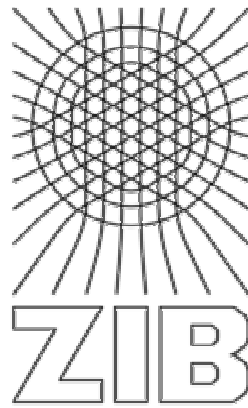

Konrad-Zuse-Zentrum
für Informationstechnik Berlin



Takustraße 7
D-14195 Berlin-Dahlem
Germany

ANDRES IMHOF

**RSWK/SWD und Faceted Browsing:
neue Möglichkeiten einer
inhaltlich-intuitiven Navigation**

**Gefördert
von der Senatsverwaltung für Wissenschaft, Forschung und Kultur des Landes Berlin,
vom Ministerium für Wissenschaft, Forschung und Kultur des Landes Brandenburg
und von den Mitgliedsbibliotheken des KOBV**

RSWK/SWD und Faceted Browsing: neue Möglichkeiten einer inhaltlich-intuitiven Navigation

Andres Imhof

Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)

ZIB Report 06-34

Abstract

Das deutschsprachige Bibliothekswesen verfügt mit den „Regeln für den Schlagwortkatalog“ (RSWK) unter Verwendung der „Schlagwortnormdatei“ (SWD) über ein Instrumentarium, welches zusammen mit einem „Faceted Browsing“ das bisher bestehende Angebot für ein Information Retrieval optimal ergänzen kann. Die Verbindung zwischen Standardvokabular (SWD) und Kettenbildung (RSWK) einerseits und eine nach Facetten-Eigenschaften gegliederte Navigation andererseits unterstützt bestmöglich eine inhaltlich bezogene Recherche. Die Stärken und Schwächen der RSWK/SWD werden erörtert und auch Klassifikationen (DDC und RVK) als mögliche Facetten diskutiert.

Keywords: RSWK, SWD, Faceted Browsing, Facetten, facettierte Navigation, Information Retrieval

CR: H.0K, 4.0

Inhalt

Einleitung.....	3
RSWK/SWD.....	3
Faceted Browsing	4
Anwendung in Forschung und Lehre.....	5
Auswahl der Facetten	5
Unzulänglichkeiten der RSWK/SWD.....	6
Vorteile der SWD in einem Faceted Browsing	7
Verwendung von Klassifikationen als Facette.....	7
Sacherschließungskonzept der Deutschen Nationalbibliothek.....	8
Der KOBV und Faceted Browsing.....	9
Zusammenfassung	9

Einleitung

Die kontinuierliche Optimierung der Informationsversorgung ist ein von allen Seiten unterstütztes Ziel in Forschung und Lehre.¹ Neben der Verfügbarkeit an Literatur ist das Organisieren derselben, aber auch das Suchen, Finden und Erhalten von Informationen angesprochen. Die Anstrengungen der vergangenen Jahre für eine verbesserte Literaturrecherche haben in der Bibliothekswelt bereits zu einigen Erfolgen geführt. Inzwischen ist es selbstverständlich geworden, über das Internet in deutschen und internationalen Online-Katalogen nach relevanter Literatur zu suchen. Weiterführende Dienste wie Dokumentenlieferung oder direkte Verlinkung zu elektronischen Volltexten sind dabei nahtlos eingebunden. Die Weiterentwicklung bibliothekarischer Dienstleistungen werden ohne Unterbrechung vorangetrieben, zumal freie Suchmaschinen wie Google im „Information Retrieval“, insbesondere durch Google-Scholar² und Google-Buchsuche³, einen beträchtlichen Konkurrenzdruck auf die Bibliotheken ausüben. Dabei drängt sich insgesamt der Eindruck auf, dass die Bibliotheken gegenüber den freien Suchmaschinen deren Entwicklungen hinterherlaufen.

Ohne es als solches wahrzunehmen, verfügen die Bibliotheken im deutschsprachigen Raum jedoch über ein Instrument, mit dem sie im Wettbewerb mit Google um die NutzerInnen wieder ihre Vorzüge herausstellen können. Es existiert längst ein Konzept, mit dessen Hilfe die Literaturrecherche inhaltlich-intuitiv gestaltet und damit ein Vorteil der Bibliotheksrecherche gegenüber kommerziellen Suchmaschinen herausgearbeitet werden kann. Das Konzept heißt „Regeln für den Schlagwortkatalog“ (RSWK)⁴ und bietet im Zusammenhang mit dem mittlerweile nicht mehr ganz so neuen Navigationskonzept „Faceted Browsing“ ungeahnte Möglichkeiten eines intuitiven Recherchewerkzeuges, das Literatur gezielt inhaltlich entdecken lässt.⁵

