

## Thesauri, Klassifikationen & Co – die Renaissance der kontrollierten Vokabulare?

PHILIPP MAYR

*Der Beitrag beschäftigt sich mit aktuellen Entwicklungen im Bereich des Terminologie Mapping zwischen kontrollierten Vokabularen. Die unterschiedlichen Typen und Ausprägungen von kontrollierten Vokabularen – Thesauri, Klassifikationen, Schlagwortnormdaten und kontrollierte Schlagwortlisten – dienen hauptsächlich der inhaltlichen Erschließung von Literatur und anderen Dokumententypen und werden u. a. in Datenbanksystemen bei der Deskriptorensuche und zum Browsing angeboten. Die Zusammenführung der heterogenen Indexierungssprachen bei der übergreifenden Datenbankrecherche (cross-database retrieval) führt jedoch zu Problemen bei der semantischen Integration der zugrunde liegenden Bestände. Der Beitrag stellt in diesem Kontext den Ansatz sowie die bisherigen Ergebnisse des Projekts „Modellbildung und Heterogenitätsbehandlung“ vor. Im Anschluss wird auf Bradford's Law of Scattering (BLS) sowie abgeleitete Einsatzmöglichkeiten des BLS im Projekt Bezug genommen: 1) BLS zur Untersuchung der Auswirkungen der Heterogenitätsbehandlung, 2) BLS bzw. Bradfordizing als alternativer, informationswissenschaftlich begründeter Zugangsmechanismus für Zeitschriftenartikel.*

### Einleitung

Nutzer sehen sich je nach Art und Umfang der Recherche immer häufiger in der Situation, auf mehrere Informationsangebote zugreifen zu müssen, um ihr Informationsbedürfnis zu befriedigen. Die alleinige Nutzung z. B. eines Online-Katalogs (OPAC) der lokalen Bibliothek führt genauso wenig zu vollständigen Ergebnissen wie die Beschränkung auf eine einzelne Fachdatenbank.<sup>1</sup> Ein Blick auf die aktuelle Informationslandschaft zeigt, dass viele Informationsangebote zwar dezentral online angeboten werden, aber zumeist unverbunden zur Verfügung stehen. So kann beispielsweise in den Sozialwissenschaften aus einer Vielzahl unterschiedlicher Literaturdatenbanken und Katalogen gewählt werden, die bislang nicht an einer Stelle zusammengefasst bzw. recherchierbar sind. Das hat zur Folge, dass mit jedem Wechsel zwischen

---

<sup>1</sup> Die Notwendigkeit, die Erzeugnisse der IuD/Fachinformation gemeinsam mit den Beständen der Bibliotheken zu sehen und innerhalb der Digitalen Bibliothek zusammenzuführen, wurde kürzlich vom Jubilar erwähnt [38].

den einzelnen Informationssystemen der Nutzer vor das Problem gestellt wird, die bereits für den OPAC formulierte Suche für eine oder mehrere Fachdatenbanken neu zu formulieren bzw. zu übersetzen. Dies erhöht die kognitive Belastung des Nutzers bei der Recherche und führt zwangsläufig zu längeren und unterbrochenen Recherchen. Auf der anderen Seite erwarten Nutzer heute die Bündelung fachlich relevanter Information an einer Stelle. Folglich wird die Integration von Informationssuche sowie die Verbindung von Angeboten unterschiedlicher Fächer zur Vereinfachung des Zugangs bei interdisziplinären Fragestellungen dringend benötigt.

Die kontrollierten Vokabulare der einzelnen Datenbanken und Kataloge sind hierbei ein wichtiges Instrument, inhaltlich erschlossene Informationsangebote zielgenau nach fachlichen Gesichtspunkten zu recherchieren [22; 23; 15]. Eine der sechs Suchtechniken des Browsing und Berrypicking Modells von Bates beschreibt diese Form des Dokumentenzugangs ausführlicher (siehe subject searching [1]). Bislang wissen vor allem erfahrene Rechercheure und Informationsspezialisten um die Möglichkeiten der kontrollierten Vokabulare [39; 36; 34], aber auch um die Problematik, dass im ungünstigsten Fall jede Datenbank die Dokumente mit einer anderen Indexierungssprache erschließt (vgl. den Begriff Heterogenität bei Krause [19; 20]). Das Problem wird insbesondere dann virulent, wenn mehrere Datenbanken parallel über eine Suchanfrage recherchiert werden sollen.

Der folgende Beitrag widmet sich zunächst der aktuellen Situation beim Integrieren von Erschließungsvokabularen unterschiedlicher Informationsangebote, insbesondere der Fachdatenbanken, Bibliothekskataloge, Sammlungen von Internetquellen, Volltextserver (vgl. dazu [24]). Dabei stehen die Heterogenität der Fachangebote und die daraus folgende Problematik für die fachübergreifende Suche im Mittelpunkt. Anschließend werden der Ansatz und die bisherigen Ergebnisse des Projekts „Modellbildung und Heterogenitätsbehandlung“ knapp beschrieben. Zum Schluss werden Bradford's Law of Scattering und die Einsatzmöglichkeiten dieser Gesetzmäßigkeit für das oben genannte Projekt ausgeführt.

### **Heterogenität und Heterogenitätsbehandlung**

Ein zentrales Problem des Information Retrieval ist die Vagheit zwischen der Anfrage eines Benutzers und den Indexierungsbegriffen, die den Inhalt der im Informationssystem gespeicherten Dokumente beschreiben.

„Jedem Bibliothekar und jedem, der sich mit Information Retrieval befasst, war schon immer klar, dass zwischen den semantischen Termen, die in der Datenbank ein Dokument charakterisieren, und dem Term, den der Benutzer anwendet, nicht immer eine 1:1-Relation besteht.“ [18]

Traditionell wird die Vagheit zwischen Anfrage- und Dokumentebene durch Verfahren zur Termerweiterung behandelt, was ‚manuell‘ durch den Nutzer (z. B. durch Verwendung eines Thesaurus<sup>2</sup> mit Synonym- oder Ober-/Unterbegriffsbeziehungen) oder in ähnlicher Weise auch durch das Informationssystem (teil-)automatisch durchgeführt werden kann (z. B. durch Vorschläge für zusätzliche und alternative Suchbegriffe). Diese Vorschlagsfunktionalität findet sich inzwischen bei nahezu allen Internetsuchmaschinen. Handelt es sich bei den zu durchsuchenden Datenbeständen um homogen erschlossene Datenbanken, sind diese Verfahren ausreichend. Anders stellt sich die Situation aber bei heterogen erschlossenen Dokumentenbeständen dar. Durch die unterschiedlichen Thesauri und anderen kontrollierten Vokabulare entsteht Vagheit/Heterogenität bereits auf der inhaltlichen Beschreibungsebene der Dokumente. Stellt man die Datenbanken einander gegenüber und vergleicht, mit welchen Begriffen gleiche Dokumente oder solche mit ähnlichem Inhalt erschlossen sind, so stellt man fest, dass für gleiche semantische Konzepte je nach Erschließungsvokabular unterschiedliche Indexierungsbegriffe verwendet wurden, die sich in ihrer semantischen Präzision unterscheiden können (siehe dazu Deskriptorenbeispiele in Tab. 1 und [8]). Die Konsequenz ist, dass die Anfrage des Nutzers nicht nur auf ein Indexierungsvokabular, sondern parallel auf mehrere abgebildet werden muss (siehe dazu Abb. 1).

