

**Analyse von maschinell generierten
Korrelationen zwischen der Regensburger
Verbundklassifikation (RVK) und der
Schlagwortnormdatei (SWD)**

Bachelorarbeit

im Fach Wissensmanagement

Studiengang Bibliotheks- und Informationsmanagement

der

Hochschule der Medien

Judith Probstmeyer

Erstprüfer: Prof. Heidrun Wiesenmüller M.A.

Zweitprüfer: Dipl. Inform. Magnus Pfeffer M.A.

Bearbeitungszeitraum: 16. März 2009 bis 15. Juni 2009

Stuttgart, Juni 2009

Kurzfassung

Die Dokumente in Bibliothekskatalogen weisen neben den formalen meist auch inhaltliche Erschließungselemente auf. Im Katalog des Südwestverbunds besitzen zahlreiche Publikationen SWD-Schlagwörter und –ketten sowie Notationen der RVK. Im Anschluss an ein Projekt zur automatischen Vergabe von RVK-Notationen an der Universitätsbibliothek Mannheim wurden aus den SWD-Schlagwörtern und RVK-Notationen der Datensätze im Verbund automatische Korrelationen generiert.

Die vorliegende Arbeit beinhaltet die Ergebnisse der Analyse dieser Korrelationen. Dabei wird sowohl auf die Konkordanz der beiden Dokumentationssprachen eingegangen, die über die SWD-gerechten Registerbegriffe der RVK ermöglicht wird, als auch auf Unterschiede, die etwa über Heterogenität entstehen. Nachdem zunächst nur die automatischen Korrelationen näher betrachtet wurden, konnten diese zudem mit den Notationen und Registerbegriffen der RVK-Online-Konkordanz verglichen werden. Abschließend werden verschiedene Möglichkeiten aufgezeigt, wie die RVK-SWD-Korrelationen in der Praxis angewandt werden können.

Schlagwörter: RVK, SWD, Korrelation, Konkordanz, Heterogenität, Automatische Erschließung

Abstract

In most cases, documents in a library catalogue are not only bibliographically described but also indexed by subject. The union catalogue of south-west Germany contains several publications with notations of the Regensburg Classification (RVK) as well as subject headings of the German subject authority file (SWD). After a project at the university library Mannheim, where a program for automatic classifying of bibliographical records using the RVK has been developed, the notations of the RVK and subject heading of the SWD have been used to create automatic correlations.

This thesis contains an analysis of these correlations and their results. Moreover, the possibility of a mapping of these two documentary languages, that exists indirectly via the index of the RVK, will be discussed and their differences due to heterogeneity. At first, only the correlations themselves are being analyzed. As a next step, these correlations have been compared directly with the notations and index terms of the RVK. The final part of this paper deals with different possibilities of how these correlations could be used in practice.

Keywords: RVK, SWD, correlation, concordance, heterogeneity, automatic indexing

Inhaltsverzeichnis

Kurzfassung	2
Abstract	2
Inhaltsverzeichnis	3
Abbildungsverzeichnis	5
Tabellenverzeichnis	5
Abkürzungsverzeichnis	6
1 Thema und Struktur	7
2 Die beiden Dokumentations Sprachen RVK und SWD	9
2.1 Die RVK.....	9
2.2 Die SWD	12
3 Heterogenität in der Erschließung	15
3.1 Einführung	15
3.2 Crosskonkordanzen	17
3.2.1 Überblick	17
3.2.2 Konkordanz von RVK und SWD	18
4 Automatische Erschließungsverfahren	20
4.1 Automatische Klassifizierung	20
4.1.1 Einordnung und geschichtlicher Abriss.....	20
4.1.2 Automatisches Klassifizieren an der UB Mannheim	22
4.2 Automatisch generierte Korrelationen zwischen verschiedenen Dokumentationssprachen	23
4.2.1 Korrelation von RVK und SWD	23
4.2.2 Verwandte Projekte	24
5 Korrelationsanalyse RVK-SWD	26
5.1 Vorbereitungen	26
5.2 Durchführung der Analyse	28
5.3 Ergebnisse.....	30
5.3.1 Automatisch generierte Korrelationen zwischen RVK und SWD	30
5.3.2 Übereinstimmungen von Konkordanz und Korrelation	36
5.3.3 Neue Konkordanzen	37
5.3.4 Neue Korrelationen	42

6	Anwendungsmöglichkeiten	45
6.1	Optimierung der RVK	45
6.2	Unterstützung bei der Sacherschließung	46
6.3	Anreicherung des SWB	47
6.3.1	Einbindung von Schlagwörtern	47
6.3.2	Einbindung von RVK-Notationen	49
7	Fazit und Ausblick	55
	Literaturverzeichnis	57
	Erklärung	60

Abbildungsverzeichnis

Abbildung 1: Die Online-Version der RVK.....	11
Abbildung 2: Die Online-Version der SWD.....	13
Abbildung 3: Das Schalenmodell	16
Abbildung 4: Umfang der RVK-Konkordanzen in den einzelnen Fachsystematiken....	30
Abbildung 5: Verteilung der Konkordanzen und Korrelationen in den drei Dateien.....	38
Abbildung 6: Entwicklung der Dateigrößen bei unterschiedlichen Schwellenwerten ...	43
Abbildung 7: Tag Cloud zur RVK-Notation AN 91500.....	48
Abbildung 8: Mögliche Darstellung von RVK-Systemstellen im OPAC	52
Abbildung 9: Grafische Darstellung der Korrelationen in Verbindung mit der RVK- Hierarchie.....	54

Tabellenverzeichnis

Tabelle 1: Zwei Publikationen aus dem SWB, die beide mit der Notation DF 7000, jedoch mit unterschiedlichen Schlagwörtern erschlossen wurden.....	14
Tabelle 2: Beispiele für Konkordanzen aus der RVK-Online	26
Tabelle 3: Beispiele für automatisch generierte Korrelationen aus dem SWB	27
Tabelle 4: Verteilung einzelner Schlagworttypen in den Korrelationen	32
Tabelle 5: Titel aus dem SWB mit der Notation EC 4730, die nicht mit dem Schlagwort "Tragik" korrelieren	33
Tabelle 6: Mögliche Konkordanzen für das Schlagwort "Strafrecht"	39
Tabelle 7: Mögliche Schlagwörter für die RVK-Notation CI 7317	46
Tabelle 8: Korrelationen zum Suchbegriff "Umweltrecht"	50

Abkürzungsverzeichnis

BLSH	British Library Subject Headings
BVB	Bibliotheksverbund Bayern
DDC	Dewey Decimal Classification
DFG	Deutsche Forschungsgemeinschaft
GHBS	Systematik der Gesamthochschulbibliotheken des Landes Nordrhein-Westfalen
LCC	Library of Congress Classification
LCSH	Library of Congress Subject Headings
LoC	Library of Congress
MARC	Machine-Readable Cataloging
OPAC	Online Public Access Catalogue
PPN	Pica Production Number
RSWK	Regeln für den Schlagwortkatalog
RVK	Regensburger Verbundklassifikation
SWB	Südwestdeutscher Bibliotheksverbund
SWD	Schlagwortnormdatei
UB	Universitätsbibliothek

1 Thema und Struktur

Inhaltlich erschlossene Dokumente können in einer Datenbank grundsätzlich besser recherchiert werden, als wenn nur eine Formalerschließung erfolgt. Doch auch im Bereich der Sacherschließung gibt es noch immer Anlass zur Verbesserung. Hier besteht u.a. schon in einem einzigen Datenbestand das Problem der heterogenen Erschließungstechniken. Weder auf nationaler Ebene noch in bestimmten Fachgebieten oder bei den allgemein gebräuchlichen Erschließungsmethoden herrscht Homogenität. Zu viele unterschiedliche Klassifikationen und Thesauri existieren parallel und sind nicht aufeinander abgestimmt. Vom Benutzer wird häufig verlangt, dass er mit dem jeweils gebräuchlichen Vokabular oder der Systematik vertraut ist. Andernfalls können zu großen Teilen relevante Informationen verloren gehen.

In Deutschland finden in wissenschaftlichen Bibliotheken neben der zunehmend verbreiteten RVK (Regensburger Verbundklassifikation) hauseigene Systematiken Anwendung. Die mit Vokabular der SWD (Schlagwortnormdatei) erschlossenen RVK-Klassen schaffen eine wichtige Verbindung (Konkordanz) zwischen der RVK als Systematik und der SWD als Thesaurus. Umgekehrt können auch in einem Bibliothekskatalog nützliche Verknüpfungen (Korrelationen) zwischen der SWD und einer Klassifikation wie der RVK hergestellt werden.

Im Rahmen eines Projektes zur automatischen Vergabe von RVK-Notationen an der UB (Universitätsbibliothek) Mannheim wurden in Ansätzen bereits Untersuchungen zum Zusammenhang zwischen RVK und SWD unternommen.¹ Dafür wurden automatische Korrelationen aus den beiden Dokumentationssprachen generiert, die dem SWB entnommen werden konnten.

Das zentrale Thema der vorliegenden Arbeit soll es sein, diese Korrelationen genauer zu untersuchen und auszuwerten. Dabei wird zunächst zum besseren Verständnis eine Beschreibung der beiden Dokumentationssprachen RVK und SWD erfolgen. Anschließend soll auf den Aspekt Heterogenität genauer eingegangen werden, z.B. um welche Art von Heterogenität es sich in diesem Falle handelt und wie sie überwunden werden kann. Die Konkordanz zwischen RVK und SWD wird in diesem Zusammenhang genauer erklärt. Danach wird im Kapitel zur automatischen Erschließung aufgezeigt, wie die Idee zur Erzeugung und Analyse der Korrelationen im Rahmen eines Projektes zur automatischen Erschließung nach RVK an der UB Mannheim entstand. Im darauf folgenden Abschnitt sollen noch einmal konkret die Korrelation zwischen RVK und SWD erläu-

¹ Pfeffer, Magnus: Classification-Weblog der UB Mannheim
URL: <http://blog.bib.uni-mannheim.de/Classification/>

tert und verwandte Projekte umrissen werden. Den Hauptteil der Arbeit stellt die Korrelationsanalyse von RVK und SWD dar. Hier werden neben einzelnen Stichproben der automatisch generierten Korrelationen auch Ergebnisse aus einem Abgleich der Korrelationen mit der RVK-SWD-Konkordanz der RVK-Online näher betrachtet. Im Anschluss daran soll geprüft werden, wie die Korrelationen gewinnbringend in der Praxis eingesetzt werden könnten, sowohl für die Benutzerseite als auch etwa intern für die Sacherschließung.

2 Die beiden Dokumentationssprachen RVK und SWD

Dokumentationssprachen dienen der inhaltlichen, z.B. der verbalen oder klassifikatorischen, Erschließung von Dokumenten. Im Folgenden werden die zwei Instrumente zur inhaltlichen Beschreibung von Publikationen näher vorgestellt, die in der vorliegenden Arbeit eingesetzt werden, nämlich die RVK und die SWD.

2.1 Die RVK

Die RVK ist heute eine der am weitesten verbreiteten Klassifikationen an wissenschaftlichen Bibliotheken in Deutschland, die außerdem Anwender in Österreich, der Schweiz und einigen anderen Teilnehmerländern verzeichnet.² Als Mitte der sechziger Jahre von der UB Regensburg entwickelte hauseigene Aufstellungssystematik fand sie zunächst vor allem Verbreitung in bayerischen Bibliotheken. Inzwischen ist sie auf Grund ihrer Universalität und Eignung als Aufstellungssystematik in vielen großen wissenschaftlichen, vornehmlich UBen, etabliert.³

Die RVK besteht aus 34 Fachsystematiken, die ihren historischen Ursprung im zweischichtigen Bibliothekssystem haben und strukturelle Parallelen zur LCC (Library of Congress Classification) aufweisen. Der erste Bereich „Allgemeines“ umfasst Themen wie „Buch- und Bibliothekswesen, Umweltschutz und Medien“. Darauf folgen die Fächer Theologie, Philosophie und Pädagogik. Der dritte Komplex beinhaltet die Philologien, anschließend finden sich die künstlerischen Fächer sowie Gesellschafts- und Sozialwissenschaften, Geschichte, Recht und Wirtschaft. Mit Geographie und Mathematik beginnt der Bereich der Naturwissenschaften. Dazu zählen des Weiteren Geologie, Physik, Chemie, Biologie und Medizin. Die Sachgebiete Landwirtschaft, Technik und Sport bilden den abschließenden Teil der Klassifikation.

Die Reihenfolge der Fachsystematiken sagt nichts über die Bedeutung der einzelnen Gruppen aus. Dennoch sind einige Bereiche umfangreicher und tiefer gegliedert als andere. Im Allgemeinen umfassen die Wissenschaftsfächer eine eigene Hauptklasse (z.B. Klasse „P“ Rechtswissenschaft). Mitunter kann es aber vorkommen, dass einzelne Fächer mehreren Hauptklassen zugeordnet (z.B. Klassen „WW“-„YZ“ Medizin) oder mehrere Fächer in einer Hauptklasse zusammengefasst sind (z.B. Klasse „V“ Chemie, Pharmazie). Durch die traditionelle Aufgliederung in Fachsystematiken kommt es besonders in neueren Fächern wie z.B. dem Thema Umweltschutz an einigen Stellen zu

² Liste der Anwender der RVK, Stand: 04.02.2009

URL: <http://www.bibliothek.uni-regensburg.de/Systematik/pdf/anwender.pdf>

³ Vgl. Lorenz, Bernd (2008), S. 10

thematischen Überschneidungen zwischen mehreren Fachsystematiken. Allein der Registerbegriff „Umweltschutz“ taucht an vier verschiedenen Systemstellen auf: „MS 1700 – Umweltschutz und -planung (soziologisch)“ im Bereich Soziologie, „QT 200 – Erhaltungsmaßnahmen“ in der Systematik Wirtschaftswissenschaften, im Notationsbereich „VN 9250 - VN 9259 – Chemische Aspekte der Umweltbelastung und des Umweltschutzes, Radioaktiver Abfall, Chemische Industrie und Umwelt, Sicherheit von Chemieanlagen und Kernkraftwerken“ im Gebiet Chemie und in der Biologie bei „WK 6000 – Umweltschutz allgemein“. Zudem ist „Umweltschutz“ zusammen mit anderen Registerbegriffen als Schlagwortkette an zahlreichen weiteren Notationen der Systematik „A – Allgemeines“ nachgewiesen. Über Siehe-auch-Hinweise und Siehe-Verweisungen wird jedoch versucht, bestimmte Themenkomplexe von anderen abzugrenzen bzw. Hinweise auf ähnliche Inhalte zu geben, um eine eindeutige Systematisierung von Dokumenten zu ermöglichen.

Die RVK ist relativ fein gegliedert und besitzt 148.407 Grundnotationen (Stand: 26.1.2008). Mit 13 Stufen wird in der Systematik Germanistik die höchste Gliederungstiefe erreicht. Die Notationen bestehen in der Regel aus einem Großbuchstaben für die erste Hauptklasse und einem weiteren für die erste Unterklasse. Darauf folgen enumerativ meist drei- bis sechsstellige arabische Zahlen. Die Notationen bilden die Hierarchie der RVK nicht durchgängig ab. Aus diesem Grund ist etwa im Gegensatz zur DDC (Dewey Decimal Classification) während einer Recherche oft keine optimale Möglichkeit zur Trunkierung gegeben. Schlüssel, die formal oder sachlich ähnliche Inhalte in verschiedenen Fachsystematiken kennzeichnen, sind auf vielfache Art vorhanden, aber eher selten und zum Teil inkonsequent angewandt. Dazu zählen z.B. Formal-, Länder-, Autoren-, Zeit- und Zahlenschlüssel.⁴

Der verbale Zugang zur RVK wird zum einen über die Klassenbenennungen ermöglicht, insbesondere aber über die Registerbegriffe, die nach der SWD und den RSWK (Regeln für den Schlagwortkatalog) angelegt sind. Diese sind zum Teil nur als allein stehende Schlagwörter in den Klassen vorhanden oder auch als Schlagwortketten und dienen der genaueren Beschreibung der Klassen. Sie befinden sich nicht nur an einzelnen „Blättern“ der RVK, sondern können auch zur Beschreibung eines ganzen Notationsbereiches an einem „RVK-Ast“ vorgefunden werden. Der Bereich „CC 7260 - CC 7266 – Angewandte Ethik (Bioethik, Medizinische Ethik, Ethik der einzelnen Wissenschaften)“ besitzt etwa die Registerbegriffe „Angewandte Ethik || Praktische Ethik || Bioethik || Medizinische Ethik“. Dem untergeordnete Blattknoten „CC 7265 – Feministische Ethik, Geschlechterforschung (Gender Studies)“ gehören außerdem die Registerbegriffe „Feministische Ethik || Geschlechterforschung / Ethik“ an. Im Rahmen eines Projektes der DFG (Deutsche Forschungsgemeinschaft) in den neunziger Jahren wurden zahlreiche

⁴ Vgl. Lorenz, Bernd (2008), S. 37-39

RVK-Klassen über zusätzliche Registerbegriffe erschlossen, doch ein Großteil der Klassifikation bleibt lediglich über die Klassenbenennungen verbal zugänglich.

Inzwischen ist die RVK auch als Online-Fassung⁵ verfügbar (Abbildung 1). In der Web-Version wird die Hierarchie der Klassifikation über eine Art Baumstruktur besser abgebildet. Der Benutzer kann in den Suchbäumen browsen oder aber über Suchmasken im oberen Teil der Webseite direkt nach Registerbegriffen, Klassenbenennungen oder Notationen recherchieren. Mit einer gefundenen Notation kann in der RVK-Online außerdem direkt nach zugehörigen Titeln im SWB, BVB (Bibliotheksverbund Bayern), OBV (Österreichischer Bibliothekenverbund) oder einer konkreten Bibliothek gesucht werden. Neben den RVK-Anwendern und Fachreferenten ist die RVK-Online auch ein hilfreiches Instrument für andere an der Klassifikation interessierte Nutzer.

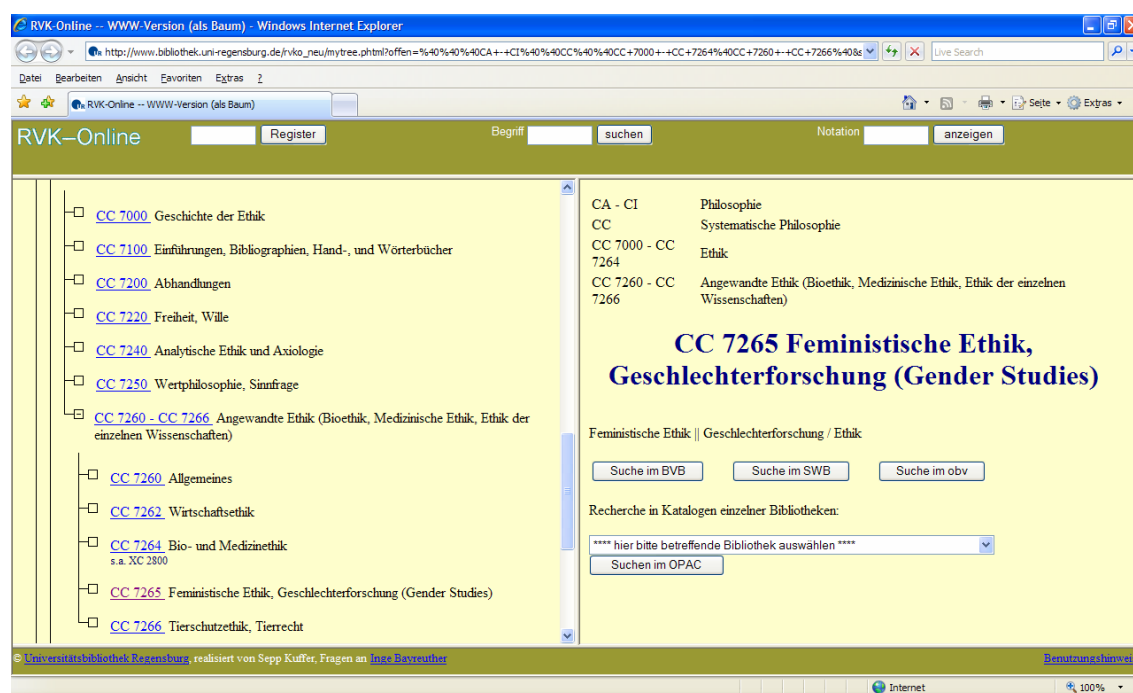


Abbildung 1: Die Online-Version der RVK

Vor Einführung der RVK-Online wurden die Bearbeiter der Systematik schon einmal über ein Programm unterstützt, das für die Fachsystematiken, die noch nicht bzw. unvollständig verbal erschlossen waren, automatisch Registerbegriffe aus der SWD vorschlug. Dadurch entfiel ein Teil der zeitintensiven Recherche nach SWD-Begriffen. Nach Beendigung des DFG-Projektes lief das EDV-unterstützte Erstellen von Registern jedoch wieder vollständig manuell ab. Grund dafür war u.a. ein auf Diskette vorliegen-

⁵ RVK-Online

URL: http://www.bibliothek.uni-regensburg.de/rvko_neu/

der und damit nicht aktualisierungsfähiger SWD-Abzug, der im Laufe der Zeit nur noch veraltete Begriffe hervorgebracht hätte.⁶

2.2 Die SWD

Die SWD ist die zentrale deutschsprachige Normdatei im Bereich der bibliothekarischen Sacherschließung und enthält Ansetzungs- und Verweisungsformen von Schlagwörtern. Sie verfolgt im Gegensatz zur Klassifikation das Ziel der inhaltlichen Feinerschließung über einen natürlichsprachigen Wortschatz und basiert auf den „RSWK“ und den „RSWK/SWD-Praxisregeln“. Auf deren Grundlage dokumentiert sie terminologisch kontrolliertes Vokabular aus allen Fachgebieten und Schlagwortkategorien. Erstellt und gepflegt wird die SWD von der Deutschen Nationalbibliothek und den Bibliotheksverbänden aus der Bundesrepublik Deutschland und Österreich.