RSWK/SWD

Die RSWK sind ein Ansatz, in kürzester Form eine Zusammenfassung vom Inhalt eines Buches abzubilden.⁶ Von grundsätzlicher Bedeutung ist die Vergabe der Schlagworte nach der Schlagwortnormdatei (SWD).⁷ Dabei werden Schlagwortketten⁸ gebildet, die aus Schlagworten verschiedener Kategorien bestehen.⁹ Die Nennung eines Schlagwortes in einer Kette hat hierbei eine bestimmte Bedeutung ebenso wie das Nicht-Vorhandensein. Handelt ein Buch laut Titelangabe beispielsweise von alten Bauernhäusern in der Schweiz, dann ist in der Schlagwortkette die Nennung der „Schweiz“ als geografische Angabe, darüber hinaus die Angabe von „Bauernhaus“, „Inventar“ und „Fachwerkbau“ als Sachschlagworte sowie die Vergabe eines Zeitschlagwortes sinnvoll. Wird die Schweiz in der Schlagwortkette nicht genannt, handelt es sich um ein Werk, das (entgegen der

¹ Grundlage dieses Artikels ist ein Vortrag auf dem vom Kooperativen Bibliotheksverbund Berlin-Brandenburg (KOBV) am 13.02.2006 veranstalteten Workshop „Verbesserung des Retrievals“: Einsatz von Suchmaschinen-Technologie in der Navigation – Konzept für ein Faceted Browsing, Volltext unter: http://www.kobv.de/deutsch/content/wir_ueber_uns/events/2006-02-13/imhof.pdf (Zugriff: 26.06.06)

² Siehe <http://scholar.google.de/>

³ Siehe <http://books.google.com/>

⁴ Siehe http://deposit.ddb.de/ep/netpub/89/96/96/967969689/_data_stat/www.dbi-berlin.de/dbi_pub/einzelpu/regelw/rswk/rswk_00.htm (Zugriff: 26.06.06)

⁵ Unter dem Begriff „Recherche“ wird in diesem Textzusammenhang sowohl die Suche, als auch das Browsing verstanden, während mit „Navigation“ nur das Browsing gemeint ist.

⁶ Vgl. Heiner-Freiling, Magda: RSWK und DDC – Sacherschließung auf zwei Beinen. In: Dialog mit Bibliotheken 17, 2005, 3, S. 4 – 13. Volltext unter: http://www.ddc-deutsch.de/literature/2005_3_RSWKundDDC.pdf, hier S. 4 (Zugriff: 26.06.06)

⁷ Siehe Grundregeln, §2, 2. http://deposit.ddb.de/ep/netpub/89/96/96/967969689/_data_stat/www.dbi-berlin.de/dbi_pub/einzelpu/regelw/rswk/rswk_02.htm#%A7%202, zur Schlagwortnormdatei siehe <http://www.ddb.de/standardisierung/normdateien/swd.htm> (jeweils Zugriff: 26.06.06)

⁸ Ebenda §13, 1. http://deposit.ddb.de/ep/netpub/89/96/96/967969689/_data_stat/www.dbi-berlin.de/dbi_pub/einzelpu/regelw/rswk/rswk_02.htm#%A7%2013 (Zugriff: 26.06.06)

⁹ Ebenda § 15, 1. http://deposit.ddb.de/ep/netpub/89/96/96/967969689/_data_stat/www.dbi-berlin.de/dbi_pub/einzelpu/regelw/rswk/rswk_02.htm#%A7%2015 (Zugriff: 26.06.06)

Angabe im Titel) die Bauernhäuser unabhängig vom Ort beschreibt. Die Schlagwortkette ist eine Aneinanderreihung von wichtigen Sachverhalten und Eigenschaften (Facetten), die erst zusammen ein Gesamtbild abgeben. Diese Beschaffenheit der RSWK/SWD kann nun mit Hilfe des Faceted Browsing ihre optimale Wirkung entfalten.

Faceted Browsing

Das Faceted Browsing oder auch die facetiierte Navigation ist eine Form des Browsens, die zunehmend in kommerziellen und wissenschaftlichen Informationsportalen Einzug hält. Sinn des Faceted Browsing ist, durch die Auswahl verschiedener Facetten ein unübersichtliches Angebot sukzessive auf das gewünschte Ziel einzuschränken. Mit Facetten sind hierbei einzelne, maßgebliche Eigenschaften der gesuchten Ergebnisse gemeint. Die Ursprünge des Gebrauchs von Facetten in der Suche nach Informationen reichen weit in das 18. Jahrhundert zurück, wobei das Wort „Facette“ erstmals in der 1933 veröffentlichten Colon Classification des Shiyali Ramamrita Ranganathan (1892 – 1972) verwendet wurde.¹⁰

Welche Funktionen machen ein Faceted Browsing aus?