Grundsätzlich tritt auch bei der automatischen Indexierung [29] dieses Problem auf, allerdings hier zwischen dem Wortschatz des Informationssuchenden und dem der Autoren der Dokumente, also dem Volltext.

Zur Behandlung der Heterogenität zwischen Indexierungsvokabularen werden am Informationszentrum Sozialwissenschaften aktuell zwei Verfahren erprobt und eingesetzt [13; 18]<sup>3</sup>:

1. Intellektuell erstellte Crosskonkordanzen zwischen Klassifikationen und Thesauri: Crosskonkordanzen sind gerichtete relevanzbewertete Relationen zwischen Termen zweier Thesauri oder Klassifikationen. Sie ermöglichen eine Übersetzung von Anfragetermen eines Erschließungssystems in ein anderes. Die Erstellung der Relationen erfolgt intellektuell, d. h. ein Terminologie-Experte vergleicht verschiedene Begriffssysteme

---

<sup>2</sup> Siehe dazu die Unterscheidung von Enduser-Thesaurus und Indexierungsthesaurus bei Blocks [3] sowie [6; 16].

<sup>3</sup> „Die These ist, dass heterogene Dokumentenbestände zuerst durch Transfermodule bilateral miteinander verbunden werden sollten (Vagheitsmodellierung auf Dokumentenebene), bevor sie in den übergeordneten Prozess der Vagheitsbehandlung zwischen Dokumenten und Anfrage (das klassische IR-Problem) eingefügt werden.“ [18].

und setzt die Terme zueinander in Beziehung. Die Verknüpfung erfolgt dabei nach einem Verfahren, das im Projekt CARMEN [7] definiert wurde und verschiedene Arten von Relationen zwischen Termen unterschiedlicher Vokabulare festlegt [25].

- Quantitativ-statistische Ansätze: Bei diesen Verfahren werden die semantischen Relationen mit Hilfe von statistischen Methoden automatisch erzeugt. Grundlage hierfür sind Parallelkorpora, d. h. Bestände, deren Dokumente gleichzeitig nach zwei Sacherschließungssystemen indiziert sind. (Weitere Informationen zu den quantitativ-statistischen Verfahren finden sich bei Hellweg et al. [13]).

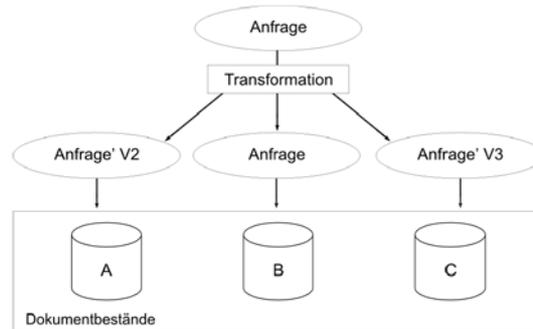


Abb. 1: Anfragetransformation bei heterogenen Datenbanken (aus [13])

Abbildung 1 zeigt abstrahiert den Ablauf einer Anfragetransformation für eine Deskriptorensuche mit einem kontrollierten Vokabular, an der drei Datenbanken beteiligt sind. Es soll angenommen werden, dass die angeschlossenen Datenbanken A, B und C mit drei verschiedenen kontrollierten Vokabularen erschlossen werden, also heterogen sind. Die Transformationskomponente unterhalb der Anfrage erkennt, dass der Anfrageterm aus dem kontrollierten Vokabular der Datenbank B stammt. Folglich transformiert der Algorithmus den Anfrageterm in die äquivalenten Deskriptoren der beiden verbundenen Vokabulare und generiert zwei weitere Anfragen Anfrage' V2 und Anfrage' V3, die an die zugehörigen Datenbanken A und C gehen.

Die qualitative Verbesserung der integrierten Suche in Datenbanken unterschiedlicher Fächer durch Heterogenitätsbehandlung hat sich erstmalig der Informationsverbund Pädagogik, Sozialwissenschaften und Psychologie (infoconnex)<sup>4</sup> zum Ziel gesetzt. So ist jede der Fachdatenbanken in infoconnex mit

<sup>4</sup> <http://www.infoconnex.de>.

einem eigenen Thesaurus erschlossen, der sich hinsichtlich Struktur, Scope und Spezialisierungsgrad wesentlich von den anderen unterscheidet. Um dem Benutzer die Möglichkeit zu geben, mit einem beliebigen – in der Regel dem ihm am besten bekannten – Thesaurus eine Anfrage zu formulieren, die dann auf die drei unterschiedlich erschlossenen Fachdatenbanken angewendet wird, wurden die Thesauri bilateral durch Crosskonkordanzen miteinander verknüpft. Die zwischen den Begriffen der Thesauri definierten Beziehungen (Relationen) erlauben es im Rechercheprozess, die Anfrage des Nutzers vom Vokabular des Ausgangsthesaurus automatisch auf andere Vokabulare abzubilden und mit den für die Zieldatenbank geeigneten Begriffen zu suchen. Die Verbesserung der Recheresituation macht sich insbesondere durch den Gewinn an Zeit bei kurzen, allgemeineren Überblicksrecherchen von Nutzern wie z. B. Studenten, Journalisten, aber auch professionellen Rechercheuren bemerkbar.

The image displays three sequential screenshots of a search interface, illustrating the transformation of a query across different thesauri. Each screenshot shows a search form with a 'Suchformular > Trefferliste' header and an 'authentifiziert' status. The search results are displayed in a table with columns for 'Pädagogik (121)', 'SOLIS (1657)', and 'Psychologie (28)'. The first screenshot shows the query 'Schlagwort=WIRTSCHAFTSTHEORIE[121]'. The second screenshot shows the query 'Schlagwort = ökonomische Theorie (1657)'. The third screenshot shows the query 'Schlagwort='ökonomische theorie'-> (ökonomie und theorien)[28]'. Each screenshot includes an 'Anfrage ändern' button.