Die terminologische Kontrolle der Begriffe erfolgt auf verschiedenen Wegen. Über die Festlegungskontrolle wird zunächst für jeden Begriff ein eindeutiger Deskriptor bestimmt. Hier hat in der Regel die gebräuchlichste Bezeichnung Vorrang. Zusammen mit diesem Deskriptor werden dann mögliche Relationen erfasst, wie z.B. Synonyme, chronologische oder hierarchische Verweisungen (Oberbegriffe) oder weitere verwandte Begriffe (Assoziationen). Des Weiteren erfolgt eine Zerlegungskontrolle, also die Überprüfung, ob der Begriff im Falle eines Kompositums so übernommen werden soll oder evtl. eine Verknüpfung mit bereits vorhandenen Schlagwörtern der SWD vorzuziehen ist. Damit zählt die SWD nach DIN 1463 grundsätzlich zu einem Thesaurus.

Auch die SWD besitzt einen kostenfreien Online- und Recherchezugang (Abbildung 2), in dem etwa nach der Ansetzungform oder verwandten Begriffen gesucht werden kann.⁷ Neben den eben genannten Relationen befinden sich im Schlagwortsatz noch weitere Angaben. Dazu zählen z.B. PPN und SWD-Nummer, um den informationstechnischen Zugriff zu erleichtern, Quellenangaben, Definitionen und Ländercodes. Weiterhin besitzt die SWD eine systematische Gliederung der Schlagwörter mit 36 Hauptgruppen. Die Systematiknummern sind ebenfalls im SWD-Satz enthalten, hierarchisch angeordnet und damit leicht recherchierbar.⁸ Etwa 80 Prozent der Schlagwörter in der SWD sind systematisch erschlossen. Dazu gehören vor allem Sach-, Körperschafts-, Personen-Schlagwörter und in bestimmten Fällen Geographika.⁹

⁶ Vgl. Lorenz, Bernd (2008), S. 51 sowie E-Mail-Auskunft von Dr. Albert Schroeder, UB Regensburg

⁷ Online-SWD

URL: <http://swb.bsz-bw.de/DB=2.104/>

⁸ Notationen der SWD-Sachgruppen

URL: http://www.d-nb.de/standardisierung/pdf/swd_syst.pdf

⁹ Vgl. Scheven, Esther (2005), S. 752

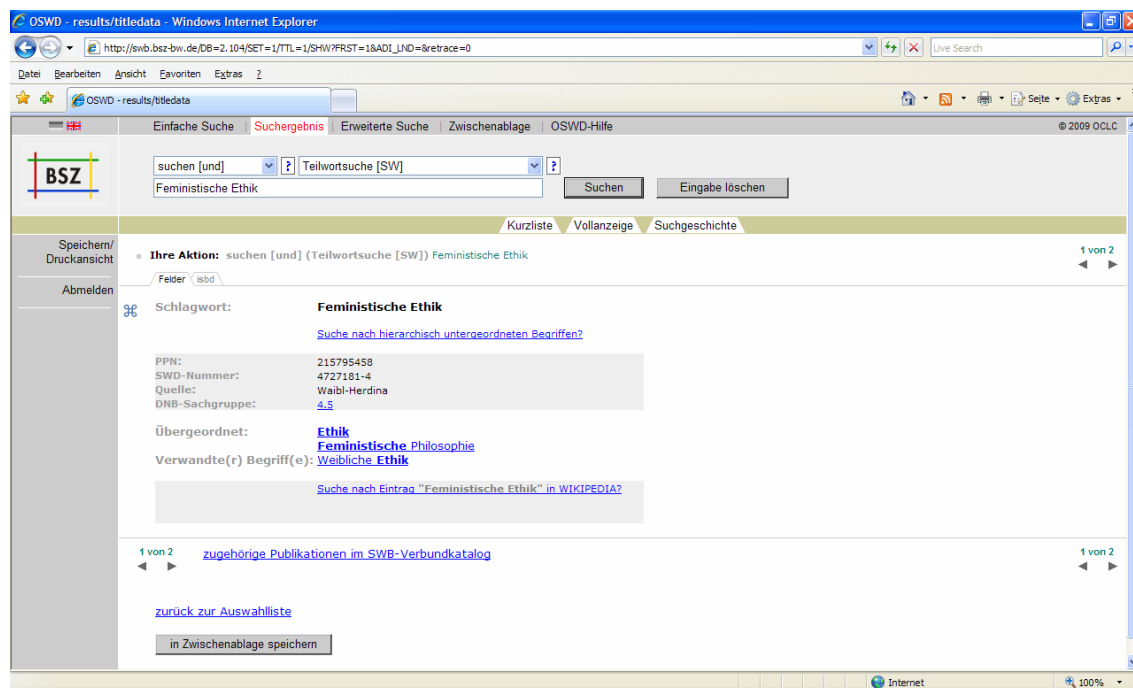


Abbildung 2: Die Online-Version der SWD

In der Regel haben die Bibliotheken über eine Kopie im Verbund Zugang zur SWD. Dadurch wird ein unkomplizierter Datenaustausch ermöglicht, und die SWD kann als kooperatives und effizientes Nachweis- und Erschließungsinstrument genutzt werden. Wenn verschiedene Bibliotheken konstant nach den RSWK und über die Normdateien verbal erschließen, wird eine größere Homogenität von verschiedenen Datenbeständen garantiert. Der Benutzer profitiert davon insbesondere während der Recherche im O-PAC (Online Public Access Catalogue).

Über mit normierten Schlagwörtern versehene Publikationen können im Bibliothekskatalog gut Genauigkeitsrecherchen vollzogen werden. Ein Dokument, das verbal über ein Schlagwort oder eine Schlagwortkette erschlossen wird, grenzt sich damit von anderen ab, obwohl diese etwa aus dem gleichen Sachgebiet stammen und eine identische Notation besitzen. Tabelle 1 zeigt zwei Publikationen, die dem SWB entnommen wurden. Obwohl sie der gleichen RVK-Klasse „DF 7000 – Allgemeines und Deutschland“ aus dem Bereich „Pädagogik/Systematische Pädagogik/Erziehungspraxis“ angehören, thematisieren sie doch zwei verschiedene Themen aus diesem Gebiet. Dies wird zum einen anhand des Titels, zum anderen auch aus den jeweiligen Schlagwörtern ersichtlich.

Titel:	<u>Erziehungskonzepte in der Schule : Praxishilfen für den Umgang mit Schülerinnen und Schülern</u> / Jürgen Bennack
RVK-Notation:	DF 7000 Pädagogik/Systematische Pädagogik/Erziehungspraxis/Allgemeines und Deutschland
Schlagwortkette:	* <u>Schulpädagogik</u> ; <u>Lehrbuch</u>
Titel:	<u>Warum unsere Kinder Tyrannen werden : oder: Die Abschaffung der Kindheit</u> / Michael Winterhoff. Unter Mitarb. von Carsten Tergast
RVK-Notation:	DF 7000 Pädagogik/Systematische Pädagogik/Erziehungspraxis/Allgemeines und Deutschland
Schlagwortkette:	* <u>Kind</u> ; <u>Verhaltensstörung</u> ; <u>Eltern</u> ; <u>Projektion <Psychologie></u> ; <u>Soziale Probleme</u> * <u>Antiautoritäre Erziehung</u> ; <u>Kritik</u> * <u>Erziehungsfehler</u> ; <u>Kind</u> ; <u>Dominanzstreben</u> ; <u>Prävention</u>

Tabelle 1: Zwei Publikationen aus dem SWB, die beide mit der Notation DF 7000, jedoch mit unterschiedlichen Schlagwörtern erschlossen wurden

3 Heterogenität in der Erschließung

3.1 Einführung

Inzwischen ist über viele Portale und Kataloge, wie etwa den KVK (Karlsruher Virtueller Katalog) oder die Suchmaschine BASE (Bielefeld Academic Search Engine), eine übergreifende Suche in mehreren verteilten Datenbeständen möglich. Diese Daten sind häufig beispielsweise auf Grund verschiedener Dokumentarten auf unterschiedliche Art inhaltlich erschlossen und damit nicht durchgängig gleichermaßen retrievalsfähig. Bibliografische Formate wie MAB (Maschinelles Austauschformat für Bibliotheken) oder MARC (Machine-Readable Cataloging) und entsprechende Schnittstellen oder die Metadaten-Initiative Dublin Core erleichtern inzwischen den Austausch zwischen verschiedenen Daten der formalen Erschließung, obwohl auch hier keine vollständige Homogenität gegeben ist. Dokumentationssprachen hingegen unterscheiden sich insbesondere in inhaltlicher Breite und Tiefe und bedürfen anderer Mittel der Angleichung. Hier würde die Standardisierung, wie z.B. die Einführung einer allgemein gebräuchlichen Klassifikation, sehr schnell an ihre Grenzen stoßen.¹⁰

Semantische Heterogenität entsteht, sobald Dokumente auf verschiedene Arten inhaltlich erschlossen sind. Dies kann sowohl bei verteilten Datenbeständen zutreffen als auch in einer geschlossenen Datenbank, in der beispielsweise verschiedene Klassifikationen oder Thesauri Anwendung finden. Diese Probleme gilt es auf anderem Wege zu lösen als über Standardisierung, die in unserer heutigen Zeit nahezu unmöglich geworden ist.

Offensichtlich wird die Problematik der Heterogenität für den Benutzer erst beim Retrieval. Bei der Nutzung einer integrierten Suche etwa über einen Verbundkatalog, entstehen häufig Konsistenzbrüche. Ein vom Benutzer gewählter Term entspricht oft nicht dem in der Datenbank terminologisch kontrollierten Vokabular. Es kann jedoch nicht vorausgesetzt werden, dass er sich vor jeder Recherche mit dem Wortschatz der Datenbank vertraut macht.

Einen Überblick zur Problematik der Heterogenität und Lösungsansätze bietet das Schalenmodell (Abbildung 3).¹¹ Es wurde 1996 für die Daten des Informationszentrum Sozialwissenschaften entwickelt und stellt Klassen von Dokumenten mit unterschiedlicher Relevanz und Inhaltserschließung dar. Über Transferkomponenten zwischen den ver-

¹⁰ Vgl. Panzer, Michael (2008), S. 62

¹¹ Vgl. Krause (2000), Jürgen, S. 206-209

schiedenen Formen der Erschließung können die Grenzen und Differenzen überwunden werden.

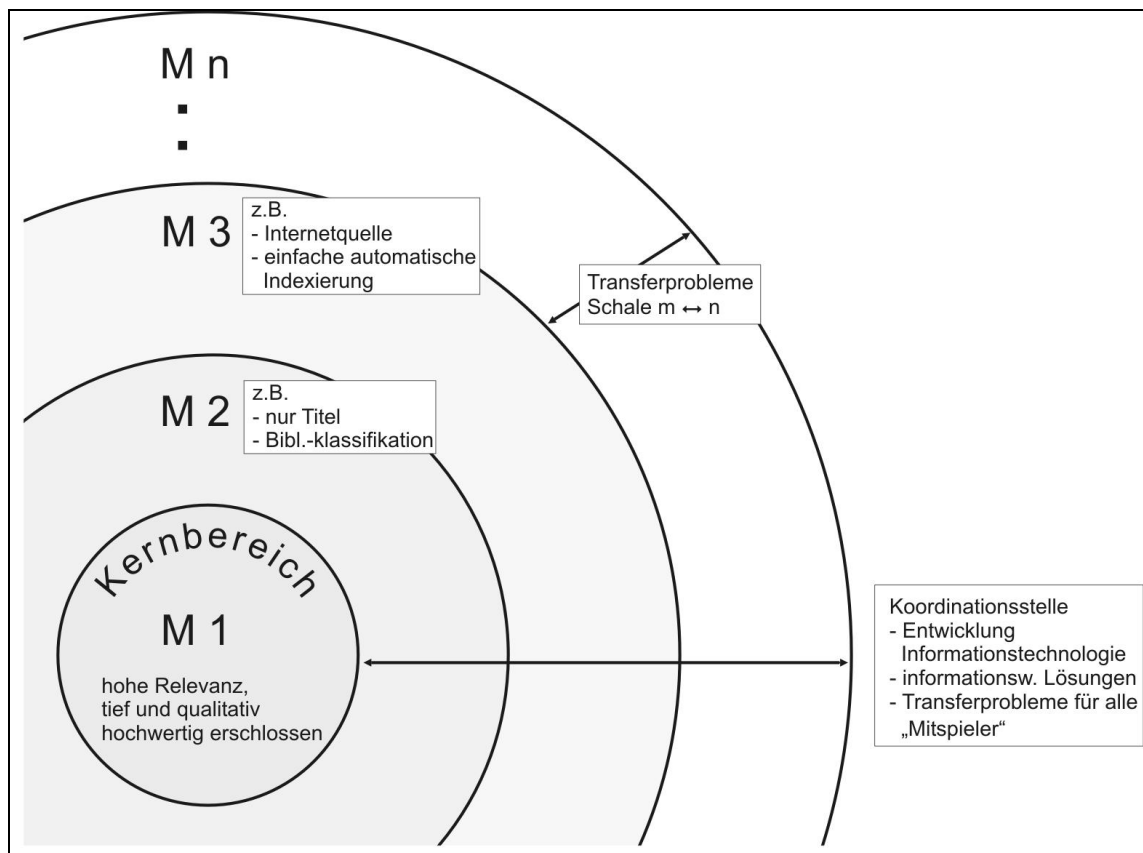


Abbildung 3: Das Schalenmodell

Der Bereich M1 stellt die besonders relevante Literatur dar, z.B. Dokumente eines Sondersammelgebietes. Die folgenden Mengen sind weniger relevant und damit vermutlich zugleich weniger qualitativ hochwertig inhaltlich erschlossen. M2 beinhaltet beispielsweise andere digitale Bibliothekskataloge, deren Bestände von Experten erschlossen wurden. M3 bezieht sich dann etwa auf Internetquellen unterschiedlicher Qualität, die automatisch indiziert wurden. Das Modell kann im Grunde aber auf alle Arten von Fachgebieten, Dokumenten und deren Anbieter projiziert werden.

Für den anspruchsvollen Nutzer sollten die Grenzen zwischen den Ebenen sichtbar sein und er sollte frei entscheiden können, in welcher Menge er suchen möchte. Andererseits sollte auch eine schalenübergreifende Suche angeboten werden, so wie es heute bereits in vielen Katalogen und Portalen realisiert wird.

Unterschiedliche Erschließungstechniken existieren aber nicht nur in den verschiedenen Schalen bzw. zwischen verschiedenen Dokumentarten, sondern ebenso in einer einzigen Schale. Mit der voranschreitenden Forschung und Entwicklung entstehen immer wieder

neue spezifische Fachthesauri und Klassifikationen, die sich immer schwerer vereinen lassen.

3.2 Crosskonkordanzen

3.2.1 Überblick

Um die Grenzen der Schichten aus dem Schalenmodell (Abbildung 3) zu überwinden und die verschiedenen Typen miteinander zu verknüpfen, aber auch um eine Verbindung zwischen den gleichwertigen Dokumenten mit unterschiedlichen Dokumentations-sprachen herzustellen, wurden verschiedene Transfermodule entwickelt. Dazu gehören qualitativ-statistische Ansätze, qualitativ-deduktive Verfahren und Crosskonkordanzen, die meist gemeinsam eingesetzt werden. Auf letztere soll in diesem Abschnitt ausführlich eingegangen werden.

Crosskonkordanzen (auch: Crosswalks oder Mapping) können Differenzen der Inhaltser-schließung überwinden, indem intellektuelle oder automatisch erzeugte Verbindungen zwischen den Termen der verschiedenen Dokumentationssprachen hergestellt werden. Sie bauen auf vorhandenen Systemen auf und sind damit bilateral und statisch. Cross-konkordanzen streben keine neue Standardisierung bestehender Terminologien an und dürfen daher keinesfalls mit Metathesauri verwechselt werden. Konkret verhelfen Crosskonkordanzen während einer Recherche einen Term des einen Systems mit dem eines anderen Systems zu verknüpfen. Im Idealfall besteht zwischen den Termen eine Synonymierelation, also eine 1:1 Beziehung. Ein Term kann aber auch mit mehreren Termen der anderen Sprache verbunden werden, d.h., es entsteht eine 1:n Relation. Wenn keine Verknüpfung zwischen einem Begriff des ersten Systems und einem ande-ren des zweiten Systems gefunden werden kann, spricht man von einer Nullrelation.¹²

Konkordanzen können zwischen Thesauri, wie dem Thesaurus Sozialwissenschaften und der SWD, zwischen Klassifikationen, wie aktuell in der Diskussion zwischen DDC und RVK, oder zwischen Dokumentationssprachen unterschiedlichen Typs, wie es im Projekt CrissCross zwischen DDC und SWD realisiert wird, erstellt werden. Dabei kön-nen fachspezifische, universale oder auch zunächst fachfremde Thesauri oder Klassifi-kationen miteinander verknüpft werden. Generell kann jedoch eine feiner unterteilte Dokumentationssprache leichter auf eine gröbere abgebildet werden.¹³

Die Vorteile einer Crosskonkordanz liegen vornehmlich in einer verbesserten Retrieval-fähigkeit. Dem Benutzer wird die Suche erleichtert, da nur ein Suchvokabular nötig ist und somit eine bessere Trefferquote erzielt werden kann. Die Prozesse laufen dabei

¹² Vgl. Krause, Jürgen (2004), S. 637-639 **übereinstimmend** Krause, Jürgen (2000), S. 209f

¹³ Vgl. Bertram, Jutta (2005), S. 253-255

meist unsichtbar für den Recherchierenden im Hintergrund ab. Ein anderer Vorzug offenbart sich für die Anwender der Konkordanz, wenn sie ihre Dokumentationssprache mit einer anderen vergleichen. Auf diesem Weg wird zumindest eine allgemeine Analyse z.B. bei thematisch ähnlichen Dokumentationssprachen möglich und Schwachstellen des jeweiligen Systems können ausfindig gemacht und überarbeitet werden.¹⁴

3.2.2 Konkordanz von RVK und SWD

Auch zwischen der RVK und der SWD besteht eine Crosskonkordanz über die Systematikregister, denn die Klassen der RVK weisen überwiegend SWD- und RSWK-gerechte Registerbegriffe auf. Bei dieser speziellen Art von Crosskonkordanz geht es aber nicht nur um die Verbindung zweier unterschiedlicher Dokumentationssprachen, sondern speziell auch um die Herstellung eines verbalen Zugangs zu den künstlichsprachigen RVK-Notationen und um die sinnvolle Ergänzung der Klassenbenennungen mit Begriffen, insbesondere dort, wo diese zu allgemein und ohne Aussagekraft sind.

So kann ein unerfahrener Nutzer die Notation „BD 5810“ keinem spezifischen Themenbereich zuordnen. Betrachtet man jedoch die Systemstelle in der RVK-Online, wird ersichtlich, dass die Notation aus der Fachsystematik „Theologie und Religionswissenschaften“ stammt. Eine Navigation durch das hierarchische Verzeichnis ergibt, dass BD 5810 das Sachgebiet „Geschichte und Kultur“ der Juden in der Neuzeit auszeichnet. Der Klassenbenennung sind hier darüber hinaus noch Registerbegriffe in Form der zwei Schlagwortketten „Judentum / Geschichte <1500-1750> || Judentum / Kultur / Geschichte <1500-1750>“ zugeordnet, weil die Thematik in diesem Umfang nicht in der Klassenbenennung auftaucht und andernfalls nur über den Kontext der Systematik bzw. die hierarchisch übergeordneten Klassen ersichtlich wird.

Die Klassenbenennungen und insbesondere Registerbegriffe sind demnach hilfreich bei der Suche nach einer Notation, die wiederum für die Recherche nach einem bestimmten Themengebiet in einem Bibliothekskatalog nützlich sein kann.

