur Einrichtung einer solchen Komponente bei den Projektpartnern sollen untersucht und in der Systemkonzeption berücksichtigt werden; dabei kann auf die Erfahrungen der USB im Rahmen verschiedener bibliothekarischer Projekte (SUBITO, JASON) sowie der FES bei der eigenen Direktbestellkomponente und die dort entwickelten Lösungen zurückgegriffen werden. Dadurch wird unmittelbar nach der Suche ein Zugang zu den Informationen selbst ermöglicht, und zwar je nach Nutzer- und Berechtigungsstatus auf den Volltext in digitaler Form, auf Kopien oder als Ausleihe (Direktbestellung).

2.4 Verkürzung der Informationskette und Nutzung der elektronischen Volltexte

Derzeit gibt es kaum einen direkten elektronischen Dokumentenverkehr zwischen Wissenschaftlern, Verlagen und Bibliotheken. Nach dem Erscheinen einer Veröffentlichung wird diese in der Regel über das Medium Papier von den Bibliotheken bzw. Informationsservicestellen beschafft,

³⁶ vgl. Wilensky 1996

elektronisch erfaßt und inhaltlich erschlossen.³⁷ Der gleiche Modalitätsbruch entsteht in der Regel nach der Recherche, da Volltexte bisher kaum für sozialwissenschaftliche Literatur zur Verfügung stehen. Dieser Ablauf ist nicht nur zeit- und kostenintensiv, sondern steht einem umfassenden und zügigen Informationszugriff entgegen.

Prototypisch soll mit den Eigenpublikationen der FES die Einbeziehung von elektronischen Volltexten erprobt werden, durch die Verlage und Wissenschaftler stärker in den Gesamtprozeß der Informationsversorgung und Inhaltserschließung integriert werden. Dadurch ergibt sich eine Verbesserung der Basis für die Inhaltserschließung, da zusätzlich der Volltext für automatische Indexierungsverfahren zur Verfügung steht (z.B. Mehrfachindexierung, Begriffsanalyse). Ferner wird die Verbreitung der Monographie bzw. der Zeitschrift auf elektronischem Weg zusätzlich zum Papierdruck und direkter Nachweis des Volltextes bei der Recherche unterstützt.

Die Einbeziehung elektronischer Zeitschriften verschiedener sozialwissenschaftlicher Verlage könnte in Rahmen eines Projektes des Programms „Global Info“ des BMBF erfolgen, hier haben sich bereits mehrere sozialwissenschaftliche Verlage interessiert gezeigt. Die elektronischen Preprints und Reports könnten im Rahmen eines DFN-Projektes zur Realisierung eines DGS-Servers und von dezentralen Servern an sozialwissenschaftlichen Fachbereichen, Spezialbibliotheken und dem IZ Sozialwissenschaften erfaßt werden. Im Gesamtsystem der angestrebten Virtuellen Bibliothek Sozialwissenschaften werden auch die Zeitschriften analytisch berücksichtigt.

2.5 Anbindung von Clearinghouse-Funktionen und Internetquellen (Planung 3. Jahr)

Durch die parallele Anbindung des sozialwissenschaftlichen Clearinghouse, das vom IZ im Rahmen der GESIS entwickelt und gepflegt wird, kann die Virtuelle Bibliothek Sozialwissenschaften um sozialwissenschaftlich relevante, inhaltlich geprüfte Links zu Internetressourcen erweitert werden.

³⁷ Das IZ erhält von einzelnen Zeitschriften die Kurzreferate (aber nicht die Volltexte der Artikel!) direkt in digitaler Form (per Diskette oder via Internet) zur Weiterverwendung in der Datenbank SOLIS.

Hier kann der Einsatz der Tools zur Aufnahme von Informationen (einschließlich Metadaten) in eine entsprechend strukturierte Datenbank geprüft werden, die gegenwärtig vom SSG Fachinformation der NSUB Göttingen entwickelt werden. In diesen Kontext kann auch der Current-contents-Dienst des IALHI eingebunden werden.

Nach Abschluß des Projektes sollen mit Hilfe des funktionsfähigen Systems weitere Informationsbestände anderer Bibliotheken, Fachinformationsstellen, Datenbankproduzenten, Verlage, wissenschaftlichen Institute und Fachgesellschaften aus dem In- und Ausland in das System der Virtuellen Bibliothek Sozialwissenschaften eingebunden werden können. Der für den Benutzer sichtbare Zugangspunkt für die Virtuelle Fachbibliothek Sozialwissenschaften wird das SSG der USB Köln sein.

2.6 Integration englischsprachiger Dokumente (Planung 3.Jahr)

Die Verfahren sind so ausgelegt, daß multilinguales Retrieval (in Deutsch und Englisch) möglich bleibt. Der IZ Thesaurus enthält neben den deutschen die englischen und russischen Alternativterme. Auch die quantitativ-statistischen Verfahren, die für die Transferkomponente des Schalenmodells getestet werden sollen, können multilinguales Retrieval durch die Analyse von Parallelkorpora in den beiden Sprachen integrieren. Eine präzisere Festlegung der Verfahren des multilingualen Retrievals erfolgt später auf der Basis des in den ersten beiden Jahren erreichten Projektstandes.