Abb. 2: Beispiel einer Anfragetransformation des Schlagworts „ökonomische Theorie“ (aus [41])

Die Abbildung 2 zeigt ein Beispiel einer Anfragetransformation durch Crosskonkordanzen. Übergreifend über alle drei Datenbanken wird eine Schlagwortsuche nach „ökonomische Theorie“ durchgeführt (mittlerer Bereich in Abb. 2). Im Thesaurus Sozialwissenschaften ist dieser Begriff ein gültiger Deskriptor. Die Anfrage „Schlagwort = ökonomische Theorie“ kann direkt an die Datenbank SOLIS weitergeleitet werden, wobei 1657 Dokumente gefunden werden.

Im Thesaurus Psychologie hingegen ist dieser Begriff kein gültiger Deskriptor. Allerdings verweisen Crosskonkordanzrelationen auf äquivalente Terme in diesem Thesaurus, die stattdessen verwendet werden können. Die Anfrage des Nutzers wird für die Datenbank PSYINDEX zur Schlagwortkombination „Ökonomie + Theorien“ transformiert, und es werden 28 Treffer gefunden. Auch im Thesaurus Bildung findet sich der Deskriptor „ökonomische Theorie“ nicht, eine Transformation zu „Schlagwort = Wirtschaftstheorie“ ergibt jedoch 121 Treffer im FIS Bildung.

Derzeit werden in infoconnex lediglich Äquivalenzrelationen genutzt, da sie in jedem Fall automatisch – also ohne Nutzerinteraktion – verwendet werden können. Der Einsatz weiterer Crosskonkordanz-Relationen (Oberbegriff, Unterbegriff, Ähnlichkeit) soll dem Nutzer später über entsprechende Elemente auf der Benutzungsoberfläche schrittweise ermöglicht werden.

*Tabelle 1: Deskriptoren des Thesaurus Sozialwissenschaften (TheSoz) aus dem Systematikbereich 3.4 „Medizin und Gesundheitswesen“ und Termtransformationen in die kontrollierten Vokabulare Thesaurus Bildung (TheBild) sowie Psyndex Terms (PsyT). Die Spalten 2, 4 und 6 zeigen die Treffermengen in den eingebundenen Datenbanken SOLIS, FIS Bildung und Psyndex (infoconnex).*

TheSoz	SOLIS	TheBild (CK)	FIS Bildung	PsyT (CK)	Psyndex
Krankheit	6450	Krankheit	1506	<i>Erkrankungen</i>	2998
öffentliches Gesundheitswesen	239	<i>Gesundheitswesen</i>	628	<i>Gesundheitsversorgungssystem</i>	1835
Arzt-Patient-Beziehung	485	-	-	<i>Ärzte + Patienten</i>	234
Ethno-medicin	72	-	-	<i>Volksmedizin</i>	204
körperliche Entwicklung	75	Körperliche Entwicklung	160	<i>Physische Entwicklung</i>	109
gesetzliche Krankenversicherung	491	<i>Krankenversicherung</i>	144	<i>Krankenversicherung von Arbeitnehmern</i>	15
Knochenkrankheit	-	Knochenkrankheit	2	<i>Osteoporose</i>	13
Impfung	7	Impfung	33	<i>Immunisierung</i>	12
Kinderlähmung	5	Kinderlähmung	7	<i>Poliomyelitis</i>	7
geriatrische Rehabilitation	1	-	-	<i>Geriatrische Patienten+ Rehabilitation</i>	123

Die in Tabelle 1 aufgezeigten Beispiele sind dem Systematikbereich 3.4 Medizin und Gesundheitswesen des Thesaurus Sozialwissenschaften entnommen. Anhand einiger Deskriptoren (z. B. Ethnomedizin, Knochenkrankheit, Kinderlähmung und geriatrische Rehabilitation), die in den Sozialwissenschaften sicherlich ausgesprochene Randgebiete repräsentieren und für die z. T. im Thesaurus Bildung keine Entsprechungen gefunden werden konnten, wird sehr gut deutlich, warum die Crosskonkordanzen in Verbindung mit der interdisziplinären Suche eine entscheidende Verbesserung für den Nutzer bewirken. Für die oben genannten Deskriptoren, die in der Datenbank SOLIS kaum Treffer generieren, bringt gerade die Crosskonkordanz zur Psychologie eine bedeutende Erhöhung der Trefferzahlen, die einem sozialwissenschaftlichen Nutzer als Mehrwertleistung automatisch angeboten werden kann.

Gänzlich ausschließen lassen sich die semantischen Unterschiede zwischen den kontrollierten Vokabularen auch durch die Heterogenitätsbehandlung nicht. Dies hat zur Folge, dass künftige Evaluationen die Qualität bzw. den Ballast (Recall/Precision) der zusätzlichen Dokumente durch die Termtransformationsverfahren auf empirischer Basis nachweisen müssen. Weiterhin sind zusätzliche Mechanismen denkbar, die zu einem Ausgleich der durch die Heterogenitätsbehandlung entstandenen Unschärfe zwischen den Vokabularen (z. B. semantische Unterschiede der Deskriptoren und Indexierungspraxis) führen. Beispielsweise ließe sich die Kernzone der Bradford-Verteilung als Browsingüberstieg nach einer vorherigen Deskriptorensuche konzipieren (siehe dazu späteres Kapitel zu Bradford's Law of Scattering).

Neben den Problemen bei der semantischen Überführung eines Vokabulars in ein anderes, die bei Doerr beschrieben werden [8], ergeben sich insbesondere bei der praktischen Arbeit Probleme, die in den strukturellen Unterschieden der beteiligten Vokabulare begründet sind. Beispielsweise ist es außerordentlich vage, einen Deskriptor aus einer unkommentierten Schlagwortliste, die keine Scope Notes und internen Verweisungen aufweist, mit einem semantisch klaren, also ausdefinierten Deskriptor eines Thesaurus zu verbinden (vgl. dazu Schwarz & Umstätter, [31] und deren Begriff des semiotischen Thesaurus [35]). Nikolai schlägt in diesem Zusammenhang vor, die Vokabulare vor der Integration (er spricht von Thesaurusföderationen) über unterschiedliche Metriken bzgl. ihrer Strukturen zu untersuchen [28]. Zu einer eher pragmatischen und groben Einteilung der in der Praxis vorkommenden Vokabulare bzw. Knowledge Organization Systems (KOS), kommt Hodge. Sie unterscheidet folgende Klassen von Vokabulartypen, die sie weiter unterteilt: "The descriptions given here provide an overview of possible systems for organizing digital libraries. The descriptions are based on characteristics such as structure and complexity, relationships among terms, and historical function. The list is not comprehensive; nor are the definitions of these terms

contained in specific standards documents. They are grouped into three general categories: term lists, which emphasize lists of terms often with definitions; classifications and categories, which emphasize the creation of subject sets; and relationship lists, which emphasize the connections between terms and concepts.” [15]