1995 waren rund 65% der RVK-Klassen mit Registerbegriffen nach RSWK und SWD versehen.¹⁵ Diese das Register ausmachenden Schlagwörter und -ketten stellen zu den vergleichsweise größeren Notationen der universalen RVK eine feinere Erschließung der Klasse dar. Trotzdem bedarf die Erschließung einer Systemstelle natürlich noch immer eines „weiteren“ Schlagwortes, als es vielleicht für ein spezifisches Dokument in einer Datenbank vergeben werden könnte, denn dort wird vornehmlich über das Prinzip des engen Schlagwortes erschlossen.¹⁶

¹⁴ Vgl. Bambey, Doris (2000), S. 254

¹⁵ Vgl. Geißelmann, Friedrich (1995), S. 132

¹⁶ Vgl. Deutsches Bibliotheksinstitut (1998), S. 202f

Gerade in diesem Punkt ergänzen sich RVK und SWD bei einer Suche im OPAC jedoch nahezu perfekt. Da eine Klasse mehrere Dokumente zu demselben Thema vereint, ist die Suche in einem Katalog mittels der Notation besonders gut für Vollständigkeitsrecherchen (Recall) geeignet. Die SWD als Thesaurus beinhaltet spezifischeres verbales Vokabular als die RVK und kann damit der Verfeinerung einer thematischen Suche (Precision) dienen. Die Vorteile beider Verfahren sollten bei einem Retrieval also ausgenutzt werden und sich hinsichtlich ihrer verschiedenen Ordnungsmöglichkeiten und ihrer Gliederungstiefe vervollständigen.¹⁷

An dieser Stelle besteht jedoch gleichzeitig auch die Schwierigkeit einer Konkordanzstellung zwischen RVK und SWD. Die strukturellen Unterschiede der beiden Dokumentationssprachen machen eine Verknüpfung schwierig und zum Teil unmöglich. Sinnvoll erscheint vor allem die Anbindung von SWD-Begriffen an die RVK, also die verbale Erweiterung des Registers. Eine Verbindung der vergleichsweise grob gegliederten RVK-Notationen an die oft sehr speziellen SWD-Schlagwörter hingegen wäre im Hinblick auf eine exakte Konkordanz zwischen den beiden Dokumentationssprachen eher problematisch. Inwieweit diese beiden Wege anderweitig gewinnbringend, z.B. bei einer Recherche im OPAC eingesetzt werden können, soll im Verlauf dieser Arbeit erörtert werden.

¹⁷ Vgl. Geißelmann, Friedrich (1995), S. 134

4 Automatische Erschließungsverfahren

Dieses Kapitel behandelt als reines automatisches Erschließungsverfahren nur das automatische Klassifizieren. In diesem Zusammenhang entstand auch die Idee zur Erstellung von RVK-SWD-Korrelationen, die ebenfalls als Instrument der automatischen Erschließung gelten können und auf die im zweiten Teil des Kapitels eingegangen werden soll.

4.1 Automatische Klassifizierung

4.1.1 Einordnung und geschichtlicher Abriss

Eine hochwertige inhaltliche Erschließung von Publikationen in Bibliotheken erfolgt bislang hauptsächlich intellektuell, d.h. von Bibliothekaren durch eine manuelle Inhaltsanalyse. Auf Grund der steigenden Anzahl von Veröffentlichungen und begrenzter Personalkapazitäten wurden schließlich in den letzten Jahren verstärkt automatische Erschließungs- bzw. konkret Klassifizierungsverfahren entwickelt, die aber noch nicht ausgereift sind und hauptsächlich als Hilfsmittel während der intellektuellen Erschließung oder bei großen Dokumentenmengen herangezogen werden.

Automatische Klassifizierung verbindet man häufig mit Dokumenten, die im Volltext vorliegen und deren Inhalt z.B. über statistische Methoden ermittelt werden kann. Der automatische Klassifizierung von nicht-digitalen Publikationen wie im Falle von gedruckten Büchern in einem Bibliotheksbestand wurde sich bisher in wenigen Projekten gewidmet. Grund dafür dürften vor allem die zum Teil erfolglosen oder zumindest schlecht dokumentierten Ergebnisse sein. Dabei entstanden erste Ansätze zum automatischen Klassifizieren von Büchern bzw. deren bibliographischen Datensätzen schon Ende der achtziger Jahre. Eine Auswahl von Projekten soll im Folgenden vorgestellt werden.

1988 wurde erstmals in Japan das „Automatic Classification Numbering-Verfahren“ getestet. Dafür wurde eine Datenbank mit Tafel-, Hilfstafel- und Registerbegriffen der Nippon Decimal Classification angelegt, die dem Katalogisierer bei der Suche nach der passenden Klasse helfen sollte. An der University of California versuchte man 1994 bibliographische MARC-Datensätze automatisch mit LCC-Notationen anzureichern. Speziell erarbeitete LCC-Klassenbeschreibungen wurden von dem Retrievalsystem Cheshire mit den LCSH (Library of Congress Subject Headings) und den Hauptsachtiteln der Datensätze abgeglichen und klassifiziert. Die Ergebnisse zeigten, dass weniger als 50 Prozent der Dokumente der richtigen Klasse zugeordnet wurden. Häufig konnten

jedoch mögliche alternative Notationen vorgeschlagen werden und somit wurde empfohlen, das Verfahren in der Praxis semi-automatisch als Hilfsinstrument für die Bibliothekare einzusetzen. Das mitunter erfolgreichste Projekt gelang 1997 der Hongkong Polytechnic University mit dem „Automatic Classification System“. Hier wurden den Klassen der DDC neue Begriffe je nach den vorher intellektuell klassifizierten Dokumenten aus Hauptsachtiteln und Kapitelüberschriften zugewiesen. Aus dem neu zu klassifizierenden Dokument und der DDC-Klasse wurde ein Ähnlichkeitsmaß bestimmt. Zudem wurden über- oder untergeordnete Klassen auf eine stärkere Ähnlichkeit überprüft und verfolgt. Schließlich konnten bis zu 90 Prozent der Titel richtig klassifiziert werden.¹⁸

Auch in Deutschland beschäftigte man sich in den neunziger Jahren an der UB Düsseldorf mit automatischer Klassifizierung von Dokumenten. Im Projekt MILOS I (Maschinelle Indexierung zur erweiterten Literaturschließung in Online-Systemen) wurden zunächst mit dem wörterbuchbasierten Indexierungsverfahren IDX Titeldaten der UB Düsseldorf automatisch indexiert.¹⁹ Während MILOS II wurde einerseits an der Optimierung von IDX gearbeitet und andererseits wurden unter Nutzung der Daten von MILOS I Versuche zur automatischen Klassifizierung unternommen. Dafür wurden die Register der Aufstellungssystematik der UB in eine Datenbank aufgenommen und mit den indexierten Titeldaten abgeglichen. Nach einigen Tests wurde jedoch klar, dass Titelstichwörter allein nicht zur automatischen Vergabe von Notationen genügen. Aus diesem Grund wurde im Rahmen des KASCADE-Projektes an der UB Düsseldorf mit dem automatischen Verfahren THEAS (Themen-Aspekt-Identifikation) experimentiert. Hier wurde mit erweiterten Titeldaten, die z.B. einem Volltext entstammten, und statistischen Verfahren sowie den Ergebnissen aus MILOS I versucht, Dokumente zu einer möglichen Thematik zuzuordnen und eine Aspektuierung durchzuführen.²⁰

Eines der letzten größeren Projekte zur automatischen Erschließung wurde in den USA durchgeführt. Die mehr als 20.000 wissenschaftlichen und z.T. manuell von Experten gesammelten Internetquellen der virtuellen Bibliothek INFOMINE²¹ sind mit LCSH verschlagwortet. Auch in diesem Fall kam ein automatisches Verfahren zum Einsatz, das den Internetressourcen Notationen der LCC zuordnete.²²

¹⁸ Vgl. Oberhauser, Otto (2005), S. 99-104

¹⁹ MILOS I/II - Kurzbeschreibung der Projektinhalte
URL: http://www.ub.uni-duesseldorf.de/home/ueber_uns/projekte/abgeschlossene_projekte/milos/mil_kurz

²⁰ Abschlußbericht zum Projekt MILOS II
URL: http://www.ub.uni-duesseldorf.de/home/ueber_uns/projekte/abgeschlossene_projekte/milos/mil_ber

²¹ INFOMINE
URL: <http://infomine.ucr.edu>

²² Vgl. Frank, Eibe/Paynter, Gordon W. (2004), S. 214-215

4.1.2 Automatisches Klassifizieren an der UB Mannheim

Auch an der UB Mannheim wurde Literatur bislang ausschließlich manuell von Fachreferenten erschlossen. Anlässlich der Umstrukturierung des zweischichtigen Bibliotheksystems aus 15 Bereichs-, Instituts- und Fakultätsbibliotheken an verschiedenen Standorten zu einem einschichtigen System aus fünf Bibliotheksbereichen wurde eine einheitliche Aufstellungssystematik nötig.²³ In diesem Rahmen wurde ein Verfahren zur automatischen Vergabe von Notationen in der UB eingeführt. Das von Magnus Pfeffer entwickelte Programm diente als Hilfsmittel und „Vorschlagsinstrument“, um eine schnelle und effiziente Umarbeitung der Bestände zu gewährleisten und die Bibliothekare auch im Anschluss daran weiter bei der Sacherschließung zu unterstützen.

Die RVK sollte an die Stelle der verschiedenen Fachsystematiken der Teilbibliotheken treten und die Bestände unter einer gemeinsamen, universalen Klassifikation zusammenführen. Ein großer Vorteil der RVK liegt in ihrer Verbreitung im deutschsprachigen Raum. Im Bestand der UB Mannheim besaßen etwa 60% der Titel auf Grund der kooperativen Katalogisierung bereits eine RVK-Notation, die dank der Möglichkeiten zur Datenübernahme von der UB Mannheim mitgenutzt werden konnten. Die restlichen Bestände ohne RVK-Notation von circa einer halben Million Titel sollten über ein EDV-gestütztes Verfahren erschlossen werden.

Mittels der Methode des fallbasierten Schließens wurde noch nicht klassifizierten Titeln automatisch eine passende Notation zugewiesen. Dies wurde ermöglicht durch einen umfangreichen Datenbestand an bereits inhaltlich erschlossenen Dokumenten. Dementsprechend konnten nur Dokumente berücksichtigt werden, die zumindest klassifikatorisch über die RVK erschlossen waren. Diese bereits gelösten Fälle sollten nun auf neue und ähnliche „Probleme“, die unklassifizierten Dokumente, übertragen werden. Über einen Abgleich der neuen „Probleme“, der unklassifizierten Dokumente, mit der Fallbasis, den bereits klassifikatorisch erschlossenen Dokumenten, wurde die ähnlichste Lösung ausgegeben. Hier war die Lösung eine Notation.²⁴

Fallbasierte Klassifizierung bedeutet also die Übernahme einer Notation eines fremden Titels, der dem noch zu klassifizierenden Titel am ähnlichsten ist. Dabei musste davon ausgegangen werden, dass alle vorhandenen RVK-Notationen im SWB richtig und vollständig waren. Um eine Verbindung zwischen den beiden Datensätzen herzustellen und diese zu vergleichen, wurden die inhaltstragenden Elemente des Dokuments ausgewählt. Dazu zählen z.B. der Hauptsachtitel und eventuelle Zusätze sowie Schlagwörter.

²³ Vgl. Benz, Christian (2008), Folie 3-5

²⁴ Vgl. Pfeffer, Magnus (2007), S. 3-8

Im Hauptsachtitel wurden alle Wörter normalisiert und Stoppwörter eliminiert, um optimale Ergebnisse zu erzielen.²⁵

Um das Verfahren zu testen, wurden 1000 zuvor bereits manuell klassifizierte Titel zufällig ausgewählt und die Notationen aus den Daten entfernt. Diese sollten nun automatisch neu klassifiziert werden und sich mit den durch Bibliothekare intellektuell vergebenen Notationen messen.²⁶ Grundsätzlich konnte ein hoher Anteil gut klassifizierter Titel erzielt werden. Über die Hälfte der Dokumente wurde richtig klassifiziert, ein Viertel annähernd richtig und weitere 10 Prozent wurden zumindest dem richtigen Fach zugeordnet. Für einen Einsatz im OPAC sind die ausschließlich automatisch gewonnenen Notationen aber noch zu fehlerhaft, und eine Alternative zur manuellen Klassifizierung stellt das automatische Verfahren nicht dar. An der UB Mannheim diente es insbesondere als Hilfsmittel bei der Retrosystematisierung der Bestände und der Ergänzung oder Korrektur von Notationen im SWB und BVB.²⁷ Inzwischen laufen auch Versuche in der Testdatenbank des SWB und mit Daten aus dem HeBIS (Hessisches Bibliotheks-Informationssystem).²⁸

4.2 Automatisch generierte Korrelationen zwischen verschiedenen Dokumentations Sprachen

4.2.1 Korrelation von RVK und SWD

Um den Zusammenhang zwischen RVK und SWD besser untersuchen zu können, wurden im Zuge der Entwicklung des in Kapitel 4.1.2 erwähnten Verfahrens zur automatischen Vergabe von RVK-Notationen über einen Abzug aus dem SWB-Verbundkatalog automatische Korrelationen zwischen den an den Titeln vorhandenen RVK-Notationen und SWD-Schlagwörtern generiert.

Diese automatisch generierten Korrelationen entstehen also aus Titeln, die sowohl klassifikatorisch über die RVK als auch verbal über die SWD erschlossen sind. Es wird demnach eine Art künstliche Konkordanz zwischen den beiden Dokumentations Sprachen erzeugt. Wie „gut“ oder „schlecht“ die Ergebnisse sind und auch inwieweit die automatisch generierten Korrelationen mit der tatsächlichen Konkordanz von RVK-Notationen und Registerbegriffen in der RVK-Online übereinstimmen, soll im Folgenden u.a. über einen Abgleich überprüft werden.

²⁵ Vgl. Pfeffer, Magnus (2007), S. 15f

²⁶ Ebd., S. 19

²⁷ Ebd., S. 28-30

²⁸ Pfeffer, Magnus: Classification-Weblog der UB Mannheim
URL: <http://blog.bib.uni-mannheim.de/Classification/>

Vor allem aber entstehen mit den automatischen Korrelationen neue Möglichkeiten der inhaltlichen Erschließung, wenn auch nicht direkt von einem automatischen Erschließungsverfahren gesprochen werden kann. Die Korrelationen können aber beispielsweise im Sacherschließungsprozess oder bei einer Einbindung in den Online-Katalog als zusätzliches Hilfsmittel zum Einsatz kommen. In diesem Fall könnte die Heterogenität der beiden Dokumentationssprachen sogar von Vorteil sein, denn gerade bei der Recherche ergänzen sich Klassifikation und Thesaurus gut. Auf welche Weise genau die Korrelationen gewinnbringend in der Praxis Anwendung finden können, soll in Kapitel 6 überprüft werden. Zuvor muss aber zumindest stichprobenhaft eine Analyse der Ergebnisse erfolgen, um den Wert der Korrelationen festzustellen.

4.2.2 Verwandte Projekte

Die Überwindung der Grenzen zwischen Thesauri und Klassifikationen ist schon seit einigen Jahren Ziel verschiedener Projekte und Institutionen. Immer wieder entstanden Ansätze zur Einbindung thesaurusartiger Strukturen in Klassifikationen, zur Entwicklung von Thesauri mit systematischem Zugang oder zur Vermischung von verschiedenen Elementen der verbalen und klassifikatorischen Erschließung. Die Analyse und Nutzung von Korrelationen, die aus einer Datenbank oder einem Bibliothekskatalog gewonnen wurden, standen dabei jedoch bisher nur selten im Mittelpunkt. Zwei Projekte, die sich zumindest beiläufig mit diesem Thema befassten, sollen im Folgenden vorgestellt werden.

Im Rahmen des 1996 durchgeführten Projektes OSIRIS (Osnabrück Intelligent Research Information System) an der UB Osnabrück sollte die Formal- und Sachrecherche über den OPAC verbessert werden. Dazu sollte beispielsweise die Anfrage des Nutzers an den Katalog, die meist nicht den gebräuchlichen Dokumentationssprachen entspricht, auf die jeweils genutzten Klassifikationen und Thesauri abgebildet werden, um die Recherche zu erleichtern. Das bedeutete, dass rein verbale Suchanfragen auch auf die genutzten Klassifikationen übertragen werden sollten, ohne dass der Nutzer mit einer Notation in Kontakt käme. Dafür wurden aus den Klassenbenennungen der GHBS (Systematik der Gesamthochschulbibliotheken des Landes Nordrhein-Westfalen) RSWK-konforme Suchbegriffe abgeleitet. Suchte man nun mit einem einfachen Stichwort im Online-Katalog, würden nach der Recherche thematische Gruppen vorgeschlagen, die im Hintergrund über die Notationen der GHBS konstruiert wurden und für Anschlussrecherchen dienlich wären. Der Zugang zu den Publikationen wäre also rein verbal. Des Weiteren sollten verschiedene Konkordanztabellen entstehen, die automatisch aus den Titeln im OPAC generiert wurden. Zum einen wurden Konkordanzen zwischen den Notationen der GHBS und RSWK-Schlagwörtern hergestellt. Diese sollten als Registerbegriffe der Klassifikation dienen und erhielten stets einen Zähler als Signifikanzmerkmal, der angab, wie häufig die Konkordanz im OPAC auftrat. Zusätzlich wurde

eine Konkordanz zwischen LCSH und BLSH (British Library Subject Headings) mit der GHBS erzeugt. Auch zwischen den Notationen von BK (Basisklassifikation), DDC und LCC wurden Konkordanzen mit der GHBS gebildet, ebenfalls mit Zähler. Schließlich sollte noch je ein deutscher und ein englischer Wortindex erstellt werden. Diese wurden zum einen aus den Klassenbenennungen der GHBS abgeleitet und zum anderen aus den extrahierten RSWK- und LCSH- bzw. BLSH-Schlagwörtern des OPACs. Neben der Optimierung der Funktionalität des Online-Katalogs stellten die Konkordanzen aber vor allem auch ein hilfreiches Instrument bei der Sacherschließung durch die Fachreferenten dar.²⁹

Weiterhin wird von der LoC (Library of Congress) das sog. "Classification Web"³⁰ unterhalten, in dem nach allen möglichen Korrelationen zwischen LCSH, LCC und DDC (kostenpflichtig) recherchiert werden kann. Diese Korrelationen entstammen dem LoC-Katalog und werden wöchentlich aktualisiert. Es können demnach nur Korrelationen gefunden werden, die auf diese Art an mindestens einem Titel im Online-Katalog der LoC vorhanden sind. DDC-Notationen sind erst seit dem Jahr 2004 eingebunden.³¹ Classification Web ist vor allem ein Hilfsmittel für Bibliothekare bei der sachlichen Erschließung. So können passende Schlagwörter bzw. Subject Headings zu einer Notation gefunden werden oder Publikationen mit Notationen der anderen Systematik versehen werden, wenn in der Bibliothek offiziell nur mit einer Klassifikation erschlossen wird. Umgekehrt kann auch von einem Subject Heading ausgehend nach einer passenden Notationsstelle gesucht werden. Die Korrelationen sind wie bei dem Projekt Osiris mit einem Schwellenwert verknüpft, der angibt, wie oft die Korrelation im Katalog der LoC auftritt. Generell sind alle drei Dokumentations Sprachen an verschiedenen Stellen im System immer wieder miteinander verknüpft, so dass ein Browsen ermöglicht wird. Von Nachteil ist, dass das Tool nahezu ausschließlich auf den Einsatz von Bibliothekaren ausgelegt ist. Für die Benutzerseite dagegen scheinen die automatischen Korrelationen keinen Vorteil mitzubringen.³²

²⁹ Vgl. Recker, Ingrid/Ronthaler, Marc/Zillmann, Hartmut (1996), S. 839-847

³⁰ Classification Web
URL: <http://classificationweb.net/>

³¹ Classification Web: Quick Start Tutorial ; LC/Dewey Correlations
URL: <http://classificationweb.net/tutorial/9dewey.html>

³² Classification Web: Quick Start Tutorial ; Subject Heading & Classification Number Correlations
URL: <http://classificationweb.net/tutorial/7subjcor.html>

5 Korrelationsanalyse RVK-SWD

5.1 Vorbereitungen

Um eine vollständige Korrelationsanalyse zwischen den Notationen der RVK und den Schlagwörtern der SWD durchzuführen, wurde zunächst je ein Abzug der RVK-Online und ein Abzug des SWB erstellt. Diese zwei Dateien bildeten den Ausgangspunkt für die Untersuchung. Da später ein direkter Vergleich zwischen den Registerbegriffen der RVK-Online und den SWD-Schlagwörtern des Verbundabzuges vollzogen werden sollte, wurden eventuelle Schlagwortketten in beiden Fällen aufgebrochen. Das führte in den Dateien meist dazu, dass einer Notation mehrere Schlagwörter zugeordnet waren.

Die Datei der RVK-Online beinhaltet die Beziehung zwischen RVK und SWD und stellte somit die Konkordanz der beiden Dokumentations Sprachen dar (Tabelle 2).