Literaturverzeichnis

- Artus, Helmut M. (1992): Graue Literatur: zum informellen Kommunikationssystem der Sozialwissenschaften. Abschlußbericht. Bonn: 1992
- Atkins, Daniel E. (1997): Report of the Santa Fe Planning Workshop on Distributed Knowledge Work Environments: Digital Libraries. 1997. <<http://www.si.umich.edu/SantaFe/report.html>> [eingesehen: 08.10.1997]
- Belkin, Nicolas J. (1996): Intelligent Information Retrieval: Whose Intelligence? In: Krause, Jürgen; Herfurt, Matthias; Marx, Jutta (Hrsg.): Herausforderungen an die Informationsgesellschaft. Konstanz 1996, S. 25-31
- BMBF (1996): Information als Rohstoff für Innovation. Programm der Bundesregierung 1996-2000. BMBF Bonn.
- Buckley, Chris; Singhal, A.; Mitra, M.; Salton, G. (1996): New Retrieval Approaches Using SMART: TREC-4. 1996, 24 S. <http://www-nlpir.nist.gov/trec/trec4.papers/Cornell_trec4.ps> [eingesehen: 19.11.1996]
- Bunzel, Jürgen (1998): The Digital Library Program of the Deutsche Forschungsgemeinschaft. In: ABI-Technik, 18, 1998, Nr.2, S.132-137
- Dennet, D.C. (1971): Intentional Systems. In: The Journal of Philosophy, Nr. 68, 1971
- DFG (Hrsg.) (1998): Weiterentwicklung der überregionalen Literaturversorgung. Memorandum (Stand: 7. Juli 1998) <<http://www.dfg.de/foerder/biblio/memo.html>> [eingesehen: 22.09.1998]
- Frisch, Elisabeth; Kluck, Michael (1997): Pretest zum Projekt German Indexing and Retrieval Testdatabase (GIRT) unter Anwendung der Retrievalsysteme Messenger und freeWAISsf. Bonn: 1997 (IZ-Arbeitsbericht Nr. 10)

-
- Gödert, W.; Lepsky, Klaus (1997): Semantische Umfeldsuche im Information Retrieval. Köln/ Düsseldorf: 1997 (Typoskript)
- Haas, S. W. (1997): Disciplinary Variation in Automatic Sublanguage Term Identification. In: Journal of the American Society for Information Science, Jg. 48, 1997, S. 67-79.
- Hofmann, Ulrich (1996): Virtuelle Organisationen - Bibliotheken an der Schwelle? In: Bibliotheksdienst, Jg. 30, 1996, H.10, S. 1674-1681
- Ingwersen, Peter (1996): The Cognitive Framework for Information Retrieval: A Paradigmatic Perspective. In: Krause, Jürgen; Herfurth, Matthias; Marx, Jutta (Hrsg.): Herausforderungen an die Informationswirtschaft. Konstanz 1996, S. 25-31
- Kluck, Michael (1998): German Indexing and Retrieval Test Data Base (GIRT): Some Results of the Pre-Test [Vortrag IRSG '98, 20th Annual Colloquium of the British Computing Society - Information Retrieval Specialist Group, Grenoble 1998]
- Krause, Jürgen (1996): Informationserschließung und -bereitstellung zwischen Deregulation, Kommerzialisierung und weltweiter Vernetzung. Bonn: 1996 (IZ-Arbeitsbericht Nr. 6)
- Krause, Jürgen; Schwänzl, Roland (1998): Global-Info Slot3: Inhaltserschließung, formale Beschreibung, Identifikation und Retrieval, Metadaten, Vernetzung. Göttingen 1998 <<http://www.global-info.org/gruppe3/doc/basis.html>> [eingesehen: 11.05.1998]
- Mandl, Thomas (1998): Vague Transformations in Information Retrieval. Konstanz 1998
- Mutschke, Peter (1995) Relationale Datenbanksysteme im Vergleich: Eine Zwischenbilanz. Bonn: 1995 (IZ-Arbeitsbericht Nr. 5)
- Nase, Annemarie (1996): Datenbanken zu Forschung und Literatur. In: Krause, Jürgen; Zimmer Monika: Informationsservice des IZ Sozialwissenschaften. Datenbankentwicklung und -nutzung, Netzwerke, Wissenschaftsforschung. Bonn: 1996

- Papenfuß, Claudia (1997) Evaluierung der Oracle RDBMS-Komponenten TextServer und WebServer.
<www.fh-augsburg.de/inf/Medoc/OracleText.html> und
<www.fh-augsburg.de/inf/Medoc/OracleWeb.html>
- Roscheisen, Martin; Baldonado, Michelle; Chang, Kevin; Gravano, Luis; Ketchpel, Steven; Paepcke Andreas (1997): The Stanford Infobus and its Service Layers. Augmenting the Internet with Higher-Level Information Management Protocols <<http://www-diglib.stanford.edu/cgi-bin/WP/get/SIDL-WP-1997-0065>> [eingesehen: 26.05.98]
- Schmiede/Wenzel (1996, 1997); URL: Darmstädter Virtueller Gesamtkatalog (DVK). Homepage. <<http://www.ifs.tu-darmstadt.de/dvk.html>>
- Shoam, Y. (1993): Agent-Oriented Programming. In: Artificial Intelligence, Nr. 60, 1993
- Wilensky, Robert (1996): Toward Work-Centered Digital Information Services. In: Computer, Mai 1996, auch unter
<<http://www.computer.org/pubs/computer/dli/r50037/r50037.htm>>
[eingesehen: 19.02.1997]
- Womser-Hacker, Christa. (1996): Das MIMOR-Modell. Mehrfach-indexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift Universität Regensburg, FG Informationswissenschaft. Dezember 1996.
- Zillmann, Harmut (1997): OSIRIS. Ein gemeinsames Projekt der Universitätsbibliothek Osnabrück und des Instituts für Semantische Informationsverarbeitung der Universität Osnabrück (ISIV). Osnabrück 1997 <<http://osiris.ub.uni-osnabrueck.de/isis/demo/start.htm>>