Auf dieser konzeptuellen Grundlage der Einteilung von Hodge und der Einführung des abstrakten Begriffs KOS lassen sich möglicherweise Modellüberlegungen anstellen. Beispielsweise ist die Verbindung einer ausgereiften Klassifikation mit anderen KOS-Typen/kontrollierten Vokabularen keine Trivialität und setzt klare Regeln und viel Erfahrung beim Terminologie-Mapping voraus [5; 17]. Renardus<sup>5</sup> und MACS<sup>6</sup> sind Beispiele für Projekte, die Crosskonkordanzen erstellt und eingesetzt haben. Die Deutsche Bibliothek, die in unterschiedlichen Crosskonkordanz-Projekten beteiligt war [11], hat dieses Jahr mit dem CRISSCROSS-Projekt<sup>7</sup> begonnen. Ziel ist hier, die Sachschlagwörter der Schlagwortnormdatei mit den Notationen der Dewey-Dezimalklassifikation (DDC) zu verbinden. Eine aktuelle Übersicht von Projekten zum Thema Interoperabilität zwischen KOS (Crosskonkordanzen und weitere Verfahren) findet sich bei Zeng & Chan [43] sowie McIlwaine [26].

### **Projekt Modellbildung und Heterogenitätsbehandlung**

Um der dezentralen und interdisziplinären Informationslandschaft zu begegnen, haben sich das Bundesministerium für Bildung und Forschung (BMBF) und die Deutsche Forschungsgemeinschaft (DFG) für die Schaffung eines generellen Wissenschaftsportals und von Fachangeboten in einem vernetzten Ansatz entschieden, wobei die Projektförderlinien der DFG zu den Virtuellen Fachbibliotheken und die des BMBF zu den Informationsverbänden zusammengeführt werden sollen. Eines dieser Projekte ist das Wissenschaftsportal *vascoda*<sup>8</sup>, das einen generellen Rechercheeinstieg auf oberster Ebene bietet und zu Fachportalen, z. B. dem Informationsverbund *infoconnex*, und einzelnen Fachdatenbanken und Bibliothekskatalogen weiterleitet [30].

Die Konsequenz aus diesem umfassenden Ansatz sind hochkomplexe Strukturen und Anforderungen bei der Integration der für *vascoda* relevanten Informationsangebote. Insbesondere stellt sich die Frage, wie der, mit der parallel zur Zahl der Angebote in *vascoda*, wachsenden Heterogenität begegnet werden soll. Sie betrifft die unterschiedliche Struktur und inhaltliche

---

<sup>5</sup> <http://www.renardus.org>.

<sup>6</sup> <http://www.ddb.de/wir/projekte/macs.htm>.

<sup>7</sup> <http://www.ddb.de/wir/projekte/crisscross.htm>.

<sup>8</sup> <http://www.vascoda.de>.

Erschließung der Dokumente, die von den einzelnen Informationsanbietern geliefert werden [24]. Rein organisatorische Lösungen, also z. B. die Standardisierung der inhaltlichen Erschließung (ein Thesaurus), sind aus vielfältigen Gründen nicht flächendeckend und kurzfristig umsetzbar. Hilfe ist hier nur von einem Rahmenkonzept zu erwarten, das organisatorische, inhaltliche und technische Maßnahmen in Beziehung setzt, Möglichkeiten zur Standardisierung aufgreift und Lösungsvorschläge für den verbleibenden Rest an Heterogenität aufzeigt (Schalenmodell bei Krause [21]).

Die Klärung dieser Fragen soll durch das Teilprojekt „Modellbildung und Heterogenitätsbehandlung“ im Kompetenznetzwerk „Neue Dienste, Standardisierung, Metadaten“ unterstützt werden, das vom BMBF gefördert und seit September 2004 am Informationszentrum Sozialwissenschaften<sup>9</sup> durchgeführt wird. Das Projekt beschäftigt sich schwerpunktmäßig mit

1. der übergreifenden Modellbildung für komplexe Informationsinfrastrukturen, u. a. am Beispiel des Wissenschaftsportals *vascoda* mit allen nachgeschalteten Ebenen,
2. Fragen zur Heterogenitätsbehandlung (Integration) als notwendige Ergänzung zur Standardisierung durch einheitliche Metadaten.

Aus dieser Perspektive heraus erarbeitet das Projekt Vorschläge – Modelle – zur Strukturierung wissenschaftlicher Informationsangebote und schafft Rahmenbedingungen und Infrastrukturen zur Förderung der semantischen Integration heterogener Informationssammlungen. Im Bereich der Heterogenitätsbehandlung fokussieren die Aktivitäten auf die Erstellung von Crosskonkordanzen zwischen fachlichen Erschließungsvokabularen zur besseren Integration innerhalb der Fächer und zwischen den Fächern, sowie auf den Aufbau eines zentralen, internetbasierten Dienstes (Heterogenitäts-Webservice) zur Nutzung dieser Terminologie Mappings in Informationsangeboten. Neben Evaluation und Weiterentwicklung von intellektuellen Verfahren zur Erstellung von Crosskonkordanzen und von statistischen Modellen zur Termtransformation hat das Projekt auch das Ziel, die bislang verfügbaren Verfahren in die Praxis zu überführen und ihre Anwendung zu unterstützen [25].

Abbildung 3 und Tabelle 2 geben den Stand der bislang fertig gestellten Crosskonkordanzen wieder. Der linke obere Bereich umfasst sozial- und politikwissenschaftliche Vokabulare, die durch Crosskonkordanzen miteinander verbunden sind (siehe Pfeile). Im unteren Bereich sind alle übrigen Vokabulare

---

<sup>9</sup> <http://www.gesis.org/Forschung/Informationstechnologie/KoMoHe.htm>.

und deren Verbindungen zusammengefasst (gestrichelte Verbindungen sind geplante Crosskonkordanzen).

Abbildung 3 zeigt, dass sich die beschriebenen Verfahren der Heterogenitätsbehandlung durchaus noch erweitert einsetzen lassen. So kann es für einen Nutzer beispielsweise sinnvoll sein, ausgehend vom DZI-Thesaurus (DZI, in Abb. 3), eine indirekte Termtransformation über den Thesaurus Sozialwissenschaften (TheSoz, in Abb. 3) in das Vokabular der Deutschen Bibliothek (SWD) zu nutzen, um die Bibliotheksbestände mit in die Suche einzubinden. Der Thesaurus Sozialwissenschaften würde dabei als Switching Language fungieren. Weitere Szenarien der Nutzung der Terminologie Mappings sind denkbar (zur Darstellung der indirekten Termtransformation bzw. Mehrschrittverfahren siehe [41]).