RVK-Notation	Schlagwort	Übergeordnete Klassenbenennungen³³
CT 8500	Fragebogen	<i>Psychologie / Tests / Verschiedene Fragebogeninstrumente</i>
CT 8500	Psychologie	<i>Psychologie / Tests / Verschiedene Fragebogeninstrumente</i>
CU 1000	Einführung	<i>Psychologie / Klinische Psychologie / Einführung, Grundfragen, Lehrbücher, Kompendien</i>
CU 1000	Klinische Psychologie	<i>Psychologie / Klinische Psychologie / Einführung, Grundfragen, Lehrbücher, Kompendien</i>
CU 2000	Psychoanalyse	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse</i>
CU 2000	Tiefenpsychologie	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse</i>
CU 2500	Psychoanalyse	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse / Adler, Alfred / Gesamtausgaben</i>
CU 2500	Tiefenpsychologie	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse / Adler, Alfred / Gesamtausgaben</i>
CU 2501	Psychoanalyse	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse / Adler, Alfred / Auszüge; Teilausgaben</i>
CU 2501	Tiefenpsychologie	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse / Adler, Alfred / Auszüge; Teilausgaben</i>
CU 2502	Psychoanalyse	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse / Adler, Alfred / Einzelwerke</i>
CU 2502	Tiefenpsychologie	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse / Adler, Alfred / Einzelwerke</i>
CU 2503	Psychoanalyse	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse / Adler, Alfred / Sekundärliteratur</i>
CU 2503	Tiefenpsychologie	<i>Psychologie / Klinische Psychologie / Tiefenpsychologie und Psychoanalyse / Adler, Alfred / Sekundärliteratur</i>

Tabelle 2: Beispiele für Konkordanzen aus der RVK-Online

³³ Die Klassenbezeichnungen wurden in allen Tabellen manuell eingefügt und sind so nicht in der Datei enthalten.

Wenn der Blattknoten selbst keinen Registerbegriff aufwies, wurde der ihm als nächstes übergeordnete Registereintrag gewählt. So besitzen beispielsweise die Notationen CU 2500, CU 2501, CU 2502 und CU 2503 in der Konkordanzdatei die gleichen Schlagwörter wie die übergeordnete Notation CU 2000: Psychoanalyse || Tiefenpsychologie. Von der RVK-Online wurde ein Abzug aus dem Jahr 2007 und ein Abzug aus dem Jahr 2009 zur Verfügung gestellt.

Die Korrelationsdatei wiederum enthält die aktuellen Rohdaten aus dem SWB, also die SWD-Schlagwörter der jeweiligen Titel mit den entsprechend korrelierenden RVK-Notationen. Es wurden folglich nur Titel einbezogen, die sowohl eine RVK-Notation als auch mindestens ein Schlagwort besaßen. Eine Ausnahme bildeten mehrbändige Werke, bei denen die RVK-Notation und die Schlagwörter vom übergeordneten Datensatz auch auf die einzelnen Bände übertragen wurden, da die Gesamtaufnahme nicht die reale Anzahl der Titel repräsentiert und die einzelnen Stücktitel meist keine zusätzliche inhaltliche Erschließung aufweisen. Weiterhin wurden in den Daten alle ungültigen RVK-Notationen, die z.B. veraltet waren oder Tippfehler aufwiesen, entfernt und auch Form-, Zeit- und Körperschaftsschlagwörter ausgewertet. Ein Datensatz besteht hier im Vergleich zur Datei mit der RVK-Online-Konkordanz nicht nur aus der Verbindung zwischen Notation und dem zugehörigen Schlagwort (Tabelle 3). Neben der PPN (Pica Production Number) des Schlagworts sind außerdem drei weitere Elemente vorhanden. Das Feld „Anzahl Titel mit RVK-Notation“ gibt die Anzahl der Titel an, die diese RVK-Notation im SWB aufweisen. Das nachfolgende Feld verzeichnet die Anzahl der Titel mit eben dieser Notation, die außerdem das entsprechende Schlagwort besitzen. Aus diesen beiden Angaben wird ein dritter Wert berechnet, der Schwellenwert. Dieser gibt mit einem Wertebereich zwischen Null und Eins die Wahrscheinlichkeit an, mit der ein Schlagwort zusammen mit einer Notation vorkommt. Je höher der Wert desto wahrscheinlicher ist demzufolge die Korrelation. Am Beispiel der Tabelle 3 bedeutet dies, dass die Notation BC 8770 insgesamt an 61 Titeln im SWB vorhanden ist, die zudem gleichzeitig verbal erschlossen wurden. Von diesen wiederum besitzen 49 Publikationen das Schlagwort „Atlas“. Somit liegt die Wahrscheinlichkeit, dass BC 8770 zusammen mit dem Schlagwort „Atlas“ auftritt, bei etwa 80 Prozent.

RVK-Notation	Schlagwort	PPN Schlagwort	Anzahl Titel mit RVK-Notation	Davon Titel mit Schlagwort	Schwellenwert	Hauptgruppe
BC 8770	Atlas	208834664	61	49	0.80328	<i>Theologie und Religionswissenschaften</i>
WC 1000	Biologie	208867503	125	125	1.00000	<i>Biologie</i>
ZN 4850	Transistor	209136111	94	53	0.56383	<i>Technik</i>
HD 232	Wirtschaftssprache	209511818	1418	741	0.52257	<i>Anglistik. Amerikanistik</i>
US 6100	Sonne	20911441X	33	25	0.75758	<i>Physik</i>

Tabelle 3: Beispiele für automatisch generierte Korrelationen aus dem SWB

Des Weiteren wurden von Magnus Pfeffer, Fachreferent für die Fächer Informatik, Mathematik, Naturwissenschaften und Technik an der UB Mannheim, zwei kurze Programme in Perl entwickelt, die eine genauere Analyse der genannten Daten ermöglichen sollten.

Das erste Programm diente der Filterung bzw. Eingrenzung der Daten auf Grund ihrer sonst zu unübersichtlichen Größe. Mit Hilfe des Programms konnte zum einen die Mindestanzahl der Titel, die eine RVK-Notation besitzen, angegeben werden. So konnten etwa Notationen herausgefiltert werden, die nur wenige Male auftraten und möglicherweise nicht ausreichend statistisch signifikant für die Analyse gewesen wären. Ferner konnte eigenständig ein bestimmter Schwellenwert festgelegt werden, um Korrelationen mit niedrigem Quotienten zu entfernen und nur bedeutende Verbindungen von Notation und Schlagwort auszugeben.

Ein zweites Programm führte schließlich einen Abgleich zwischen den Daten der RVK-Online (Konkordanzen) und den Rohdaten aus dem SWB (Korrelationen) durch. Dabei wurden jeweils die Notationen und entsprechenden Schlagwörter der beiden Dateien eingelesen und verglichen. Daraufhin entstanden drei unterschiedliche Gruppen. Wenn die gleiche Verbindung von Notation und Schlagwort in den Daten der RVK-Online wie auch in den Daten des SWB gefunden werden konnte, entstand eine erste Datei, die folglich die Übereinstimmungen aus den beiden Ausgangsdateien vereinte und damit eine Schnittmenge abbildete. Die Korrelationen, die in den Verbunddaten auftauchten, nicht aber solchermaßen als Konkordanz in den Daten der RVK-Online vorhanden waren, wurden in eine zweite Datei geschrieben. Die dritte Datei beinhaltete die übrig gebliebenen Werte, d.h. die Konkordanzen von RVK-Notation und Schlagwort, die in der RVK-Online auftauchten, jedoch nicht in dieser Konstellation im Verbund vorhanden waren.

5.2 Durchführung der Analyse

Für die Analyse der Daten wurden die in Kapitel 5.1 genannten Dateien und Programme herangezogen, und diese zum einen über Cygwin/Unix und zum anderen mit Microsoft Excel bearbeitet. Mittels des Filter-Programms wurden Dateien mit verschiedenen Schwellenwerten und Mindest-Titelanzahl erzeugt und dokumentiert.

Die Analyse der automatisch generierten Korrelationen und der drei neu entstandenen Dateien sollte stichprobenartig durchgeführt werden. Dabei war die Gruppe, die die Schnittmenge von Konkordanz und Korrelation beinhaltete, vor allem wichtig, um einen möglichst objektiven Schwellenwert zu bestimmen und anschließend mit den gewünschten Daten weiterzuarbeiten. Die beiden übrigen Dateien jedoch waren interessanter, weil über sie mögliche neue Schlagwörter, entweder als Anreicherung für den SWB oder als Optimierung der RVK, ausfindig gemacht werden konnten. Inwiefern

diese Begriffe in der Praxis tatsächlich geeignet sein würden, z.B. in Abhängigkeit von ihrem Schwellenwert, sollte während der Analyse überprüft werden.

Zudem wurden im Rahmen der Auswertung die Größen der drei Gruppen ermittelt. Die erste Erwartung war, dass die Datei, die die Übereinstimmung von Korrelation und Konkordanz enthält, am größten sein würde, geht man von einer beidseitigen hochwertigen verbalen und klassifikatorischen Erschließung aus. Die beiden anderen Dateien sollten also übersichtliche Werte annehmen.

Ferner sollten auch die Größen der Fachsystematiken innerhalb der drei verschiedenen Dateien ermittelt werden, um z.B. festzustellen, wo Konkordanzen oder Korrelationen besonders ausgeprägt sind und welche Register in der RVK dahingehend noch überarbeitet oder erweitert werden könnten. Während der Analyse kam z.B. zu Tage, dass fast 2/3 der Konkordanzen aus der RVK-Online der Gruppe R – Geographie angehören (Abbildung 4). Um die Datenmengen zu reduzieren und die anderen Gruppen besser analysieren zu können, wurde diese Fachsystematik in einigen Fällen aus den Ergebnissen entfernt. Die Ursache für den enormen Umfang liegt einerseits an den sehr breit und tief gegliederten Klassen und den entsprechend zahlreichen Notationen.³⁴ Außerdem kommen in R – Geografie zahlreiche Schlüsselungen zum Einsatz. In der gedruckten RVK-Version der Gruppe R³⁵ findet man zunächst mehrere Seiten des Sachschlüssels S1R vor. Im sich anschließenden Hauptteil ist dann an jeder einzelnen Stelle in Rot der Hinweis „+S1R“ vorhanden. Die Notationen sind folglich alle geschlüsselt. Grundsätzlich kann jede Stelle mit jedem Schlüssel kombiniert werden. Bei über 300 Schlüsseln und mehr als 400 Systemstellen ergeben sich also über 120.000 Notationen. Andererseits befinden sich an den Klassen oft unzählige Registerbegriffe. Die Fachsystematik Geographie wurde als eine der ersten mit SWD-gerechten Registerbegriffen versehen.³⁶ Eine Analyse der Konkordanzen in der RVK-Online ergab zudem, dass in der Gruppe R durchschnittlich weitaus mehr Registerbegriffe pro Notation vorhanden sind als in den übrigen Fachsystematiken. Aus einem Protokoll der „Arbeitsgruppe Klassifikation der Kommission für Erschließung“ vom 16.01.2006 geht beispielsweise hervor, dass die Notation RC 70235 mit 210 Ketten die höchste Anzahl an Schlagwörtern und –ketten in

³⁴ Als Beispiel sei hier die Notation RF 70726 angeführt: R – Geographie / RC – RZ – Regionale Geographie / RC 10000 – Europa / RC 15000 – Mitteleuropa / RC 20000 – Deutschland / RC 25000 – Deutschland (Westliche Länder) / RE 10000 – Süddeutschland / RF 10000 – Bayern / RF 40000 – Bayern <Süd> (Südbayern) / RF 70000 - RF 70918 – Moränenlandschaften des diluvialen Lechgletschers, Isargletschers und Loisachgletschers (Iller/Lech- und Loisach/Isar-Gletschers) mit Ammergau und Starnberger See <Region> (Würmseebcken) / RF 70103 - RF 70850 – Allgemeine Geographie / RF 70540 - RF 70850 – Anthropogeographie / RF 70645 - RF 70789 – Wirtschafts-, Handels- und Verkehrsgeographie / RF 70654 - RF 70735 – Wirtschaftsgeographie / RF 70714 – Industriegewirtschaftsgeographie / RF 70720 – Industriezweige / RF 70726 – Konsumgüterindustrie

³⁵ Regensburger Verbundklassifikation : 16 ; Geographie (R)
URL: http://www.bibliothek.uni-regensburg.de/rvko/pdf/rvko_R.pdf

³⁶ Rundbrief zur Regensburger Verbundklassifikation (1997), S. 3
URL: http://www.bibliothek.uni-regensburg.de/Systematik/rundbriefe/1997_01.pdf

der RVK aufweist.³⁷ Darin eingeschlossen waren nur einige der möglichen Permutationen. Inzwischen sind diese Registerbegriffe aber überarbeitet worden.

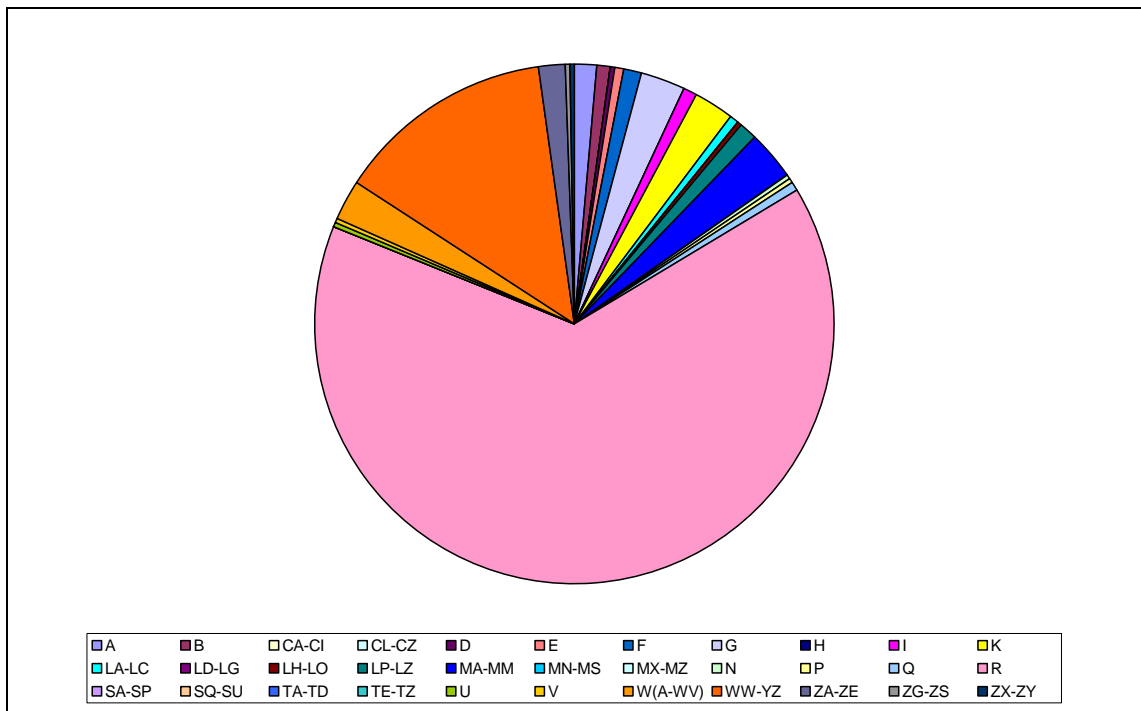


Abbildung 4: Umfang der RVK-Konkordanzen in den einzelnen Fachsystematiken

Zusätzlich wurde neben einem aktuellen Abzug der RVK-Online aus dem Jahr 2009 ein Abzug von 2007 zur Verfügung gestellt. Auf diese Weise sollte überprüft werden, ob in den zwei Jahren sichtbare Veränderungen vorgenommen wurden. Schließlich sind Dokumentations Sprachen ständigen Erweiterungen und Korrekturen unterworfen und nicht nur aufwendig zu erstellen, sondern auch zu pflegen. Es stellte sich jedoch heraus, dass die zwei Abzüge sich nur unerheblich voneinander unterschieden. Auf eine ausführlichere Analyse wurde deswegen verzichtet.

5.3 Ergebnisse

5.3.1 Automatisch generierte Korrelationen zwischen RVK und SWD

Zunächst wurden nur die automatisch generierten Korrelationen aus RVK-Notation und SWD-Schlagwort aus dem SWB stichprobenartig analysiert. Ziel war es, unterschiedliche Korrelationstypen zu ermitteln und zu bewerten, z.B. in Hinblick auf eine weitere Nutzung. Für diese Analyse wurde eine Datei gewählt, in der die Korrelationen einen

³⁷ Arbeitsgruppe Klassifikation der Kommission für Erschließung (2006), S. 6
 URL: http://www.bibliothek.uni-regensburg.de/Systematik/protokolle/2006_01_16.pdf

Mindest-Schwellenwert von 0,3 aufwiesen und in der pro Notation mindestens zehn Publikationen vorhanden waren. Trotz dieser Einschränkungen beinhaltete die Datei immer noch mehr als 60.000 Verbindungen.

Während der Auswertung wurde festgestellt, dass z.B. bestimmte Schlagworttypen weniger sinnvoll oder statistisch signifikant sind als andere. Dazu gehören z.B. Formschlagwörter, denn sie beschreiben ein Dokument nicht nach inhaltlichen, im Sinne von thematischen, Aspekten, sondern nach der Form der Publikation. Bei einer Recherche mit kontrolliertem Schlagwortvokabular in einem Bibliothekskatalog würde man ein Formschlagwort deswegen vermutlich nur in Zusammenhang mit einem anderen inhaltstragenden Schlagwort benutzen. Zudem werden die Formschlagwörter in den Konkordanzen der RVK-Online nicht konsequent eingesetzt.

So finden formale Aspekte in der RVK meist über Schlüssel Anwendung, wie im Notationsbereich „XD 2801 - XD 2828 – Formalschlüssel“ der Fachsystematik Medizin. Dabei sind die letzten zwei Ziffern ausschlaggebend für die Zuordnung zu einer bestimmten Dokumentart. Das Schlagwort „Kongress“ korreliert u.a. mit den Notationen XD 2805, XD 3105 und XD 3205 aus der Systematik „XD – Medizinische Mikrobiologie“ mit Schwellenwerten zwischen 0,6 und 0,9. Der Schlüssel 05 entspricht hier auch genau der Klasse für Kongresse, Tagungsberichte, Lehrgänge und Symposien. Würde man jedoch in der RVK-Online all diesen Notationen den einzelnen Begriff „Kongress“ zuordnen, entstünde schnell ein sehr umfangreiches und unübersichtliches Register. Darüber hinaus gibt es in der Klasse XD weitere formale Kategorien für Kongresse, z.B. XD 4205, XD 5205, XD 6205 u.ä., die allerdings, nähme man den Schwellenwert als Indikator für die Bedeutung einer Korrelation, wegen eines zu geringen Schwellenwertes in den automatischen Korrelationen nicht sofort signifikant zu sein scheinen. Daher stellt sich die Frage, ob Formalschlüssel überhaupt mit Registerbegriffen erschlossen werden sollten.

Zudem konnte bei einer Analyse einzelner Schlagworttypen in den Korrelationen festgestellt werden, dass bei hohen Schwellenwerten vor allem geographische, Personen- und Formschlagwörtern sowie Sprachen vorkommen (Tabelle 4). Bei einem Schwellenwert von 1,0 und einer Mindestanzahl von 20 Titeln pro Notation war durchschnittlich nur jedes vierte Schlagwort ein Sachschlagwort. Nach stichprobenartigen Untersuchungen fanden sich unter diesen Sachschlagwörtern wiederum insbesondere Allgemeinbegriffe wie „Biologie“, „Geologie“, „Literatur“, „Organische Chemie“ oder „Pflanzen“. Enge Schlagwörter waren zunehmend bei kleiner werdendem Schwellenwert vorhanden. Außerdem stieg die Anzahl der Sachschlagwörter bei sinkendem Schwellenwert an und die anderen Schlagworttypen nahmen parallel dazu ab. Hierin spiegelt sich der generelle Unterschied zwischen einer Klassifikation und einem Thesaurus wider. Spezifischere Schlagwörter treten eher selten an den breit angelegten Notationen und in hohen Korrelationsbereichen auf. In den niedrigeren Schwellenwertbe-

reichen dagegen kommen sie häufiger vor, jedoch nur mit einem geringen Quotienten auf Grund der zahlreichen verwandten Begriffe, die nicht immer eindeutig und übergreifend auf alle Publikationen einer Notation angewandt werden können. Wenn eine hohe Korrelation bei einem engen Schlagwort gemessen werden kann, dann ist die Systematikstelle oft sehr fein gegliedert. Formschlagwörter hingegen treten offensichtlich relativ unbeeinflusst vom Schwellenwert an allen denkbaren Korrelationen auf, denn sie sind unabhängig von der Spezifität des Themas und „kennzeichnen die Erscheinungsweise, die Art der Darstellung, die physische Form eines Dokuments (den Dokumenttyp) sowie in Einzelfällen das Niveau der Darstellung“.³⁸

Schwellenwert	Personen-Schlagwörter	Geographische Schlagwörter	Form-Schlagwörter	Sprachen-Schlagwörter	Rest (z.B. Sachschlagwörter)
0,2	2 %	12 %	23 %	2 %	61 %
0,58<=x<0,62	5 %	15 %	15 %	7 %	58 %
1,0	18 %	16 %	26 %	17 %	23 %

Tabelle 4: Verteilung einzelner Schlagworttypen in den Korrelationen

Nicht alle dieser Korrelationen sind jedoch gleich „gut“ oder „schlecht“. Der Wert einer Korrelation lässt sich vermutlich oft an der Höhe des Schwellenwertes festmachen. Ein hoher Quotient steht demzufolge für häufiges gemeinsames Auftreten der Beziehung aus Notation und Schlagwort. Daraus ließe sich pauschal schließen, dass die Korrelation besonders „gut“ ist. Umgekehrt würde ein niedriger Schwellenwert eine „schlechte“ Korrelation indizieren. Im Folgenden soll diese These anhand einiger Beispiele geprüft werden.

Gute Korrelationen lassen sich etwa an dem Schlagwort „Tragik“ finden. Es korreliert mit der Notation EC 3970 aus der Fachsystematik Literaturwissenschaft und hier konkret dem Bereich „Stilformen / Tragik“ in 12 von 12 Fällen, d.h. mit einer Wahrscheinlichkeit von 100 Prozent. Das bedeutet, dass im SWB alle Dokumente mit dieser Notation auch das passende Schlagwort „Tragik“ erhielten. Auch im Vergleich mit der RVK-Online ist diese Korrelation sehr gut, denn die Systemstelle entspricht genau dieser Stilform und das Schlagwort ist dort auch als Registerbegriff vorhanden. Weniger gut dagegen korreliert „Tragik“ mit der Notation „EC 4730“ für Publikationen zum Thema „Tragödie“, denn nur 11 von 36 Publikationen wurden hier mit dem Begriff

³⁸ Regeln für den Schlagwortkatalog: Formschlagwörter : § 501 – Definition
 URL: http://deposit.ddb.de/ep/netpub/89/96/96/967969689/_data_stat/www.dbi-berlin.de/dbi_pub/einzelpu/regelw/rswk/rswk_07.htm

verschlagwortet. Das entspricht einem prozentualen Anteil von etwa 30 Prozent und bedeutet, dass nicht alle Dokumente zur „Tragödie“ ebenfalls direkt den Bereich Tragik behandeln (Tabelle 5). Eine Erklärung dafür lässt sich z.B. bei näherer Betrachtung der Begriffe in der SWD und deren Beziehungen zueinander ableiten, denn „Tragik“ ist der Oberbegriff zur „Tragödie“, womit „Tragödie“ als noch spezifischeres bzw. engeres Schlagwort gelten kann als „Tragik“.

Titel:	<u>Tragedy : a short introduction</u> / Rebecca Bushnell
RVK-Notation:	<u>EC 4730</u> Allgemeine und vergleichende Sprach- und Literaturwissenschaft. Indogermanistik. Außereuropäische Sprachen und Literaturen/Allgemeine und vergleichende Literaturwissenschaft./Poetik/Gattungspoetik/Drama/Tragödie/Allgemeines
Schlagwortkette:	* <u>Tragödie</u> ; <u>Einführung</u>
Titel:	<u>Tragedy and tragic theory : an analytical guide</u> / Richard H. Palmer
RVK-Notation:	<u>EC 7510</u> Allgemeine und vergleichende Sprach- und Literaturwissenschaft. Indogermanistik. Außereuropäische Sprachen und Literaturen/Allgemeine und vergleichende Literaturwissenschaft./Literaturgeschichte/Geschichte der Literaturgattungen (literarischen Gattungen)/Drama/Tragödie/Allgemeines/Gesamtdarstellungen <u>EC 4730</u> Allgemeine und vergleichende Sprach- und Literaturwissenschaft. Indogermanistik. Außereuropäische Sprachen und Literaturen/Allgemeine und vergleichende Literaturwissenschaft./Poetik/Gattungspoetik/Drama/Tragödie/Allgemeines
Schlagwortkette:	* <u>Tragödie</u>
Titel:	<u>Tragische Handlungsstrukturen : eine Theorie der Tragödie</u> / Rolf Breuer
RVK-Notation:	<u>EC 4730</u> Allgemeine und vergleichende Sprach- und Literaturwissenschaft. Indogermanistik. Außereuropäische Sprachen und Literaturen/Allgemeine und vergleichende Literaturwissenschaft./Poetik/Gattungspoetik/Drama/Tragödie/Allgemeines
Schlagwortkette:	* <u>Tragödientheorie</u> * <u>Tragödie</u>

Tabelle 5: Titel aus dem SWB mit der Notation EC 4730, die nicht mit dem Schlagwort "Tragik" korrelieren

In vielen Fällen lässt sich zudem an Korrelationen mit hohen Schwellenwerten das Sachgebiet ablesen. Die beiden Schlagwörter „Ostafrika“ und „Pflanzen“ korrelieren beispielsweise mit der Notation „WL 1217“ bei 227 Titeln im SWB zu etwa 100 Prozent. Daraus ließe sich schließen, dass es sich um eine Systemstelle für Pflanzen in Ostafrika handeln muss. Und tatsächlich lässt sich in der RVK-Online über die Hierarchiestufen genau das Thema Botanik in Ostafrika eingrenzen: „W – Biologie / WL – Spezielle Botanik (Systematik) / WL 1101 - WL 1299 – Floren und Bestimmungsbücher in regionaler Begrenzung / WL 1210 – Afrika / WL 1217 – Ostafrika insgesamt“.

Dagegen bedeuten niedrige Korrelationen, dass ein Schlagwort nur selten an der entsprechenden Notation im SWB vorhanden ist und dementsprechend nur eine vage Verbindung zwischen der Systemstelle und dem verbalen Erschließungsbegriff hergestellt werden kann, so z.B. im Fall der Notation „AC 13000“ und dem Schlagwort „Alche-

mie“. Zwischen ihnen besteht eine Korrelation mit einem Schwellenwert von 0,3. Dies ist wenig erstaunlich, denn die Klasse „AC 13000“ beinhaltet eigentlich Bibliographien der Naturwissenschaften, Medizin und Technik. „Alchemie“ als Registerbegriff ist aber nur in der Systematik für Ethnologie im Notationsbereich „LC 53000 - LC 53695 – Alchemie“ und in der Fachsystematik Chemie und Pharmazie bei „VB 2330 - VB 2339 – Alchemie“ vorhanden. Die Korrelation ist also im Vergleich zur RVK-Online „schlecht“, aber immer noch vergleichsweise hoch. Das fällt vor allem dann auf, wenn man die anderen Schlagwörter, die in der Korrelation von „AC 13000“ vorkommen, betrachtet. „Medizin“ (Schwellenwert 0,1), „Naturwissenschaften“ (0,4) und „Technik“ (0,2) scheinen hinsichtlich ihres Schwellenwertes teilweise sogar weniger bedeutend zu sein als „Alchemie“, dabei tauchen sie sogar als Registerbegriffe an „AC 13000“ auf. Daher wäre zu schlussfolgern, dass die Notation um den Registerbegriff „Alchemie“ ergänzt werden müsste oder, da alle Titel mit der Korrelation aus „AC 13000“ und „Alchemie“ im SWB eine zweite RVK-Notation „VB 2334 - Bibliographien, Dokumentationen“ aus dem Bereich Alchemie besitzen, hier möglicherweise eine überflüssige Doppelstelle vorliegt. Hinzu kommt, dass unter „VB 2334“ eine weitere englischsprachige Publikation gefunden wird, die die Zweitnotation „AC 13000“ nicht besitzt, und die eben formulierte Vermutung weiter bestätigt.

Hohe Korrelationen sind demnach nicht prinzipiell „gute“ oder niedrige „schlechte“ Korrelationen, und nicht immer kann die Bedeutung der Verknüpfung von Notation und Schlagwort aus dem Schwellenwert abgelesen werden. Obwohl etwa der Bereich „LH 40000 - LH 49910“ internationale Kunstaussstellungskataloge umfasst, besitzen im SWB nicht alle Titel in diesem Notationsbereich das Schlagwort „Ausstellung“ und damit einhergehend einen entsprechend hohen Schwellenwert. Ein Viertel der Korrelationen weist sogar einen Schwellenwert von unter 0,75 auf. „LH 44330“, eine Klasse für Kunstaussstellungen in Kassel, hat z.B. nur einen Schwellenwert von 0,32. Betrachtet man einige Dokumente unter dieser Notation im Verbundkatalog genauer, stellt man fest, dass sich vor allem Publikationen zur „Documenta“ finden lassen. Die Documenta ist eine regelmäßig stattfindende, große Kunstaussstellung in Kassel. Aus diesem Grund wird das Körperschaftsschlagwort „Documenta“ in Verbindung mit weiteren Zeit- und Ortsangaben angewandt und nicht das vergleichsweise weite Schlagwort „Ausstellung“, wie es bei Publikationen anderer, kleinerer Ausstellungen der Fall ist. Daher liefert die Beziehung „LH 44330 – Ausstellung“ in diesem Fall nur einen geringen Schwellenwert, obwohl erwartungsgemäß eine „gute“ Korrelation zustande kommen müsste. Hier spiegelt sich also die Problematik der Verschiedenheit der beiden Dokumentationssprachen wider. Während in der RVK eher Allgemeinbegriffe ins Register aufgenommen werden, besteht mittels verbaler Erschließung die Möglichkeit Individualbegriffe einzusetzen.

Ein weiteres Beispiel für einen geringen Schwellenwert, der dennoch eine „gute“ Korrelation repräsentiert, stellt die Klasse „WM 5000 – Embryologie“ von Pflanzen der Fach-

systematik Biologie dar. Hier wird nur ein Schwellenwert von 0,3 zwischen der Notation „WM 5000“ und dem Schlagwort „Embryologie“ erzielt. Obwohl alle Titel im SWB sich generell mit der Embryonalentwicklung beschäftigen, wurden hier viele verschiedene verwandte Schlagwörter vergeben. Neben „Embryologie“ existieren z.B. noch „Phylogenie“, „Ontogenie“, „Entwicklungsbiologie“, „Embryonalentwicklung“ oder „Molekulargenetik“. Recherchierte man im Verbundkatalog in der erweiterten Suche, also nur mit „WM 5000“ und „Embryologie“, würden vermutlich relevante Titel aus dem Bereich Pflanzenembryologie auf der Strecke bleiben. Diese Schlagwörter könnten also durchaus auch für die Verbesserung der Thesauri-Relationen genutzt werden.

Umgekehrt lassen sich immer wieder Korrelationen mit hohen Schwellenwerten finden, wo die Konstellation aus Notation und Schlagwort zunächst ungewöhnlich zu sein scheint. „CE 6800“ beispielsweise ist die Systemstelle für Gesamtwerke Johannes Keplers und korreliert mit den Sachschlagwörtern „Optik“ und „Astronomie“ zu 100 Prozent. Die RVK-Klassen „CA – CI“ umfassen das Fach Philosophie, ein derart thematisch verwandtes Schlagwort wie „Philosophie“ ist an der Notation „CE 6800“ im SWB jedoch nicht zu finden. Da Johannes Kepler sich aber mit allen drei Bereichen beschäftigte und Optik, Astronomie und Philosophie übergreifend erforschte, erscheint der Zusammenhang nachvollziehbar. Weiterhin sollte überlegt werden, ob die Korrelation nicht sinnvoll genutzt werden könnte, indem man dem RVK-Notationsbereich einer Person nicht nur das entsprechende Personenschlagwort zuordnet, sondern z.B. auch Sachschlagwörter zu dem Gebiet, mit dem sich jene Person befasst hat.

Im Katalog können demzufolge sehr spezielle Korrelationen vorkommen, die in der RVK-Online auf diese Weise nicht möglich sind und zunächst missverständlich erscheinen. Die Theorie der RVK wird mitunter nicht in der Praxis im SWB abgebildet. Der Schwellenwert kann demnach nur teilweise ein Indikator für die Bedeutung der automatisch generierten Korrelationen sein.

In eine andere Richtung geht die Überlegung, Notationen genauer zu betrachten, die überdurchschnittlich viele Titel aufweisen. Aus einer hohen Anzahl von Titeln an einer einzelnen Notation kann möglicherweise geschlossen werden, dass die Klasse weiter untergliedert werden müsste und nicht mehr aktuell oder zu allgemein ist. Gerade in den Naturwissenschaften, wo durch Forschung und Entwicklung immer neue Wissensbereiche in den Vordergrund treten, sollten regelmäßige Erweiterungen der Systematik erfolgen. Etwa im Bereich der Informatik wäre es denkbar, Notationen nach weiteren inhaltlichen oder formalen Gesichtspunkten zu untergliedern. Einige Systemstellen aus dem Bereich der Programmiersprachen (ST 250) weisen zum Teil mehrere hundert oder tausend Titel im SWB auf. Die Übersichtlichkeit der Titel für den Nutzer wird damit nicht nur am Katalog erschwert, sondern auch vor Ort beim Browsen am Regal. Für die Notation „ST 250 J35“ der Programmiersprache „Java“ werden z.B. fast 1000 Titel im SWB gefunden, mehrbändige Werke nicht eingeschlossen. Hier könnte also beispielsweise

eine Untergliederung nach formalen (Einführungen, Gesamtdarstellungen, Übungsbücher etc.) oder weiteren sachlichen Aspekten (Java, Javascript) vorgenommen werden. Auch im Bereich „Datenverarbeitung im Bibliothekswesen“ (AN 73000) könnte über eine Erweiterung nachgedacht werden, denn zu diesem Blattknoten lassen sich im SWB über 400 Publikationen finden. Darunter fallen z.B. Themen wie Web 2.0, Digitalisierung oder Virtuelle Bibliotheken, die als Unterpunkte im Hinblick auf moderne Trends und Entwicklungen in diesem Bereich durchaus sinnvoll wären. Diese Überarbeitungen können insbesondere in Spezial- oder großen Universalbibliotheken bei der Aufstellung zugunsten aktueller Wissenschaftsbereiche sinnvoll sein. Natürlich kann eine Klassifikation es aber nicht leisten, für jede Art von Publikation eine exakte Systemstelle zur Verfügung zu stellen.

Abschließend brachte eine Analyse zum Umfang der Schlagwörter an den Notationen nochmals den generellen Unterschied zwischen klassifikatorischer und verbaler Sacherschließung zutage. Von etwa 239.000 gezählten verschiedenen Notationen mit über einer Million zugehörigen Registerbegriffen in der RVK-Online fanden rund 120.000 Notationen Anwendung im SWB. Zu diesen Notationen kamen im Verbundkatalog mehr als 3 Millionen verschiedene Schlagwörter der Sacherschließung. Während in der RVK pro Notation rund 4 Registerbegriffe vorhanden sind, findet man an den Titeln einer Notation im SWB im Durchschnitt 27 verschiedene Schlagwörter. Eine Konkordanz der beiden Dokumentationssprachen erscheint bei diesen Differenzen zunächst kaum erreichbar. Dennoch sind regelmäßige Überarbeitungen und Weiterentwicklungen wie etwa der Anreicherung der RVK-Register mit SWD-Vokabular unabdingbar.

5.3.2 Übereinstimmungen von Konkordanz und Korrelation

Entgegen den Erwartungen enthält die Datei mit den Übereinstimmungen zwischen den Konkordanzen aus der RVK-Online und den Korrelationen aus dem SWB nicht die Mehrheit der Datensätze, sondern nur einen geringfügigen Anteil. Unter Einbezug aller Fachsystematiken der RVK sind nur ca. ein Prozent der Kombinationen von Notation und Schlagwort sowohl in der RVK-Online vorhanden als auch im SWB. Obwohl dieses Ergebnis zunächst überraschen mag, erscheint es nach Betrachtung der anderen Daten (Kapitel 5.3.3 und 5.3.4) wiederum plausibel. Hier spiegelt sich der grundsätzliche Unterschied der beiden Dokumentationssprachen wider. Während im SWB vielfältige Dokumente erschlossen werden und man dafür häufig sehr spezifische bzw. enge Schlagwörter verwendet, sind die Notationen der RVK vielmehr allgemein bzw. weit gefasst. Aus diesem Grund kann es nur wenige Übereinstimmungen zwischen der Korrelation und der Konkordanz geben.

Die in der Datei enthaltenen Datensätze sind daher unterschiedlichster Art. Der Notation „ST 250 P42 – Programmiersprachen P“ aus der Gruppe Informatik gehören beispielsweise die Schlagwörter „PL 1 || PL/M“ an. Wegen der Spezifität dieser Begriffe sind im

SWB 19 von 19 Titeln deswegen womöglich ebenfalls mit dem Schlagwort „PL 1“ versehen. Demgegenüber steht etwa die Konkordanz „WC 1000 – Biologie“, die zwar einen sehr allgemeinen Bereich umfasst, bei der jedoch ebenfalls eine Korrelation von 100 Prozent zustande kommt (125 Schlagwörter an 125 Notationen). Dieser hohe Schwellenwert kommt im SWB allein durch das mehrbändige Werk „Handbuch der biologischen Arbeitsmethoden / unter Mitarbeit von zahlreichen Fachgenossen hrsg. von Emil Abderhalden“ zustande. Ohne dieses Werk würde die Notation WC 1000 mit der RVK-Klassenbenennung „Hoppe-Seyler-Thierfelder“ aus dem Notationsbereich „Hand- und Lehrbücher“ für Biologie einen geringeren Quotienten ausweisen.

5.3.3 Neue Konkordanzen

Eine zweite Datei dokumentiert die Korrelationen aus dem Verbund, die aber so nicht als Konkordanz in der RVK-Online auftauchen. Diese Begriffe könnten dementsprechend der Erweiterung oder Optimierung der Register dienen und der UB Regensburg als Vorschläge gemeldet werden.³⁹

Pauschal könnte man alle Kombinationen von Schlagwort und Notation für die Erweiterung der Register wählen, die besonders häufig auftreten und einen hohen Schwellenwert besitzen, d.h., bei denen die Wahrscheinlichkeit einer Korrelation und somit voraussichtlich einer sinnvollen Konkordanz besonders hoch ist. Schließt man Notationen, die weniger als 10 Mal auftauchen aus und filtert die Korrelationen mit einem Schwellenwert unter 0,3 heraus, bleiben bei einem aktuellen Abzug von 2009 noch ca. 53.000 neue Konkordanzen übrig. Insgesamt sind das nur 4,8 Prozent nach dem Abgleich aller Konkordanzen und Korrelationen (Abbildung 5). Von diesen besitzen 3.105 Notationen einen Schwellenwert von 1,0 und wären demnach voraussichtlich besonders gut für Neueinträge im Register der RVK geeignet. Ob tatsächlich alle Schlagwörter mit einem hohen Korrelationsquotienten ohne weiteres genutzt werden können, soll an einigen Beispielen überprüft werden.

³⁹ Universitätsbibliothek Regensburg: Neumeldung von Notationen
URL: http://www.bibliothek.uni-regensburg.de/rvko_neu/formular/neumeldung.phtml

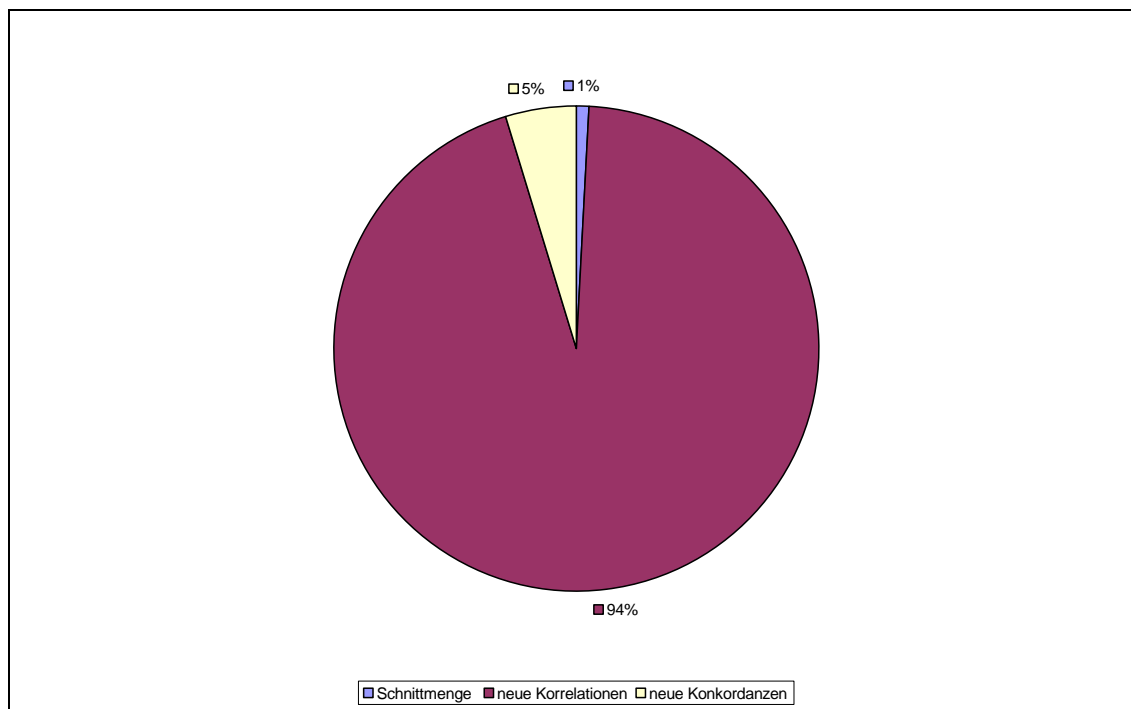


Abbildung 5: Verteilung der Konkordanzen und Korrelationen in den drei Dateien

Ein erstes Beispiel für eine sinnvolle Registerergänzung stellt etwa der Begriff „Rondo“ für die Klasse „LR 54410 – Rondeau, Rondo“ (Schwellenwert von 1,0) dar. Da „Rondo“ bislang überhaupt nicht in der Registersuche erscheint, bleibt diese Systemstelle somit vorerst verborgen. Bei einer Begriffsrecherche in der RVK-Online lassen sich aber auch noch zwei weitere Notationen mit der Klassenbenennung „Rondeau, Rondo“ finden: „LT 54410“ und „LW 54410“. Auch hier wäre eine Ergänzung mit dem Begriff „Rondo“ denkbar.