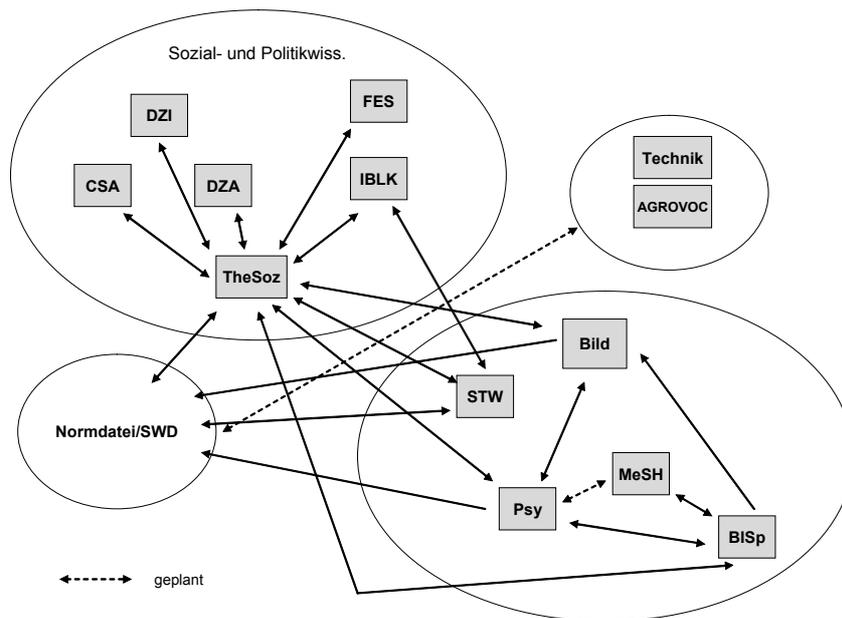


Abb. 3: Netz der Crosskonkordanzen im Projekt „Modellbildung und Heterogenitätsbehandlung“, Stand Ende 2005

*Tabelle 2: Übersicht der verbundenen kontrollierten Vokabulare mit Auflösung der Kürzel aus Abb. 3*

Kürzel	Name des Vokabulars	Größe d. Vok.
DZI	Thesaurus des Deutschen Instituts für soziale Fragen	ca. 2 000
CSA	CSA Thesaurus	ca. 4 000
TheSoz	Thesaurus Sozialwissenschaften (IZ)	ca. 7 500
DZA	Thesaurus des Deutschen Zentrums für Altersfragen	ca. 2 000
IBLK	Thesaurus „Internationale Beziehungen und Länderkunde“ (Euro-Thesaurus)	ca. 9 000
FES	Deskriptoren der Friedrich-Ebert Stiftung	ca. 4 000
SWD	Schlagwortnormdatei der DDB (Ausschnitte)	ca. 500 000
STW	Standard Thesaurus Wirtschaft	ca. 5 600
Psy	Psyndex Terms	ca. 5 300
Bild	Thesaurus Bildung	ca. 55 000
MeSH	MeSH Deskriptoren (Ausschnitte)	ca. 22 000
BISp	Sport Deskriptoren des BISp	ca. 7 500
Agrovoc	Agrovoc Thesaurus (geplant)	ca. 14 000
Technik-T.	div. Technikthesauri (geplant)	Keine Angaben

### Schalenmodell

Für die Modellbildung im Projekt stellt insbesondere das bereits erwähnte Schalenmodell konzeptuell weiterführende Überlegungen an [21]. Das Schalenmodell entwickelt eine konzeptuelle Lösung, die hilft, die Informationsangebote, die bzgl. der Erschließung heterogen sind, strukturiert und reguliert anzuordnen (siehe Niveaus der Datenrelevanz und Erschließung, Trennbarkeit der Schalen). Das Schalenmodell geht ähnlich wie Umstätter ([38], S. 1442) davon aus, dass wissenschaftliche Informationsangebote aus der Bibliotheks- und Dokumentationswelt in der Digitalen Bibliothek vereinigt werden müssen (vgl. dazu [32]). Folgende essentielle Bedingungen müssen erfüllt sein, um in das Schalenmodell integriert werden zu können: fachwissenschaftliche Relevanz der Quellen und koordinierte und kontrollierte Integration der deregulierten Angebote. Wünschenswert ist weiterhin ein konsistenter und tief erschlossener Kernbereich (Schale 1), der sich von weiteren Schalen bzgl. der Qualität der Inhaltserschließung unterscheidet. Es bleibt schließlich den

Informationsanbietern bzw. der Koordinierungsinstanz überlassen, wie viele Schalen angesetzt und welche Kriterien für die Definition der Schalen genutzt werden. So kommt Krause aus Sicht eines Fachdatenbankanbieters zu folgenden Schaleneinteilungen für die Sozialwissenschaften:

„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.

Die zweite Schale lockert die Relevanzbedingungen und parallel dazu die Anforderungen an die Qualität der Inhaltserschließung.

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.“ [21]

Vorteil des Schalenmodells ist neben der Selbstbestimmung der Teilbereiche, dass die Schalendefinitionen alle wichtigen Erschließungsmodelle einbeziehen. Neben den beiden intellektuellen Erschließungstraditionen: generelle Sacherschließung der Bibliotheken durch Normdaten und Klassifikationen und fachspezifische Indexierung der IuD durch Thesauri, findet auch die automatische Erschließung, z. B. durch Suchmaschinentheorie, Eingang in das oben vorgeschlagene Modell. Damit ist das Modell selbst für ‚verschmutzte‘, rein automatisch erschlossene Dokumente der äußeren Schalen offen. Alternativ-Modelle wie z. B. die Standardisierung der kontrollierten Vokabulare in Form eines Metathesaurus – für den Bereich der Medizin z. B. das Unified Medical Language System (UMLS<sup>10</sup>) – scheiden als Modell für vascoda im Grunde aus, da es sich bei vascoda um einen fächerübergreifenden Ansatz handelt. Es wird nicht angenommen, dass es finanziell, organisatorisch und politisch möglich und sinnvoll ist, einen gemeinsamen Thesaurus für jedes Fach bzw. alle Fächer durchzusetzen.

In Analogie zu dem Schalenmodell (Kernbereiche der Forschungsliteratur) ist Bradford's Law of Scattering (BLS) zu sehen. Die bei den Bibliothekswissenschaftlern wohlbekannte informatrische Gesetzmäßigkeit des BLS beschreibt ebenfalls Kernzonen, allerdings anhand von Häufigkeitsverteilungen in den Fachzeitschriften. Auf den ersten Blick ist kein direkter Bezug zu den oben beschriebenen Verfahren der Heterogenitätsbehandlung zu sehen. Allerdings lassen sich mindestens zwei Szenarien beim Einsatz der Komponenten zur Heterogenitätsbehandlung denken, in denen die Annahmen des BLS zum Tragen kommen können.

---

<sup>10</sup> <http://www.nlm.nih.gov/research/umls>.