Ein ähnlicher Fall tritt bei der Kombination PH 2600 – Strafrecht ein (Schwellenwert von 0,98). Der Registerbegriff „Strafrecht“ ist zwar im Bereich Umweltschutz an der Notation „AR 29600 – Umwelt und Bürgerliches Recht, Strafrecht“ und an verschiedenen Stellen in der Gruppe „BR – Kirchenrecht“ auffindbar, aber in Zusammenhang mit der Fachsystematik „P – Rechtswissenschaft“ nur bei „PH 6000 – Gesetzestexte“ als Kette „Strafrecht / Quelle“ vorhanden. Die Systematik P scheint generell sehr begrenzt verbal erschlossen zu sein. Deshalb liegen in der Datei mit den möglichen neuen Konkordanzen viele weitere Notationen aus der Gruppe P mit verschiedenen Schwellenwerten vor, bei denen eine Ergänzung der Benennung „Strafrecht“ möglich wäre. Tabelle 6 zeigt Korrelationen mit einem Schwellenwert von mindestens 0,5 und bei denen die RVK-Notation an wenigstens 20 Titeln vorkommt. In der RVK-Online stehen Registerbegriffe die mehrere Notationen umfassen grundsätzlich an der hierarchisch höchstmöglichen Stelle und vererben den Begriff an die untergeordneten Klassen weiter. Daher wäre es überlegenswert, ob das Schlagwort „Strafrecht“ nicht bereits an den Anfang der Thematik Strafrecht gestellt werden sollte, also etwa in Verbindung mit dem Notations-

bereich „PH 2500 – PH 3980“ oder eventuell sogar mit der gesamten Klasse PH. Verfährt man nach diesem Prinzip, sollte der Begriff dann jedoch von der Notation PH 6000 entfernt werden, um das Register nicht unnötig aufzublähen.

RVK-Notation	Klassenbenennung	Anzahl Titel mit RVK-Notation	Schlagwort	Davon Titel mit Schlagwort	Schwellenwert
PC 2100	<i>Quellenverzeichnisse, Gesetzesverzeichnisse</i>	110	Strafrecht	57	0.51818
PC 2300	<i>Informatorische Handbücher</i>	111	Strafrecht	56	0.50450
PH 2220	<i>Sammelwerke</i>	52	Strafrecht	34	0.65385
PH 2260	<i>Wissenschaft, Forschung, Methodik, Allgemeine Prinzipien, System</i>	40	Strafrecht	21	0.52500
PH 2290	<i>Komplexe Untersuchungen</i>	43	Strafrecht	24	0.55814
PH 2300	<i>Gesamtdarstellungen</i>	39	Strafrecht	24	0.61538
PH 2360	<i>Zeit der Kodifikationen</i>	46	Strafrecht	35	0.76087
PH 2600	<i>Gesamtdarstellungen: Handbücher, Lehrbücher, Einführungen, Grundrisse</i>	100	Strafrecht	98	0.98000
PH 2620	<i>Fall- und Klausurensammlungen</i>	292	Strafrecht	262	0.89726
PH 3000	<i>Gesamtdarstellungen</i>	248	Strafrecht	241	0.97177
PH 3500	<i>Gesamtdarstellungen</i>	339	Strafrecht	321	0.94690
PH 3860	<i>Schwangerschaftsabbruch</i>	35	Strafrecht	20	0.57143
PH 3980	<i>Sonstige Einzeldarstellungen zum Besonderen Teil des StGB</i>	50	Strafrecht	27	0.54000
PH 6103	<i>Kommentare, Hand- und Lehrbücher zum materiellen Strafrecht</i>	75	Strafrecht	63	0.84000
PH 6113	<i>Kommentare, Hand- und Lehrbücher zum materiellen Strafrecht</i>	76	Strafrecht	57	0.75000
PH 6134	<i>Einzeldarstellungen zum materiellen Strafrecht</i>	50	Strafrecht	31	0.62000
PO 4279	<i>Einzeldarstellungen</i>	39	Strafrecht	23	0.58974

Tabelle 6: Mögliche Konkordanzen für das Schlagwort "Strafrecht"

Die ermittelten Schlagwörter müssen jedoch nicht unbedingt nur auf Klassen hinweisen, denen noch keine Registerbegriffe angehören. Einige der neu gefundenen Schlagwörter stellen z.B. verwandte Begriffe oder Synonyme zu anderen bereits vorhandenen Schlagwörtern dar. Für die Systemstelle „KN 1305 – Lehrbuch, Sprachführer, Aufgabensammlung“ existieren momentan z.B. folgende Ketten: Polnischunterricht / Aufgabensammlung || Polnischunterricht / Lehrbuch || Polnischunterricht / Sprachführer. Unter den automatisch generierten Korrelationen befindet sich zu dieser Notation das Schlagwort „Polnisch“, das hier durchaus, etwa in Kombination mit „Lehrbuch“ als Kette zu „Polnisch / Lehrbuch“, ergänzt werden könnte. „Polnischunterricht / Lehr-

buch“ könnte auch zu der Annahme führen, dass es sich um methodische oder didaktische Materialien zum Polnischunterricht handelt und nicht um Dokumente für Lernende. Die Registerbegriffe der Notation „KG 1305 - Lehrbücher, Sprachführer, Aufgabensammlungen“ für den Bereich Russisch in der RVK-Online sind genau nach diesem Muster angelegt: Russischunterricht / Aufgabensammlung || Russischunterricht / Sprachführer || Russisch / Lehrbuch.

Die neu gefundenen Konkordanzen zeigen aber nicht nur an, welche Register bisher generell wenig erschlossen wurden oder wo Synonyme denkbar sind, sondern können des Weiteren unzureichende Permutationen aufdecken. Der Vorschlag aus der Datei (Schwellenwert 1,0 bei 34 von 34 Titeln), das Schlagwort „Deutschland <DDR>“ mit der Systemstelle „DQ 4805 – Sekundarstufe I: Hauptschule (frühere DDR: Polytechnische Oberschule-Oberstufe)“ zu verknüpfen, scheint zunächst nicht sinnvoll. Schließlich ist dieser Begriff bereits bei „DQ 4800 - DQ 4820 Deutschland (DDR) (1949-1990)“ mit der Schlagwortkette „Lehrplan / Deutschland <DDR>“ vorhanden und über die Registersuche auffindbar. Recherchiert man die Begriffe aber in umgekehrter Reihenfolge „Deutschland <DDR> / Lehrplan“ im Registerindex, erhält man keinen Treffer. In diesem Fall wäre es sinnvoll, diese entsprechende Permutation zu ergänzen. Auf die gleiche Weise wurde bei den deutschen Bundesländern vorgegangen. Die untergeordneten Knotenpunkte des Bereichs „DQ 3100 - DQ 4620 – Lehrpläne in den Bundesländern“ besitzen alle die Schlagwortkette „Lehrplan / [Name des Bundeslandes]“ mit der permutierten Form „[Name des Bundeslandes] / Lehrplan“, wie z.B. „DQ 3700 - DQ 3720 – Hessen“ mit „Hessen / Lehrplan || Lehrplan / Hessen“.

In einem letzten Beispiel für eine sinnvolle Konkordanz spiegelt sich die notwendige Überarbeitung der Klassen- und Registerbenennungen nach der SWD wider. Bei einem Blick in die RVK-Online wird deutlich, dass die automatisch generierte Korrelation „AR 13240 – Nationalpark“ (Schwellenwert 0,7561) nur von der Konkordanz abweicht, weil der Registerbegriff wie auch die Klassenbenennung in der Pluralform, nämlich „Nationalparks“, vorliegen.

Einige der denkbaren neuen Konkordanzen sollten jedoch nicht einfach ohne weiteres übernommen werden. Ein hoher Schwellenwert bei einer großen Anzahl übereinstimmender Titel bedeutet nicht zwingend eine sinnvolle Ergänzung. Über den Aufbau des Registers kann beispielsweise festgestellt werden, ob zusätzliche Registerbegriffe wirklich nötig sind. Bei vielen Personenschlagwörtern tritt etwa der Fall ein, dass eine Konkordanz zwischen einer Notation eines Blattknotens und dem Namen jener Persönlichkeit vorgeschlagen wird. Dabei erscheint dieser häufig schon an einer übergeordneten Systemstelle. Als ein Beispiel sei hier die Notation „GK 8630“ mit dem Registerbegriff „Schiller, Friedrich von“ angeführt. Diese Stelle umfasst Sekundärliteratur zum Thema „Leben und Werk. Gesamtdarstellungen“ Schillers und besitzt keine weiteren Registerbegriffe in der RVK-Online. Die Person Friedrich Schillers schließt insgesamt den No-

tationsbereich „GK 8000 - GK 9067“ ein. Eine Suche im Register ergibt eine Verlinkung zur ersten möglichen Notation GK 8000, die ihre Eigenschaften den darunter liegenden Klassen vererbt. Diese folgenden Systematikstellen ebenfalls mit „Schiller, Friedrich von“ zu verschlagworten, würden vielmehr Ballast für das Register bedeuten, als eine hilfreiche Ergänzung zu sein. Beispielsweise sind im Notationsbereich „GL 2500 – GL 9951“ für Schriftsteller aus dem 19. Jahrhundert in der Datei 50 neue Konkordanzan mit einem Schwellenwert zwischen 0,5 und 1,0 vorgeschlagen. Tatsächlich besitzen alle Autoren aber bereits ein Personenschlagwort in ihrem jeweiligen übergeordneten Notationsbereich.

Ein weiterer abwegiger Vorschlag für eine Konkordanz mit einem Schwellenwert von 0,5 besteht zwischen der Klasse „AP 57900 – Beeinflussung, Filmerziehung“ und dem Schlagwort „Filmkritik“. Obwohl der Quotient zunächst nicht unbedingt aussagekräftig erscheint, ist es erstaunlich, dass die Hälfte der Publikationen diese Verbindung aufweist. Schließlich entspricht das Thema Filmkritik eigentlich genau der Systematikstelle „AP 57700 – Filmkritik“. Hier stellt sich die Frage, ob die Klassen der RVK nicht eindeutig genug voneinander abgegrenzt sind, d.h., wie richtig eigentlich die Konkordanz in der RVK-Online ist und/oder ob möglicherweise falsch klassifiziert wurde.

Einen anderen Fall stellt das Beispiel „RK 15226 – Kartenarten, Kartenwerke, Atlanten, Faltblätter“ für Alpenländer dar. In 142 von 142 Fällen ist hier das Schlagwort „Alpenverein“ zugeordnet. Da es sich jedoch um allgemeine Atlanten und Kartenwerke der Alpen handelt und nicht alle Publikationen in dieser Gruppe unbedingt mit dem Alpenverein in Zusammenhang stehen, ist „Alpenverein“ eventuell bereits ein zu spezielles und enges Schlagwort für das Register. Eine aktuelle Recherche im SWB ergibt zudem, dass die Publikation „Alpenatlas : society - economy - environment / Ulrike Tappeiner; Axel Borsdorf; Erich Tasser“ aus dem Jahr 2008 ebenfalls unter dieser Notation gefunden wird, aber nicht das Schlagwort „Alpenverein“ besitzt, denn sie wurde durch einen Verlag veröffentlicht und steht in keinerlei Verbindung zum Alpenverein. Der Registerbegriff „Alpenverein“ wird damit für die Notation „RK 15226 hinfällig“, da das Schlagwort nicht mehr der Beschreibung aller Eigenschaften der Klassenobjekte entsprechen würde. Der „übermächtige“ Urheber bewirkte daher eine zu hohe Korrelation und eine Verzerrung der Ergebnisse.

Wie bereits in Kapitel 5.3.1 festgehalten wurde, bilden Formschlagwörter generell kaum sinnvolle Beziehungen mit den Notationen, insbesondere dann nicht, wenn kein weiteres inhaltstragendes Schlagwort vorhanden ist oder ergänzt wird und das eigentliche Thema nur aus dem Kontext der RVK ermittelt werden kann. In der RVK-Online sind Formschlagwörter als Registerbegriffe eher unregelmäßig vorhanden. In einigen Fällen sind sie an formalen Klassen auffindbar (Bsp. „ZA 31000 – Aufsatzsammlungen“ in der Fachsystematik „Land- und Forstwirtschaft“), ein anderes Mal wiederum nicht (Bsp. „PY 135 – Aufsatzsammlungen“ aus dem Bereich „Rechtswissenschaft / Historische

Kanonistik“ ohne zusätzlichen Registerbegriff). Ein drittes Beispiel verdeutlicht, dass manche Forms Schlagwörter relativ zusammenhangslos in der RVK eingesetzt werden. „EU 630 – Philosophische Hintergründe usw.“ zur „Geschichte der Indogermanistik“ weist mit der Aufzählung „usw.“ nicht nur eine ungenaue Klassenbeschreibung auf, sondern mit dem Schlagwort „Aufsatzsammlung“ in der Kette „Indogermanistik / Geschichte / Aufsatzsammlung“ außerdem eine unzutreffende Registerbeschreibung. Die Übernahme der Forms Schlagwörter aus den Korrelationen zur Ergänzung des RVK-Registers ist deswegen nicht sinnvoll.

5.3.4 Neue Korrelationen

Die dritte Datei enthält die Konkordanzen aus der RVK-Online, zu denen keine oder nur geringe entsprechende Korrelationen aus dem SWB nachgewiesen werden konnten. Folglich könnten diese neuen Korrelationen im Verbund eine wichtige Rolle spielen und entweder auf Benutzerseite der Anreicherung der Daten im OPAC dienen oder dem Fachreferenten bei der Erschließung zugute kommen.

Auffällig ist die Größe dieser Gruppe im Vergleich zu den beiden vorigen. Bei einer Mindestanzahl von 10 Titeln pro Notation und einem Schwellenwert von mindestens 0,3 umfasst die Datei immer noch mehr als eine Million Konkordanzen, die nicht als Korrelation im Verbund vorhanden sind. Das sind annähernd 95 Prozent aller Schlagwort-Notation-Relationen (Abbildung 5 in Kapitel 5.3.3).

Sobald der Schwellenwert weiter erhöht wird, steigt entsprechend auch die Zahl der neuen möglichen Korrelationen in der Datei, denn die Werte der anderen beiden sinken parallel dazu ab (Abbildung 6). Ungewöhnlich ist auch die Anzahl der Korrelationen der Fachsystematik R (Geographie), die zeitweise mehr als zwei Drittel ausmachen (zu möglichen Ursachen s. Kapitel 5.2).

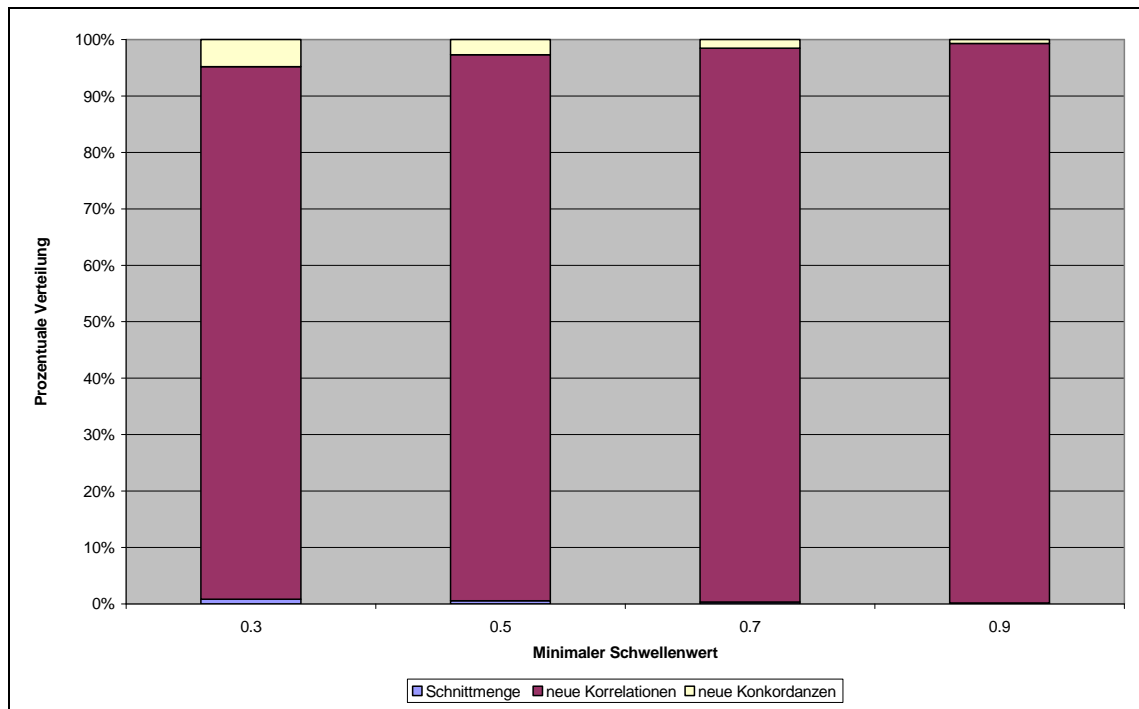


Abbildung 6: Entwicklung der Dateigrößen bei unterschiedlichen Schwellenwerten

Die Ursache für diese außerordentlich große Anzahl an RVK-Konkordanzen könnte in der grundsätzlichen Verschiedenheit der beiden Dokumentationssprachen liegen. Während die Notationen der RVK eher breit angelegt und allgemein gehalten sind, besitzen die Dokumente im SWB oft sehr spezifische, enge Schlagwörter. Hinzu kommt, dass, wie bereits in Kapitel 5.1 erwähnt, Blattknoten ohne Registereintrag stets den als nächstes übergeordneten Begriff erhalten. Da große Teile der RVK noch nicht verbal erschlossen sind, weisen zahlreiche aufeinander folgende Notationen die gleichen Schlagwörter auf. In der Fachsystematik M für Politologie, Soziologie und Militärwissenschaft besitzen fast 90 Prozent der Notationen die Schlagwörter „Politisches System, Schwarzafrika, Afrika“. Dies betrifft die Notationen aus dem Bereich „MI 10.000 – MI 69.999“ und ist darauf zurückzuführen, dass diese Klassen keine Registerbegriffe besitzen. Fast allen Notationen wurden deswegen die Begriffe der übergeordneten Klasse „MI 10.000 – Afrika / Politisches System || Schwarzafrika / Politisches System“ zugewiesen. Die Wahrscheinlichkeit aber, dass die Titel im SWB mehrheitlich diese ohnehin schon sehr weiten Schlagwörter zusammen mit den dagegen sehr spezifischen und tief gegliederten Notationen aufweisen, ist gering. So sind die 25 Titel im SWB, die die Notation „MI 16086“ (Politischer Wandel in Algerien) besitzen und von denen wiederum nur 17 auch nach RSWK erschlossen sind, mit vielfältigen Schlagwörtern wie „Algerien“, „Fundamentalismus“, „Unterdrückung“, „Bürgerkrieg“, „Friedensbemühung“, „Frankreich“ usw. versehen. „Politisches System“ ist nur an einem Datensatz vorhanden, „Afrika“ und „Schwarzafrika“ hingegen nirgends, denn es handelt sich um die geographisch sehr viel exaktere Region „Algerien“. Würde man also die Blattknoten

ohne Registerbegriffe unter einer Notation zu einem virtuellen Knoten zusammenfassen, wäre die Datei mit den Konkordanzen der RVK-Online vermutlich weniger umfangreich und es kämen damit vermutlich weniger neue Korrelationen, dafür jedoch mehr neue Konkordanzen zustande.

6 Anwendungsmöglichkeiten

Die automatisch generierten Korrelationen besitzen in der Praxis einen hohen Nutzen. Im folgenden Kapitel werden verschiedene Anwendungsmöglichkeiten der Korrelationen und der später weiterbearbeiteten Dateien vorgestellt.

6.1 Optimierung der RVK

Die neuen Konkordanzen eignen sich gut, um die Klassen der RVK mit Registerbegriffen zu ergänzen und die Arbeit an den Registern zu vereinfachen. Interesse an einer Verbindung zwischen RVK und SWD besteht ohnehin schon seit längerem:

„Wünschenswert erscheint die Erstellung einer RVK-SWD-Konkordanz als Ergänzung zu DDC-SWD-Konkordanz, die bereits in Arbeit ist, um die bessere Integration von systematischer und verbaler Sacherschließung weiter voranzubringen.“⁴⁰

„In vielen Fachsystematiken fehlen noch entsprechende Register. Die Koordinierungsstelle ist hierbei auf die Mitarbeit der Anwenderbibliotheken angewiesen, da sie dies personell nicht leisten kann. Die Suche nach Registerbegriffen wurden von den Anwesenden als wichtiges Angebot in RVK-Online bestätigt.“⁴¹

Die Vorteile einer Überarbeitung bzw. Erweiterung der RVK-Registerbegriffe liegen auf der Hand. Da die meisten Bibliotheken zur verbalen Erschließung von Dokumenten auf die RSWK und SWD zurückgreifen, erscheint es nur logisch für den verbalen Teil der RVK ebenfalls SWD-Vokabular zu verwenden. Auf diese Weise könnten dem Benutzer z.B. auch die Registerbegriffe bei einer Recherche angeboten oder sogar in den OPAC integriert werden. Damit würde dem Nutzer die Anwendung der Klassifikation näher gebracht und erleichtert. Besäße die RVK ein vollständiges Register mit kontrolliertem Vokabular, wäre auch eine Konkordanz zu einer anderen Dokumentationssprache wie der DDC im Projekt CrissCross⁴² denkbar.

Im Allgemeinen könnte natürlich intellektuell geprüft werden, ob die neu gefundenen Schlagwörter für das Register der RVK sinnvoll genutzt werden können. Bei einem Umfang von mehreren tausend Notationen würde dies jedoch erst einmal einen erheblichen Aufwand bedeuten. Ein automatisches Verfahren zur Auswahl in Frage kommender Schlagwörter könnte hier Abhilfe schaffen. So könnte es im ersten Schritt überprü-

⁴⁰ Raffelt, Albert (2008), S. 220

URL: http://www.freidok.uni-freiburg.de/volltexte/5000/pdf/bibliothek_aussen_innen.pdf

⁴¹ Protokoll des RVK-Anwendertreffens, S. 6

URL: http://www2.bsz-bw.de/cms/public/arbeit/agsach/Protokoll_RVK_Anwendertreffen_2008.pdf

⁴² DFG-Projekt CrissCross

URL : <http://linux2.fbi.fh-koeln.de/crisscross/index.html>

fen, ob die Notation schon Registerbegriffe besitzt oder nicht. Anfangs könnten z.B. Klassen ohne Registerbegriffe Priorität haben. Zusätzlich könnten bestimmte Schlagworttypen eliminiert und nur Korrelationen ab einem ausreichend hohen Schwellenwert berücksichtigt werden. Anschließend würde anhand des Registeraufbaus geprüft werden, ob das Schlagwort bereits als Registerbegriff in einer über- oder nebengeordneten Klasse auftaucht, wie im Fall „GK 8630 – Schiller, Friedrich von“ oder bei „AP 57900 – Filmkritik“, um diese Benennung ebenfalls aus der Liste zu entfernen.

6.2 Unterstützung bei der Sacherschließung

Die automatisch generierten Korrelationen zwischen SWD-Schlagwörtern und RVK-Notationen aus dem SWB könnten als Hilfsmittel für die verbale Sacherschließung dienen. Wenn ein Titel im ersten Schritt also auf automatischem oder intellektuellem Weg eine oder mehrere RVK-Notationen erhält, könnten dazu sämtliche korrelierende Schlagwörter mit ihren Schwellenwerten, PPNs und Angaben zur Titelanzahl als Vorschläge ausgegeben werden. Tabelle 7 zeigt beispielhaft alle Korrelationen, die für die Klasse „CI 7317 – Abhandlungen, Studien“ des Philosophen Miguel de Unamuno zustande gekommen sind.

RVK-Notation	Anzahl Titel mit RVK-Notation	PPN Schlagwort	Schlagwort	Davon Titel mit Schlagwort	Schwellenwert
CI 7317	12	209195061	Menschenbild	1	0.08333
CI 7317	12	208845542	Angst	1	0.08333
CI 7317	12	209182377	Erziehungsphilosophie	1	0.08333
CI 7317	12	208838821	Ästhetik	1	0.08333
CI 7317	12	210035137	Tragik	1	0.08333
CI 7317	12	209960655	Polemik	2	0.16667
CI 7317	12	20954158X	Existentialismus	2	0.16667
CI 7317	12	209066873	Philosophie	5	0.41667
CI 7317	12	209140755	Unamuno, Miguel de	12	1.00000

Tabelle 7: Mögliche Schlagwörter für die RVK-Notation CI 7317

Da an den Notationen mitunter sehr viele Schlagwörter vorhanden sind, könnten diese beispielsweise aufsteigend nach Schwellenwerten oder Titelanzahl sortiert werden, um so bei der Auswahl der Schlagwörter mit den relevantesten beginnen zu können. Dafür sollten die Korrelationen in einer gesonderten Datenbank abgelegt werden, die die aktuellen Daten aus dem SWB übernimmt bzw. sich regelmäßig parallel zum SWB aktualisiert, um immer den neuesten Abzug vorliegen zu haben.

Umgekehrt könnte die gleiche Datenbank mit den automatisch generierten Korrelationen auch eine Suchmaske besitzen, die es erlaubt, mehrere Schlagwörter mittels Bool-

scher Operatoren zu suchen. Auf diese Weise wäre es möglich, gleichzeitig unterschiedliche Schlagwörter zu recherchieren und eine dazu passende Notation zu ermitteln.

Schließlich könnte diese Datenbank wie das sog. „Classification Web“ (s. Kapitel 4.2.2) frei oder zu kommerziellen Zwecken online zugänglich gemacht werden.

6.3 Anreicherung des SWB

Die automatisch erzeugten RVK-SWD-Korrelationen sind nicht nur hilfreich bei der Vergabe von Schlagwörtern, sondern können auch auf verschiedene Arten in den OPAC integriert werden und dem Benutzer zugute kommen. Kapitel 6.3 zeigt unterschiedliche Möglichkeiten des Einsatzes der Korrelationen bei der Recherche.

6.3.1 Einbindung von Schlagwörtern

In Zeiten des Web 2.0 kommt es immer öfter darauf an, den Bibliothekskatalog für den Recherchierenden attraktiver, insbesondere interaktiver, zu gestalten. Eine alleinige Einbindung der neuen Korrelationen in einen Titeldatensatz, z.B. über ein zusätzliches Feld, mit dem Ziel der Termerweiterung scheint deswegen nicht besonders sinnvoll und produziert womöglich eher Ballast bei einer Recherche. Allein mit der Notation „UB 5000 – Populäre Naturwissenschaften, allgemein“ aus dem Bereich Physik sind 93 verschiedene Korrelationen möglich. Insbesondere Allgemeinbegriffe wie „Alltag“, „Physik“ oder „Geschichte“ würden die ursprünglich präzise verbale Erschließung der Publikation überflüssig machen. Individualbegriffe wie „UFO“, „Getreidefeld“ oder „Parität <Physik>“ hingegen würden den Benutzer bei einer Recherche möglicherweise in eine falsche Richtung bzw. zu den falschen Titeln weisen, auch wenn die Schwellenwerte in diesen Beispielen sehr niedrig ausfallen.

Stattdessen könnte das Mapping indirekt beispielsweise über eine „Tag Cloud“ in den OPAC eingebunden werden, um dem Benutzer Hinweise für Anschlussrecherchen zu liefern. Die Vorschläge in Form von Tags bzw. Schlagwörtern könnten immer dann erscheinen, wenn ein Benutzer einen Titel im Katalog aufruft. Besitzt das Dokument eine oder mehrere RVK-Notationen, können die mit dieser Notation in der Korrelationsdatei verknüpften Schlagwörter als zusätzliches Recherchetool angezeigt werden. Auf diese Weise werden die Titel nicht mit irrelevanten Schlagwörtern zugeschwemmt, und gleichzeitig werden die Recherchefunktionen flexibler gestaltet. Nicht zuletzt kann dadurch auch die SWD auf eine attraktivere Art in den Online-Katalog integriert werden.

Tag Clouds ermöglichen ferner die visuelle Darstellung der unterschiedlichen Gewichtung der abgebildeten Schlagwörter. Je wichtiger oder häufiger ein Tag ist, desto größer oder farbiger wird er beispielsweise angezeigt. Diese Funktion kann über die Auswer-

tung des Korrelationsquotienten am jeweiligen Schlagwort bewirkt werden. Bei einem Wert von 0,9 wird das Schlagwort X also größer angezeigt als das Schlagwort Y mit einem Quotienten von 0,3. Auch hier wäre es denkbar, einen Mindestwert für den Quotienten und eine Mindestanzahl an Titel im Voraus zu definieren. Da die Schlagwörter in der Wolke mit Links hinterlegt sind, kann über einen Klick auf den gewünschten Begriff außerdem sofort eine neue (Schlagwort-)Recherche ausgelöst werden.

Abbildung 7 zeigt eine manuell erstellte Tag Cloud, die sämtliche Schlagwörter der Notation AN 91500 zum Thema „Ausbildung im Bibliotheks- und Informationswesen“ beinhaltet. Nur die Formschlagwörter wurden entfernt. Die Farben dienen der Abgrenzung der einzelnen Begriffe.



Abbildung 7: Tag Cloud zur RVK-Notation AN 91500

Betrachtet man die Tag Cloud der Abbildung 7 genauer, wird schnell ersichtlich um welches Thema es sich handelt. Die unterschiedlich großen Schlagwörter repräsentieren die Publikationen im SWB verhältnismäßig logisch und nachvollziehbar. Das Schlagwort „Niedersachsen“ allerdings erscheint mit seiner Größe zunächst ungewöhnlich. Diese auffällige Größe ist jedoch auf das mehrbändige Werk „Modellversuch Konzeption und Entwicklung von Studiengängen im Bereich Bibliothek, Information und Dokumentation“ im SWB zurückzuführen, das vom Niedersächsischen Ministerium für Wissenschaft und Kunst publiziert wurde und sich vermutlich vornehmlich mit einem Modellversuch dieser Studiengänge im Land Niedersachsen befasst. Aber auch die unscheinbaren Begriffe sind nicht zu unterschätzen. Sie zeigen z.T. sehr spezielle Themenbereiche auf dem Gebiet zu Ausbildung und Studium im Bibliotheks- und Informationswesen auf, wie z.B. „Zeiteinteilung“ und „Online-Informationssystem“. Am letzten Beispiel wird offensichtlich, dass die Schlagwörter auch Hinweise auf ganz neue Themengebiete liefern können. Die Publikation, die mit „Online-Informationssysteme“

verschlagnotiert wurde, besitzt tatsächlich nur dieses einzelne Schlagwort, gleichzeitig aber insgesamt vier RVK-Notationen, u.a. aus dem Bereich Informatik.

Die Tag Cloud ließe sich später auf verschiedene Arten realisieren. Sie könnte beispielsweise gleich im Anschluss an eine Recherche für sämtliche Notationen in der Ergebnisliste erscheinen. Problematisch wäre dann nur die Darstellung bei großen Treffermengen, zumal ein Titel allein bereits mehrere Notationen gleichzeitig aufweisen kann und oft eine hohe Anzahl von Schlagwörtern pro Notation in den Korrelationen möglich ist. Daher ist es wahrscheinlich effektiver, wenn die Tag Cloud im Katalog erscheint, sobald ein konkreter Titeldatensatz aufgerufen wird.

Vorstellbar wäre auch, die Registerbegriffe der RVK-Notationen einzubinden, insofern die Klasse Registerbegriffe aufweist und diese nicht bereits in den Korrelationen als Schlagwörter vorhanden sind. Sie könnten beispielsweise mit einer anderen Farbe und in einer einheitlichen Größe in der Tag Cloud angezeigt werden, da in diesem Fall kein Schwellenwert zur Größenbestimmung gegeben ist. Mit dieser Methode ließe sich z.B. auf die RVK verweisen, die dem Benutzer somit näher gebracht werden könnte.

6.3.2 Einbindung von RVK-Notationen

Neben den Schlagwörtern könnte auch die RVK selbst unter Einbeziehung der automatisch generierten Korrelationen auf eine andere Art gewinnbringend in den Katalog integriert werden als nur über die Möglichkeit zur Recherche nach Notationen. Die Suche mit Notationen ist vielen Nutzern mitunter immer noch unbekannt oder zu umständlich, weil sie sich nicht mit der Systematik auskennen. Dabei gestatten Recherchen über eine Klassifikation eher thematisch vollständige Ergebnislisten als beispielsweise ein einfaches Stichwort. In diesem Fall würde mittels der Korrelationen zwischen RVK-Notation und SWD-Schlagwort also nicht wie im Beispiel der Tag Cloud von der Notation auf verschiedene Schlagwörter geschlossen, sondern umgekehrt. Dem Suchenden würde die RVK damit z.B. neben der aktuellen Ergebnismenge präsentiert werden, ohne dass er selbst eine Recherche über Notationen initiiert hätte.

Sobald im Online-Katalog eine verbale Rechercheanfrage aufgegeben wird, würde das entsprechend gesuchte Stichwort mit den Schlagwörtern in den Korrelationsdaten abgeglichen werden. Es wäre denkbar, an dieser Stelle auch linguistische Methoden wie Stemming oder Ähnlichkeitssuchfunktionen einzusetzen, um durchgängig Treffer zu erhalten, sollte der gewählte Suchbegriff nicht den Vorgaben der SWD entsprechen und als Schlagwort in den Korrelationsdaten vorhanden sein. Wird das Suchwort dann in den Schlagwörtern der Korrelationsdaten gefunden, können entsprechend Notationen zugeordnet werden. Diese würden dann als Vorschläge für bestimmte Themengebiete an den OPAC bzw. an den Benutzer mit der Suchanfrage zurückgeliefert werden. Auf diese Weise können über den Einstieg mit einem Suchbegriff möglicherweise auch

Themengebiete in der RVK aufgedeckt werden, die keine feste Systemstelle besitzen, so z.B. im Fall neuerer Wissenschaftsfächer wie Kulturwissenschaften oder Umweltschutz. Der Schwellenwert kann auch hier wieder als eine Art „Ranking-Werkzeug“ eingesetzt werden. Notationen, die häufiger zusammen mit einem bestimmten Schlagwort zusammen auftreten, stehen z.B. am Anfang der Vorschlagsliste, Notationen mit niedrigeren Korrelationsquotienten folgen in absteigender Reihenfolge. Bei zu vielen Vorschlägen könnte prinzipiell eine untere Grenze gesetzt werden. Auf diese Weise könnte der Benutzer mit den angebotenen Notationen weitere Recherchen durchführen, ohne sich in der Systematik auskennen zu müssen.

An einem Beispiel soll demonstriert werden, wie die Darstellung der entsprechenden Korrelationen im OPAC erfolgen könnte. Wurde etwa mit dem Stichwort „Umweltrecht“ im Katalog gesucht, würde dieser Begriff mit den Schlagwörtern in den Korrelationen abgeglichen werden. Tabelle 8 zeigt Korrelationen, die zu diesem Suchbegriff gefunden wurden. Dabei wurden alle Korrelationen herausgefiltert, die einen Schwellenwert unter 0,3 besaßen und bei denen die RVK-Notation an weniger als 20 Titeln im SWB vorkam, um unnötigen Ballast zu vermeiden.

RVK-Notation	Übergeordnete Klassenbezeichnungen	Schlagwort	Anzahl Titel mit RVK-Notation	Davon Titel mit Schlagwort	Schwellenwert
PN 792	<i>Rechtswissenschaft / Verwaltungswissenschaften und Verwaltungsrecht / Besonderes Verwaltungsrecht / Recht der Naturnutzung und des Umweltschutzes / Umweltrecht / Gesetzessammlungen, Entscheidungssammlungen</i>	Umweltrecht	77	66	0.85714
AR 29000	<i>Allgemeines / Naturschutz und Umweltschutz / Umweltrecht / Allgemeines</i>	Umweltrecht	81	60	0.74074
PN 794	<i>Rechtswissenschaft / [...] / Umweltrecht / Kommentare, Gesamtdarstellungen, Handbücher, Wörterbücher</i>	Umweltrecht	160	110	0.68750
PN 795	<i>Rechtswissenschaft / [...] / Umweltrecht / Allgemeinproblematik und Methodenfragen</i>	Umweltrecht	174	69	0.39655
PN 798	<i>Rechtswissenschaft / [...] / Umweltrecht / Rechtsschutz</i>	Umweltrecht	23	9	0.39130
PN 796	<i>Rechtswissenschaft / [...] / Umweltrecht / Umweltverfassung, Umweltverwaltung</i>	Umweltrecht	109	42	0.38532
PN 793	<i>Rechtswissenschaft / [...] / Umweltrecht / Materialien, Sammelwerke, Serien</i>	Umweltrecht	252	87	0.34524

Tabelle 8: Korrelationen zum Suchbegriff "Umweltrecht"

In Abbildung 8 wird am Beispiel der Korrelationen zum Schlagwort „Umweltrecht“ eine mögliche Darstellungsform der RVK-Systematikstellen im OPAC vorgestellt. Diese Liste soll dem Nutzer Vorschläge zu weiteren und insbesondere möglichst vollständigen Recherchen unterbreiten. An dieser Stelle soll also keine Eingrenzung der Treffermenge erfolgen, sondern mit dem Klick auf eine Systemstelle eine komplette Ergebnisliste der Titel mit den entsprechenden Notationen aufgerufen werden. Die Ergebnisse sind absteigend je nach Höhe des Schwellenwerts gerankt. Bei der Darstellung der Klassen bzw. der Klassenbeschreibungen wurde in diesem Beispiel aus Gründen der Übersichtlichkeit nur die Hauptklasse angegeben. Die Punkte sollen zeigen, dass es sich hierbei nicht nur um Titel aus dem Gebiet „Rechtswissenschaften“ handelt, sondern noch weiter untergliedert wird. Auf die drei Punkte folgt die Notation, an der das Schlagwort „Umweltrecht“ in den Korrelationsdaten gefunden wurde. Sie soll der Abgrenzung zu den anderen Sachgebieten mit der gleichen Hauptklasse dienen, die gleichzeitig in der Liste erscheinen, wie in diesem Fall „Rechtswissenschaft“, und könnte den Benutzer möglicherweise mit dem Aufbau der Systematik vertraut machen. Die vollständige Hierarchie kann eingesehen werden, indem mit dem Mauszeiger über den jeweiligen Hyperlink gefahren wird, der über ein zusätzliches gelbes Feld die hierarchische Gliederung in Form von Klassenbenennungen bis zur jeweils angegebenen Notation offenbart. Eine sofortige Anzeige der vollständigen Klassenbenennungen in dem blau hinterlegten Feld wäre vermutlich zu umfangreich und würde „den Bildschirm sprengen“. In Klammern hinter der RVK-Notation befindet sich die Anzahl der Titel, die im SWB zu jener Notation vorhanden sind und über den hinterlegten Hyperlink direkt aufgerufen werden können. Sie weichen in der Regel von den Zahlen in den Korrelationsdaten mit der Spalte „Anzahl Titel mit RVK-Notation“ ab, da in den automatisch generierten Korrelationen nur diejenigen Publikationen erfasst wurden, die sowohl klassifikatorisch über die RVK als auch verbal über die SWD erschlossen wurden. Unter dem Punkt „Mehr anzeigen“ könnten außerdem weitere Systemstellen zu finden sein, die beispielsweise nicht erscheinen, weil der Schwellenwert unter 0,3 liegt.

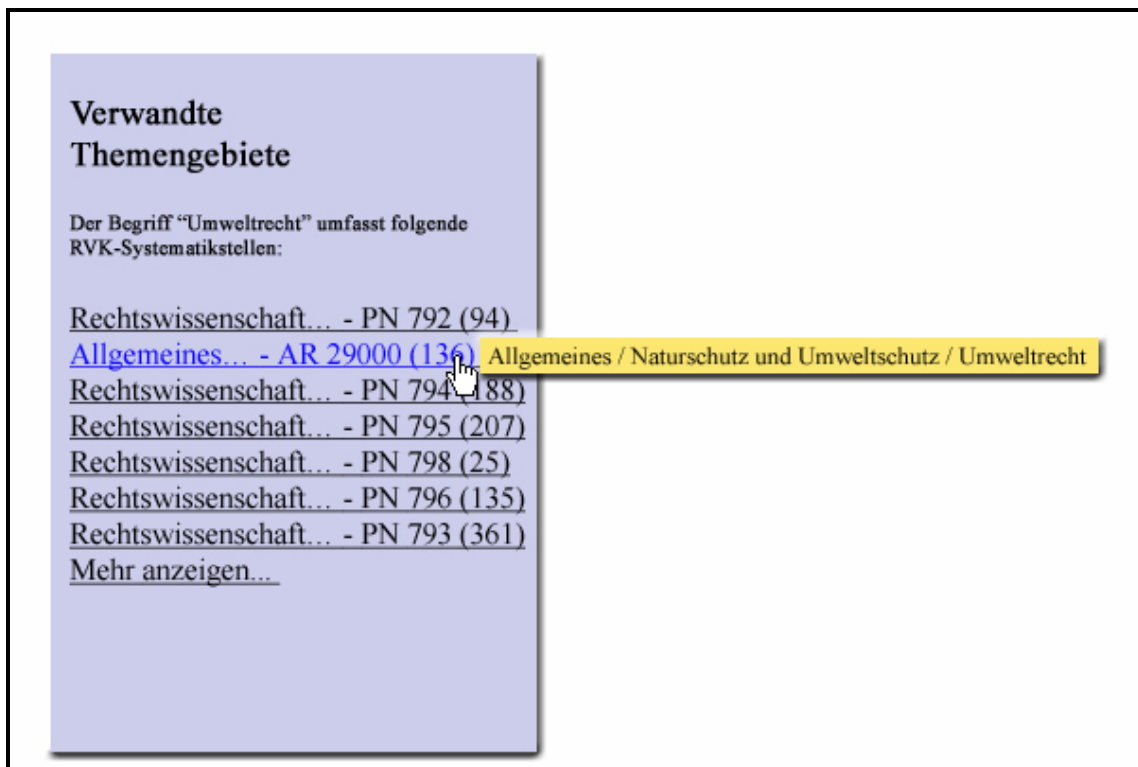


Abbildung 8: Mögliche Darstellung von RVK-Systemstellen im OPAC

Auf diese Art wird dem Suchenden ein unkomplizierter verbaler Zugang zur RVK angeboten und die Anschlussrecherche erleichtert. Neben der Möglichkeit, die Felder mit Hyperlinks zu den Ergebnismengen zu hinterlegen, könnte alternativ oder auch zusätzlich, z.B. über ein kleines Symbol hinter der Anzahl der Treffer, ein Link zu Bibscout⁴³ eingefügt werden. Bibscout verzeichnet und strukturiert die Bestände des SWBs nach der RVK. Über die einzelnen Hauptklassen kann sich zur gewünschten Notation durchgeklickt oder ganz einfach gebrowsed werden. So könnte der Nutzer bei der Suche nach der Notation „AR 29000“ feststellen, dass diese Notation nur „Allgemeines“ zum Thema Umweltrecht umfasst und im Anschluss noch weitere, sehr viel spezifischere Klassen wie „AR 29040 – Naturschutzrecht, Recht der Landschaftspflege, Tierschutzrecht“, „AR 29200 – Lebensmittelrecht“ oder „AR 29240 – Recht der Umweltchemikalien und Biozide“ folgen.