### Einsatzmöglichkeiten von Bradford's Law of Scattering

Das BLS [4] gilt als eine der anwendbarsten und robustesten informationswissenschaftlichen Gesetzmäßigkeiten (siehe auch Lotka's Law, Zipf's Law usw. [12]), obwohl die Aussagen dieses Gesetzes nicht unumstritten sind [37; 14]. Das informetrische BLS beschreibt die gesetzmäßige Verteilung bzw. Konzentration von Artikeln zu bestimmten Forschungsgebieten oder „Subjects“<sup>11</sup> in Fachzeitschriften. Die Verteilungs- bzw. Konzentrationsmuster sind in den einzelnen Wissenschaftsdisziplinen unterschiedlich stark aufzufinden, wobei die Konzentration auf wenige Kernzeitschriften (Kernbereich, siehe Bradford Zone 1) in den Naturwissenschaften bislang am besten zu beobachten sind.

Umstätter beschreibt die Gesetzmäßigkeit wie folgt:

„Die Grundaussage im BLS war, dass sich eine immer gleiche Menge an Literatur auf verschiedene, thematisch geordnete Zeitschriften wie  $n^0:n^1:n^2:n^3$  ... verteilt. Schon an dieser Schreibweise ist erkennbar, dass es sich lediglich um eine Präzisierung der 80:20 Regel handelt. Wenn wir somit in einer thematisch zentralen Zeitschrift z. B. pro Jahr 50 für uns relevante Aufsätze finden (z. B. „Origin of Life“), und etwa 30 Kernzeitschriften brauchen um 100 zu finden, sowie etwa 27 000 um 200 zu finden, so wäre das BLS erfüllt.“ [37].

Bisher sind in der Literatur unterschiedliche Einsatzmöglichkeiten<sup>12</sup> des BLS beschrieben, von denen hier nur einige genannt werden sollen:

- der Aufbau von Kollektionen,
- die Auswahl von Journals (z. B. bei Thomson Scientific)<sup>13</sup>,
- die Messung der Abdeckung von Bibliographien/Datenbanken und
- die Rechercheunterstützung [42].

Im Sinne der Rechercheunterstützung gehen wir davon aus, dass das Verteilungsmuster des BLS (Kernzone mit hoher Artikelhäufigkeit) für den Benutzer positive Auswirkungen beim Information Retrieval haben kann. Die Such- und Browsingtechniken des Berry Picking Modells [1], aber auch andere Information Seeking Modelle [10] scheinen als Ansatzpunkte für das BLS durchaus geeignet.

---

<sup>11</sup> Als ein Subject kann z. B. Virologie innerhalb der Biologie oder Information Retrieval innerhalb der Informationswissenschaft angesehen werden.

<sup>12</sup> Eine ausführlichere Übersicht bieten beispielsweise Nicolaisen & Hjørland [27].

<sup>13</sup> <http://scientific.thomson.com/knowtrend/essays/selectionofmaterial/journalselection/>.

Auch bei Bates findet sich bereits eine Aussage zu Einsatzmöglichkeiten des BLS für die Informationssuche:

„... the key point is that the distribution tells us that information is neither randomly scattered, nor handily concentrated in a single location. Instead, information scatters in a characteristic pattern, a pattern that should have obvious implications for how that information can most successfully and efficiently be sought.“ [2]

Mindestens zwei für das Projekt „Modellbildung und Heterogenitätsbehandlung“ relevante Einsatzmöglichkeiten lassen sich für das BLS beschreiben.

1. BLS zur Rechercheunterstützung: BLS bzw. Bradfordizing [42] ließe sich als ein alternativer, informationswissenschaftlich begründeter Zugangsmechanismus für Zeitschriftenartikel modellieren. In Anlehnung an die Suchtechniken Browsing, Directed Searching und Linking bei Bates [2] ist folgender Überstieg zwischen den Suchmodi Directed Searching (Deskriptorensuche) und Browsing denkbar: Im Anschluss an eine Deskriptorensuche könnte automatisch ein Bradfordizing der Ergebnismenge vorgenommen werden, um den Kernbereich der Zeitschriften zum weiteren Browsing anzubieten (vgl. journal run bei Bates [1]). Der Nutzer wechselt damit von einer herkömmlichen Term-Suche in einen Browsing-Modus. Die Crosskondordanzen bzw. weiteren Verfahren der Heterogenitätsbehandlung könnten dazu eingesetzt werden, die Anzahl der Datenbanken für die Deskriptorensuche deutlich zu erweitern und somit für eine vollständigere und möglichst interdisziplinäre Dokumentenmenge zu sorgen. Außerdem ließen sich über die Bradford-Zonen (insbesondere Kernzone 1) neue deskriptorenbasierte Browsing-Zugänge generieren, die für bestimmte Forschungsgebiete eines Fachportals<sup>14</sup> die Dokumente innerhalb der Kernzone aggregieren und damit vereinfacht zugänglich machen. Des Weiteren ließen sich anhand der gekoppelten Suchmodi innerhalb der Bradford-Zonen unter Umständen Serendipity-Effekte erzielen. Dies müsste insbesondere in Benutzertests untersucht werden.
2. BLS zur Untersuchung der Auswirkungen der Heterogenitätsbehandlung: Die Bradford-Verteilungen lassen sich auf die durch die Heterogenitätskomponenten deutlich erweiterten und interdisziplinären Dokumentenbestände anwenden und informatrisch auf Veränderungen hin untersuchen. Durch die Anwendung der Term-Transformationsverfahren auf konkrete „Subjects“ lassen sich deutlich veränderte Verteilungen

---

<sup>14</sup> Z. B. die Themenbereiche der Sozialwissenschaftlichen Fachinformationsdienste (soFid) des IZ für die Sozialwissenschaften, <http://www.geis.org/Information/soFid/index.htm>.

(Verteilung mit deutlich mehr Streuung/Scattering, sog. long tail distributions) produzieren. So wird die Verteilung der Zeitschriften z. B. für das Subject „Informationsgesellschaft“ gegenüber den erweiterten Kontexten (Sozialwissenschaften; Sozialwissenschaften + Wirtschaftswissenschaften; Sozialwissenschaften + Wirtschaftswissenschaften + Kulturwissenschaften, usw.) messbar längere Verteilungen generieren, die damit ein Plus an Vollständigkeit und wahrscheinlich auch Interdisziplinarität der Trefferdaten empirisch nachweisen. Damit wäre ein alternatives Evaluationsverfahren für Term-Transformationen skizziert, das ohne vorherige Relevanzbewertung auskommt. Weiterhin ließen sich die unterschiedlichen Operationalisierungen einzelner Subjects vergleichen und gegeneinander abwägen [27]. Als Vergleichswert kann die Deskriptorensuche ohne Term-Transformation dienen.

Zuletzt soll noch auf eine weitere aus 1. abgeleitete Einsatzmöglichkeit des BLS hingewiesen werden. Wie bereits bei den Problemen zur Heterogenitätsbehandlung erwähnt, ließe sich quasi als Ausgleich für die Verluste bei der Term-Transformation, insbesondere bei den noch nicht erprobten Mehrschrittverfahren, ein Browsingzugang der Ergebnismenge durch Bradfordizing der Dokumente vornehmen. Die Vagheit der Termtransformation und die damit zwangsläufig verbundene Absenkung der Precision ließe sich unter Umständen durch ein entsprechendes Browsingangebot für die Kernzone wieder aufheben.

### **Ausblick**

Die dargestellten Verfahren der Heterogenitätsbehandlung und viele weitere Einsatzmöglichkeiten der kontrollierten Vokabulare zeigen, dass heute auf die erschlossenen Dokumentenbestände, sowohl der Bibliotheken als auch der IuD, unter keinen Umständen verzichtet werden kann. Wir können somit auf die Frage im Titel des Beitrags positiv antworten und sehen die kontrollierten Vokabulare in einer Schlüsselrolle. Die entscheidende Frage wird sein, wie sich die Dokumentenräume dieser beiden Erschließungsmodelle mit der automatischen Indexierung verbinden lassen und wie neue Informationsangebote, die keine intellektuelle Erschließung aufweisen, in ein Gesamtmodell integriert werden können.

Umstätter, der in vielen Veröffentlichungen zur Digitalen Bibliothek [39; 40; 9] auf die Bedeutung der kontrollierten Vokabulare, der Sacherschließung und insbesondere der Thesauri und Ontologien (vgl. dazu [40; 33]) hingewiesen hat, legt mit seinem facettenreichen Werk den Grundstock für Forschungsarbeiten an der Schnittstelle zwischen Bibliothekswissenschaft, Dokumentation und Informationswissenschaft.

### Literatur und Internetquellen

- 1 BATES, M. J. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review*, 13, 5, 407-424. URL: <http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>.
- 2 BATES, M. J. (2002). Speculations on Browsing, Directed Searching, and Linking in Relation to the Bradford Distribution. In H. Bruce, R. Fidel, P. Ingwersen, P. Vakkari (Eds.), *Fourth International Conference on Conceptions of Library and Information Science (CoLIS 4)* (S. 137-150). URL: [http://www.gseis.ucla.edu/faculty/bates/articles/Searching\\_Bradford-m020430.html](http://www.gseis.ucla.edu/faculty/bates/articles/Searching_Bradford-m020430.html).
- 3 BLOCKS, D. (2004). *A qualitative study of thesaurus integration for end-user searching*. University of Glamorgan/Prifysgol Morgannwg Dissertation. URL: [http://www.comp.glam.ac.uk/~FACET/dblocks/DBlocks\\_ThesisOnline\\_Main.html](http://www.comp.glam.ac.uk/~FACET/dblocks/DBlocks_ThesisOnline_Main.html).
- 4 BRADFORD, S. C. (1934). Sources of information on specific subjects. *Engineering*, 26, 85-86.
- 5 BUCKLAND, M. & GEY, F. (1999). Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies. *D-Lib Magazine* 5, 1. URL: <http://www.dlib.org/dlib/january99/buckland/01buckland.html>.
- 6 BURKART, M. (2004). Thesaurus. In R. Kuhlen, T. Seeger & D. Strauch (Eds.): *Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis; Bd. 1* (S. 141-154). München: Saur.
- 7 CARMEN-Projekt: CARMEN AP12, Crosskonkordanzen von Klassifikationen und Thesauri, 2002. URL: <http://www.bibliothek.uni-regensburg.de/projects/carmen12/>.
- 8 DOERR, M. (2001). Semantic Problems of Thesaurus Mapping. *Journal of Digital Information* 1, 8. URL: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>.
- 9 EWERT, G. & UMSTÄTTER, W. (1997). *Lehrbuch der Bibliotheksverwaltung*. Stuttgart: Hiersemann.
- 10 Fisher, K. E., Erdelez, S. & McKechnie, L. (Eds.) (2005). *Theories of Information Behavior*. Medford, New Jersey: Information Today, Inc.

- 11 FREYRE, E. & NAUDI, M. (2001). MACS: subject access across languages and networks. In I. C. McIlwaine (Ed.) *IFLA Satellite Meeting* (S. 3-10). Dublin, Ohio: Saur.
- 12 GARFIELD, E. (1980). Bradford's Law and Related Statistical Patterns. *Current Contents* 4, 19, 476-483. URL: <http://www.garfield.library.upenn.edu/essays/v4p476y1979-80.pdf>.
- 13 HELLWEG, H., KRAUSE, J., MANDL, T., MARX, J., MÜLLER, M. N. O., MUTSCHKE, P. & STRÖTGEN, R. (2001). *Treatment of Semantic Heterogeneity in Information Retrieval*. Bonn: IZ Sozialwissenschaften. URL: [http://www.gesis.org/Publikationen/Berichte/IZ\\_Arbeitsberichte/pdf/ab\\_23.pdf](http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_23.pdf).
- 14 HJØRLAND, B. & NICOLAISEN, J. (2005). Bradford's law of scattering: ambiguities in the concept of "subject". In F. Crestani & I. Ruthven (Eds.), *5th International Conference on Conceptions of Library and Information Science*. Berlin: Springer. URL: [http://www.db.dk/jni/Articles/Hjorland&Nicolaisen\(2005\).pdf](http://www.db.dk/jni/Articles/Hjorland&Nicolaisen(2005).pdf).
- 15 HODGE, G. (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. URL: <http://www.clir.org/pubs/reports/pub91/contents.html>.
- 16 KNORZ, G. (2004). Informationsaufbereitung II: Indexieren. In R. Kuhlen, T. Seeger & D. Strauch (Eds.), *Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis; Bd. 1* (S. 179-188). München: Saur.
- 17 KOCH, T., NEUROTH, H. & DAY, M. (2001). Renardus: Cross-browsing European subject gateways via a common classification system (DDC). In I. C. McIlwaine (Ed.), *IFLA Satellite Meeting* (25-33). Dublin, Ohio: Saur. URL: <http://www.ukoln.ac.uk/metadata/renardus/papers/ifla-satellite/ifla-satellite.pdf>.
- 18 KRAUSE, J. (2003). *Standardisierung von der Heterogenität her denken: Zum Entwicklungsstand Bilateraler Transferkomponenten für digitale Fachbibliotheken*. Bonn: IZ Sozialwissenschaften. URL: [http://www.gesis.org/Publikationen/Berichte/IZ\\_Arbeitsberichte/pdf/ab\\_28.pdf](http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_28.pdf).
- 19 KRAUSE, J. (2004). Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. *ZfBB: Zeitschrift für Bibliothekswesen und Bibliographie* 51, 2, 76-89.