Weiterhin könnte die Anzeige der RVK-Systematikstellen um das sog. „Faceted Browsing“ erweitert werden. Mit dieser Methode würden nun die Ergebnisse weiter eingegrenzt werden, statt wie bisher Ansätze für weitere Recherchen zu liefern. Dies scheint vor allem dann sinnvoll, wenn an einer Notation große Mengen an Treffern gefunden werden, wie bei „PN 793“ mit mehr als 350 Dokumenten im SWB. Dafür könnten z.B. die Schlagwörter der einzelnen Titel dienen, die ohnehin eine noch feinere Untergliederung

⁴³ BibScout

URL: <http://titan.bsz-bw.de/bibscout/>

zung im Vergleich zur Klassifikation darstellen. Sie könnten entweder als Schlagwortketten angeboten werden, einzeln alphabetisch geordnet oder je nach Häufigkeit in der jeweiligen Klasse aufgelistet werden. An dieser Stelle könnte erneut die Tag Cloud zum Einsatz kommen und die Schlagwörter grafisch strukturiert präsentieren.

Die obige Darstellung scheint jedoch zumindest in Bezug auf die Veranschaulichung der Hierarchie und die Umsetzung der Klassenbenennungen und Notationen noch verbesserungsfähig. Ein weiteres Beispiel stellt dar, wie diese Aspekte noch besser umgesetzt werden könnten. Abbildung 9 zeigt eine grafische Darstellung der Korrelationen innerhalb der Systematik der RVK. Auch hier treten wieder alle Notationen auf, die unter dem Begriff „Umweltrecht“ in den Korrelationsdaten gefunden werden konnten. Sie sind schwarz gefärbt und unterschiedlich groß je nach Höhe des Schwellenwertes, wie auch die Schlagwörter in der Tag Cloud. Die übrigen grauen Begriffe sind die jeweils übergeordneten Klassenbenennungen, die über Linien miteinander verbunden sind. Auf diese Weise wird nicht nur die Hierarchie der Klassifikation eindeutig dargestellt, sondern es können auch die Klassenbeschreibungen vollständig angezeigt und nachvollzogen werden. Mit einem Klick auf die entsprechende Systemstelle werden wieder die entsprechenden Ergebnisse aufgerufen.

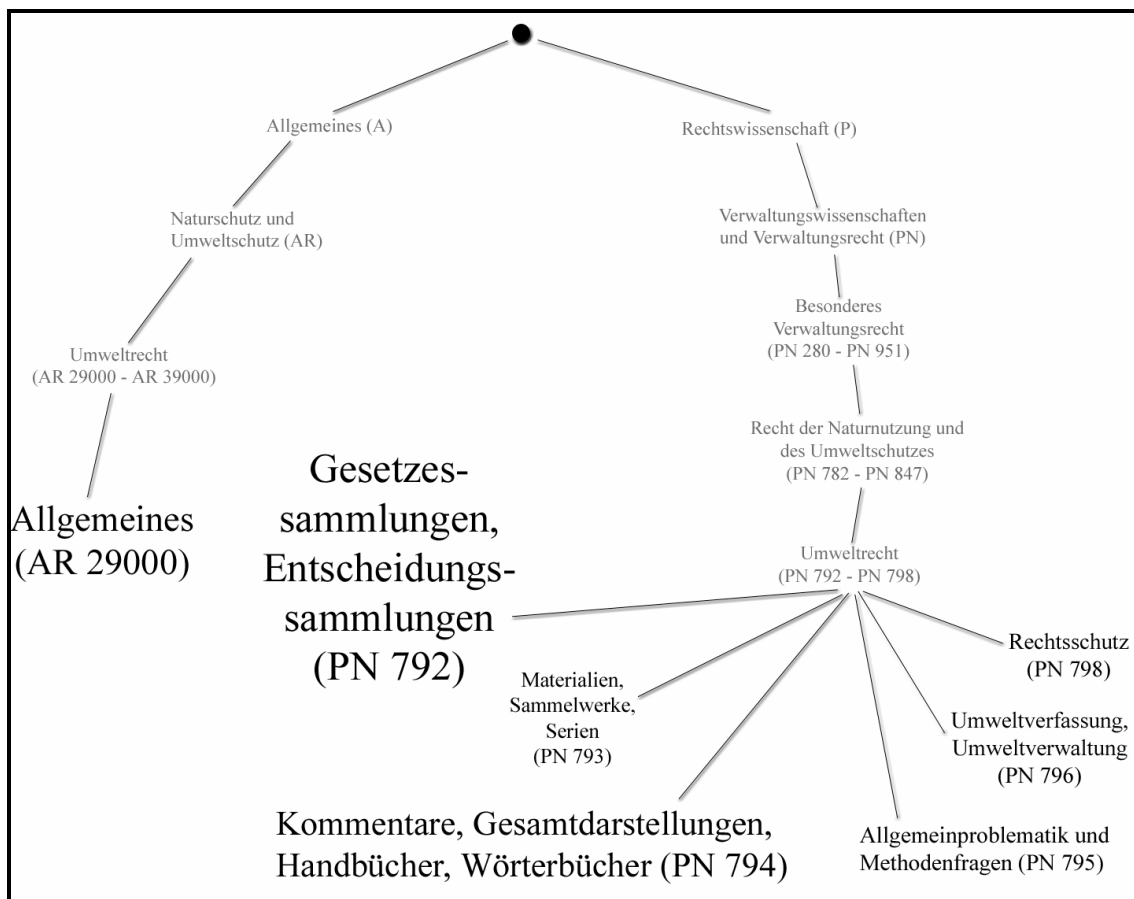


Abbildung 9: Grafische Darstellung der Korrelationen in Verbindung mit der RVK-Hierarchie

Zudem wäre es überlegenswert, ob nicht an dieser Stelle die komplette RVK hinterlegt werden könnte. Somit hätte der Nutzer die Möglichkeit, auch über- oder nebengeordnete Systemstellen zu verfolgen. Ein Klick auf „Umweltrecht (AR 29000 – 39000)“ würde also bewirken, dass tatsächlich alle Systemstellen aus diesem Bereich angezeigt würden, und man könnte wie in der RVK-Online browsen. Dafür wären auch Techniken des grafisch-assoziativen Browsers denkbar, der im Grunde über ein „Netz“ verschiedene verwandte Begriffe mit unterschiedlichen Farben oder Größen anzeigt und auf weitere Recherchemöglichkeiten hinführen soll. Wählt man einen Begriff aus diesem Netz aus, rückt dieser ins Zentrum und versammelt um sich neue verwandte Stichwörter, während die vorigen Begriffe in den Hintergrund rücken. Ähnlich ließe sich auch die RVK im OPAC nachbilden. So bliebe die Darstellung übersichtlich und man wüsste, an welcher Systemstelle man sich befindet, ohne andere Äste der Klassifikation wie in der RVK-Online manuell schließen zu müssen.

7 Fazit und Ausblick

Automatisch generierte Korrelationen zwischen einem Klassifikationssystem und einem Thesaurus können, wie im letzten Kapitel deutlich wurde, auf vielfältige Art und Weise auf Seiten des Nutzers oder der Bibliothek zum Einsatz kommen. Obwohl bereits in einigen Projekten mit Korrelationen zwischen zwei verschiedenen Dokumentations-sprachen gearbeitet wurde (Kapitel 4.2.2), stand die Analyse und Nutzung dieser bislang kaum im Mittelpunkt der Untersuchungen.

Die Untersuchung einzelner Korrelationen zwischen RVK und SWD ergab schließlich sehr unterschiedliche Ergebnisse. Einige Korrelationen waren sehr gut und gaben Anlass zu weiteren Untersuchungen und Einsatzmöglichkeiten, andere Korrelationen konnten von vornherein von weiteren Betrachtungen ausgeschlossen werden. Dazu gehörten z.B. die Formschlagwörter, die auf Grund ihrer geringen sachlichen Aussagekraft sowie dem inkonsequenten Erscheinen in SWB und RVK-Online aus den Ergebnissen zu entfernen gewesen wären. Es wurde außerdem festgestellt, dass die Verbindungen aus Notation und Schlagwort nicht immer vorhersehbar sind und der Schwellenwert nicht die Aussagekraft über den Wert der Korrelation besitzt, wie zunächst vermutet. Während des Vergleichs der Korrelationen mit der Konkordanz aus der RVK-Online wurden zudem die grundsätzlichen Unterschiede zwischen den zwei Dokumentationssprachen bzw. deren heterogenes Verhalten zueinander deutlich.

Daraus resultierten wiederum Ideen zum praktischen Einsatz der automatisch generierten Korrelationen, denn der Abgleich von Korrelationen und Konkordanz ermöglichte es, u.a. neue Konkordanzen und damit neue Registerbegriffe für die RVK-Online zu sammeln. Mit einem vollständigen RVK-Register würde nicht nur die Arbeit bei der Klassifizierung von Dokumenten selbst leichter fallen, sondern es würden auf diese Weise auch Konkordanzen zu anderen Dokumentationssprachen, insbesondere anderen großen Klassifikationssystemen, wie z.B. der DDC, leichter durchführbar.

Auf der anderen Seite eignen sich die Korrelationen hervorragend als Sacherschließungshilfe. Im Erschließungsprozess könnten sie als Vorschlagsinstrument für Schlagwörter zu einer bestimmten Notation oder umgekehrt dienen.

Beide Dokumentationssprachen eignen sich außerdem als sehr gute Suchinstrumente und sollten jedem Benutzer zur Verfügung gestellt werden. Obwohl die RVK und die SWD zumindest im deutschsprachigen Raum weit verbreitet sind und grundsätzlich in Katalogen wissenschaftlicher Bibliotheken als Recherchemöglichkeit angeboten werden, sind sie noch zu wenig genutzt. Zu oft sind sie nur über die erweiterte Suche zugänglich oder anderweitig leicht zu übersehen. Hinzu kommt, dass den meisten Suchen-

den die Recherche mit Notationen und kontrolliertem Vokabular unbekannt ist und auch Hilfstexte oder Tutorials noch zu selten genutzt werden.

Die automatisch generierten Korrelationen zwischen der RVK und SWD könnten Sachrecherchen unter Einsatz der Dokumentationssprachen zukünftig erleichtern und insbesondere in Zeiten des Web 2.0 zugunsten optimierter Rechercheinstrumente eingesetzt werden. Im Rahmen der Arbeit wurden verschiedene Möglichkeiten aufgezeigt automatisch generierte Korrelationen in einen Bibliothekskatalog zu integrieren. Über eine Tag Cloud ließen sich etwa Schlagwörter anzeigen, die dem Benutzer bei Anschlussrecherchen dienlich sein könnten. Aber auch die Hierarchie der RVK selbst könnte mit Elementen eines grafisch-assoziativen Browsers auf eine attraktivere und verständlichere Art in den OPAC eingebunden werden.

Dem Nutzer sollten grundsätzlich auch im wissenschaftlichen Bibliotheksbereich mehr Möglichkeiten geboten werden, Anschlussrecherchen auszuführen. In öffentlichen Bibliotheken sind Web 2.0-Funktionen wie Tag Clouds und grafisch-assoziative Browser dagegen schon häufiger im Einsatz.

Literaturverzeichnis

Abschlußbericht zum Projekt MILOS II

URL: http://www.ub.uni-duesseldorf.de/home/ueber_uns/projekte/abgeschlossene_projekte/milos/mil_ber
(Zugriff am 09.06.2009)

Arbeitsgruppe Klassifikation der Kommission für Erschließung: Protokoll vom 16.01.2006

URL: http://www.bibliothek.uni-regensburg.de/Systematik/protokolle/2006_01_16.pdf
(Zugriff am 03.06.2009)

Bambey, Doris: Thesauri und Klassifikationen im Internet : neue Herausforderungen für klassische Werkzeuge. In: Neisser, Horst u.a. (Hrsg.): Information und Öffentlichkeit (1. Gemeinsamer Kongress von BDB und DGI). Leipzig u.a. : Dinges u. Frick, 2000, S. 245-255

Benz, Christian: Von der formalen zur realen Einschichtigkeit der Universitätsbibliothek Mannheim. 2008

URL: http://www.opus-bayern.de/bib-info/volltexte/2008/523/pdf/benz_mannheim_2008.pdf (Zugriff am 03.06.2009)

Bertram, Jutta: Einführung in die inhaltliche Erschließung: Grundlagen – Methoden – Instrumente. Würzburg : Ergon, 2005

BibScout: Das Internet-Verzeichnis der Medienbestände im Südwestdeutschen Bibliotheksverbund (SWB), systematisch geordnet nach der Regensburger Verbundklassifikation

URL: <http://titan.bsz-bw.de/bibscout/> (Zugriff am 03.06.2009)

Classification Web

URL: <http://classificationweb.net/> (Zugriff am 03.06.2009)

Classification Web: Quick Start Tutorial ; LC/Dewey Correlations

URL: <http://classificationweb.net/tutorial/9dewey.html> (Zugriff am 03.06.2009)

Classification Web: Quick Start Tutorial ; Subject Heading & Classification Number Correlations

URL: <http://classificationweb.net/tutorial/7subjcor.html> (Zugriff am 03.06.2009)

Deutsches Bibliotheksinstitut: Klassifikationen für wissenschaftliche Bibliotheken: Analysen, Empfehlungen, Modelle. Berlin : Dt. Bibliotheksinstitut, 1998

DFG-Projekt CrissCross

URL : <http://linux2.fbi.fh-koeln.de/crisscross/index.html> (Zugriff am 09.06.2009)

Frank, Eibe/Paynter, Gordon W.: Predicting Library of Congress Classifications from Library of Congress Subject Headings. In: Journal of the American Society for Information Science and Technology 55 (2004) 3, S. 214-227

Geißelmann, Friedrich: Online-Version einer Aufstellungssystematik. In: Havekost, Hermann (Hrsg.): Aufbau und Erschließung begrifflicher Datenbanken : Beiträge zur bibliothekarischen Klassifikation. Oldenburg : BIS-Verl., 1995, S. 131-152

INFOMINE : Scholarly Internet Resource Collections
URL: <http://infomine.ucr.edu/> (Zugriff am 09.06.2009)

Krause, Jürgen: Sacherschließung in virtuellen Bibliotheken : Standardisierung versus Heterogenität. In: Rützel-Banz, Margit (Hrsg.): 89. Deutscher Bibliothekartag in Freiburg im Breisgau 1999 : grenzenlos in die Zukunft (ZfBB (2000) Sonderheft 77). Frankfurt am Main: Klostermann, S. 202 – 212

Krause, Jürgen: Standardisierung und Heterogenität. In: Kuhlen, Rainer u.a. (Hrsg.): Grundlagen der praktischen Information und Dokumentation. 5. Aufl. München : Saur, 2004, S. 635-642

Liste der Anwender der RVK, Stand: 04.02.2009
URL: <http://www.bibliothek.uni-regensburg.de/Systematik/pdf/anwender.pdf> (Zugriff am 03.06.2009)

Lorenz, Bernd: Handbuch zur Regensburger Verbundklassifikation : Materialien zur Einführung. 2. Aufl. Wiesbaden : Harrassowitz, 2008

Manecke, Hans-Jürgen: Klassifikationen, Klassieren. In: Kuhlen, Rainer u.a. (Hrsg.): Grundlagen der praktischen Information und Dokumentation. 5. Aufl. München : Saur, 2004, S. 127-140

MILOS I/II - Kurzbeschreibung der Projektinhalte
URL: http://www.ub.uni-duesseldorf.de/home/ueber_uns/projekte/abgeschlossene_projekte/milos/mil_kurz (Zugriff am 09.06.2009)

Notationen der SWD-Sachgruppen
URL: http://www.d-nb.de/standardisierung/pdf/swd_syst.pdf (Zugriff am 03.06.2009)

Oberhauser, Otto: Automatisches Klassifizieren : Entwicklungsstand – Methodik – Anwendungsbereiche. Frankfurt am Main u.a. : Lang, 2005

Online-SWD
URL: <http://swb.bsz-bw.de/DB=2.104/> (Zugriff am 03.06.2009)

Panzer, Michael: Semantische Integration heterogener und unterschiedlichsprachiger Wissensorganisationssysteme: CrissCross und jenseits. In: Heinz Peter Ohly u.a. (Hrsg.): Kompatibilität, Medien und Ethik in der Wissensorganisation. Proceedings der 10. Tagung der Deutschen Sektion der Internationalen Gesellschaft für Wissensorganisation Wien, 3. - 5. Juli 2006. Würzburg : Ergon, 2008, S. 61-69

Pfeffer, Magnus: Automatische Vergabe von RVK-Notationen anhand von bibliografischen Daten mittels fallbasiertem Schließen. 2007

Pfeffer, Magnus: Classification-Weblog der UB Mannheim

URL: <http://blog.bib.uni-mannheim.de/Classification/> (Zugriff am 03.06.2009)

Protokoll des RVK-Anwendertreffens 2008

URL: [http://www2.bsz-](http://www2.bsz-bw.de/cms/public/arbeit/agsach/Protokoll_RVK_Anwendertreffen_2008.pdf)

[bw.de/cms/public/arbeit/agsach/Protokoll_RVK_Anwendertreffen_2008.pdf](http://www2.bsz-bw.de/cms/public/arbeit/agsach/Protokoll_RVK_Anwendertreffen_2008.pdf) (Zugriff am 09.06.2009)

Raffelt, Albert (Hrsg.): Die Bibliothek von außen und von innen : Aspekte Freiburger Bibliotheksarbeit ; für Bärbel Schubel. Freiburg im Breisgau : Universitätsbibliothek, 2008

URL: [http://www.freidok.uni-](http://www.freidok.uni-freiburg.de/volltexte/5000/pdf/bibliothek_aussen_innen.pdf)

[freiburg.de/volltexte/5000/pdf/bibliothek_aussen_innen.pdf](http://www.freidok.uni-freiburg.de/volltexte/5000/pdf/bibliothek_aussen_innen.pdf) (Zugriff am 03.06.2009)

Recker, Ingrid/Ronthaler, Marc/Zillmann, Hartmut: OSIRIS : Osnabrück Intelligent Research Information System ; ein Hyperbase Front End System für OPACs. In: Bibliotheksdienst 30 (1996) 5, S. 833-848

Regeln für den Schlagwortkatalog: Formschlagwörter : § 501 – Definition

URL: http://deposit.ddb.de/ep/netpub/89/96/96/967969689/_data_stat/www.dbi-berlin.de/dbi_pub/einzelpu/regelw/rswk/rswk_07.htm (Zugriff am 03.06.2009)

Regensburger Verbundklassifikation : 16 ; Geographie (R):

URL: http://www.bibliothek.uni-regensburg.de/rvko/pdf/rvko_R.pdf (Zugriff am 03.06.2009)

Rundbrief zur Regensburger Verbundklassifikation 13 (1997) 1

URL: http://www.bibliothek.uni-regensburg.de/Systematik/rundbriefe/1997_01.pdf (Zugriff am 09.06.2009)

RVK-Online

URL: http://www.bibliothek.uni-regensburg.de/rvko_neu/ (Zugriff am 03.06.2009)

Scheven, Esther: Effiziente Sacherschließung in schwierigen Zeiten : Gedanken zur Zukunft der SWD. In: Bibliotheksdienst 39 (2005) 6, S. 748-773

Universitätsbibliothek Regensburg: Neumeldung von Notationen

URL: http://www.bibliothek.uni-regensburg.de/rvko_neu/formular/neumeldung.phtml (Zugriff am 03.06.2009)

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Bachelorarbeit selbständig angefertigt habe. Es wurden nur die in der Arbeit ausdrücklich benannten Quellen und Hilfsmittel benutzt. Wörtlich oder sinngemäß übernommenes Gedankengut habe ich als solches kenntlich gemacht.

Ort, Datum

Unterschrift