- 20 KRAUSE, J. (2004). Standardisierung und Heterogenität. In R. Kuhlen, T. Seeger & D. Strauch (Eds.), *Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis; Bd. 1* (S. 635-641). München: Saur.
- 21 KRAUSE, J. (1996). *Informationserschließung und -bereitstellung zwischen Deregulation, Kommerzialisierung und weltweiter Vernetzung („Schalenmodell“)*. URL: [http://www.gesis.org/Publikationen/Berichte/IZ\\_Arbeitsberichte/pdf/ab6.pdf](http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab6.pdf).
- 22 LANCASTER, F. W. (1972). *Vocabulary Control for Information Retrieval*. Washington, D. C.: Information Resources Press.
- 23 MANECKE, H.-J. (2004). Klassifikation, Klassieren. In R. Kuhlen, T. Seeger & D. Strauch (Eds.), *Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis; Bd. 1* (S. 127-140). München: Saur.
- 24 MAYR, P. (2006). *Informationsangebote für das Wissenschaftsportal vascoda – eine Bestandsaufnahme*. Bonn: IZ Sozialwissenschaften. URL: [http://www.gesis.org/Publikationen/Berichte/IZ\\_Arbeitsberichte/pdf/ab\\_37.pdf](http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_37.pdf).
- 25 MAYR, P., STEMPFHUBER, M. & WALTER, A.-K. (2005). Auf dem Weg zum wissenschaftlichen Fachportal – Modellbildung und Integration heterogener Informationssammlungen. In M. Ockenfeld (Ed.), *27. DGI-Online-Tagung*. Frankfurt a. M.: DGI. URL: [http://www.ib.hu-berlin.de/~mayr\\_etal\\_dgi05.pdf](http://www.ib.hu-berlin.de/~mayr_etal_dgi05.pdf).
- 26 McIlwaine, I. C. (Ed.) (2003). *Subject Retrieval in a Networked Environment*. München: Saur.
- 27 NICOLAISEN, J. & HJØRLAND, B. (2007). Practical potentials of Bradford's law: A critical examination of the received view. *Journal of Documentation* 63 (im Druck).
- 28 NIKOLAI, R. (2002). *Thesaurusföderationen: Ein Rahmenwerk für die flexible Integration von heterogenen, autonomen Thesauri*. Universität Fridericiana zu Karlsruhe Hilden (Rheinland), Dissertation, URL: [www.ubka.uni-karlsruhe.de/indexer-vvv/2003/informatik/6](http://www.ubka.uni-karlsruhe.de/indexer-vvv/2003/informatik/6)
- 29 NOHR, H. (2004). Theorie des Information Retrieval II: Automatische Indexierung. In R. Kuhlen, T. Seeger & D. Strauch (Eds.), *Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis; Bd. 1* (S. 215-225). München: Saur.

- 30 PIANOS, T. (2005). Was macht vascoda? Vision und Wirklichkeit. *Zeitschrift für Bibliothekswesen und Bibliographie* 52, 2, 67-78.
- 31 SCHWARZ, I. & UMSTÄTTER, W. (1999) Die vernachlässigten Aspekte des Thesaurus: dokumentarische, pragmatische, semantische und syntaktische Einblicke. *Nachrichten für Dokumentation* 50, 4, 197-203. URL: <http://www.ib.hu-berlin.de/~wumsta/infopub/pub1996f/pub110.html>.
- 32 SOERGEL, D. (2002). A Framework for Digital Library Research: Broadening the Vision. *D-Lib Magazine* 8, 12. URL: <http://www.dlib.org/dlib/december02/soergel/12soergel.html>.
- 33 SOERGEL, D., LAUSER, B., LIANG, A., FISSEHA, F., KEIZER, J. & KATZ, S. (2004). Reengineering Thesauri for New Applications: the AGROVOC Example. *Journal of Digital Information* 4, 4. URL: <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Soergel/>.
- 34 SVENONIUS, E. (2000). *The intellectual foundation of information organization*. Cambridge, Mass.: MIT Press.
- 35 UMSTÄTTER, W. (ohne Jahr). *Semiotischer Thesaurus*. URL: <http://www.ib.hu-berlin.de/%7Ewumsta/infopub/semiothes/lexicon/default/index.html>.
- 36 UMSTÄTTER, W. (1993). Die Kunst der Frage. *Bibliotheksdienst* 27, 8, 1180-1191. URL: <http://www.ib.hu-berlin.de/~wumsta/pub75.html>.
- 37 UMSTÄTTER, W. (2005). Anmerkungen zu Birger Hjørland und Jeppe Nicolaisen: Bradford's Law of Scattering: Ambiguities in the Concept of "Subject". *Libreas*, 3. URL: [http://www.ib.hu-berlin.de/%7Elibreas/libreas\\_neu/ausgabe3/008ums.htm](http://www.ib.hu-berlin.de/%7Elibreas/libreas_neu/ausgabe3/008ums.htm).
- 38 UMSTÄTTER, W. (2005). Bibliographie, Kataloge, Suchmaschinen. Das Ende der Dokumentation als modernes Bibliothekswesen. *Bibliotheksdienst* 39, 11, 1442-1456. URL: [http://www.zlb.de/aktivitaeten/bd\\_neu/heftinhalte2005/Bibliotheken011105.pdf](http://www.zlb.de/aktivitaeten/bd_neu/heftinhalte2005/Bibliotheken011105.pdf).
- 39 UMSTÄTTER, W. (2006). Macht die Digitalisierung die Bibliotheken überflüssig? URL: <http://www.ib.hu-berlin.de/~wumsta/infopub/pub2001f/GFKL06.pdf>.
- 40 UMSTÄTTER, W. & WAGNER-DÖBLER, R. (2005). *Einführung in die Katalogkunde: Vom Zettelkatalog zur Suchmaschine*. Stuttgart: Hierse-mann.

- 41 WALTER, A.-K., MAYR, P., STEMPFHUBER, M. & BALLAY, A. (2006). Crosskonkordanzen als Mittel zur Heterogenitätsbehandlung in Informationssystemen.. In M. Stempfhuber (Ed.), *In die Zukunft publizieren - 11. IuK-Jahrestagung* (S. 205-225). Bonn: IZ Sozialwissenschaften.
- 42 WHITE, H. D. (1981). 'Bradfordizing' search output: how it would help online users. *Online Review* 5, 1, 47-54.
- 43 ZENG, M. L. & CHAN, L. M. (2004). Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems. *Journal of the American Society for Information Science and Technology* 55, 3, 377-395.

Die zitierten Internetquellen wurden zuletzt am 27.07.2006 aufgerufen.