- Nach jeder Auswahl einer Facette werden alle übrigen Facetten neu berechnet. Das Angebot an Facetten wird mit jeder weiteren Auswahl geringer.
- Innerhalb der Navigation wird grundsätzlich angezeigt, welche Facetten bereits ausgewählt sind.
- Jede Auswahl einer Facette führt zu mindestens einem Treffer. Facetten ohne Inhalt werden nicht angeboten.
- Zu jeder angebotenen Facette wird die Anzahl der enthaltenen Treffer angezeigt.
- Jede ausgewählte Facette kann wieder entfernt werden, wonach wiederum die angebotenen Facetten neu generiert werden. Die Ergebnisliste wird wieder länger.¹¹

An der University of California in Berkeley wurde 2003 eine Nutzerstudie in Bezug auf eine umfangreiche Bildsammlung mit Faceted Browsing publiziert.¹² Das Projekt mit dem Namen Flamenco¹³ mündete in eine konkrete Software, die seit 2006 unter BSD-Lizenz¹⁴ in der Release 1.0 als Open-Source zum Download angeboten wird.¹⁵ Das Faceted Browsing fand vielfachen Eingang in Suchmaschinen wie Endeca¹⁶, FAST¹⁷ u.a. Viele kommerzielle Portale haben den Nutzen dieser Navigation für sich entdeckt. Eines der prominentesten Beispiel dürfte das Online-Auktionshaus eBay sein, das eine etwas einfachere Form, und zwar ein hierarchisch gegliedertes, facetiiertes Browsen mit einer eigenen Klassifikation anbietet.¹⁸ Hier kann nicht jede Facette wieder entfernt, sondern immer nur wieder in die oberen Hierarchiestufen der Kategorie zurückgekehrt werden.

¹⁰ Einen knappen historischen Abriss der facetiierten Suche findet man bei Maple, Amanda: Faceted Access: A Review Of The Literature. 1995. Volltext unter: http://library.music.indiana.edu/tech_s/mla/facacc.rev (Zugriff: 26.06.06)

¹¹ Vgl. Denton, William: How to Make a Faceted Classification an Put It On the Web. Nov. 2003, siehe <http://www.miskatonic.org/library/facet-web-howto.html>, Kap. 4.2 (Zugriff: 26.06.06)

¹² Siehe Yee, Ka-Ping; Swearingen, Kirsten; Li Kevin; Hearst, Marti: Faceted metadata for image search and browsing. In: Proceedings of CHI 2003. p. 401 – 408, ACM. Volltext unter: <http://bailando.sims.berkeley.edu/papers/flamenco-chi03.pdf> (Zugriff: 26.06.06)

¹³ FLexible information Access using MEtadata in Novel COmbinations, siehe <http://flamenco.berkeley.edu/> (Zugriff: 26.06.06)

¹⁴ Siehe <http://flamenco.berkeley.edu/license.html> (Zugriff 26.06.06)

¹⁵ Siehe ebenda <http://flamenco.berkeley.edu/download.html> (Zugriff: 26.06.06)

¹⁶ Siehe <http://endeca.com/>

¹⁷ Siehe <http://www.fast.no>

¹⁸ Siehe <http://www.ebay.de>

Anwendung in Forschung und Lehre

Für den wissenschaftlichen Bereich ist diese Navigationsart gleichermaßen interessant. In exzellenter Art und Weise zeigt dies der Online-Katalog der New York City State University (NCSU)¹⁹, der mit der Suchmaschine von Endeca betrieben wird. Nach einer Suche werden die Ergebnisliste und parallel dazu verschiedene Facetten angezeigt, um gegebenenfalls anhand dieser zu browsen. Als Facetten werde u.a. angeboten: Sachschlagwort, Genre der Veröffentlichung, Format, Bibliothek, Region, Epoche, Sprache und Autor vermischt mit Körperschaft. Die Bibliothek der NCSU ist dem anvisierten Ziel schon recht nahe und führt exemplarisch vor, was für deutschsprachige Online-Kataloge hier vorgestellt werden soll.

In Deutschland betreibt das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz)²⁰ eine Suchmaschine mit der FAST-Technologie für den Dreiländerkatalog.²¹ Im Dreiländerkatalog werden ebenfalls Einschränkungsmöglichkeiten nach Facetten angeboten. Man kann diesbezüglich jedoch nicht von einem Faceted Browsing sprechen, denn ein entscheidendes Kriterium wird nicht erfüllt: Die zur Einschränkung ausgewählten Facetten werden nicht angezeigt, so dass darüber eine ausgewählte Facette nicht wieder entfernt werden kann. Die NutzerInnen sind so nicht in der Lage, die Auswahl abwechselnd einzuschränken und wieder zu erweitern. Insgesamt richtet sich die Auswahl der Facetten auf wichtige Metadatenfelder wie Erscheinungsdatum, Schlagwort, Sprache, Publikationstyp, Autor und Körperschaft, die der Katalog anbietet. Alle Schlagworttypen werden im Gegensatz zur NCSU als eine Gesamtheit aufgefasst und nur in einer Facette zusammen aufgeführt. Autor und Körperschaft werden dagegen im Dreiländerkatalog einzeln als Facetten angeboten, während die NCSU diese in einer Facette vereint. Beide Systeme bieten die Sprache einer Publikation als Facette an und auch der Zeitraum wird sehr geschickt in groben Zeiträumen angegeben, wobei im Dreiländerkatalog sogar der Zeitraum schrittweise spezifiziert werden kann.

Auswahl der Facetten

Für den Aufbau eines Faceted Browsing auf einem Online-Katalog bleibt die jeweilige Entscheidung für spezifische Facetten ausschlaggebend.²² Im Folgenden wird erörtert, welche Metadaten eines Online-Kataloges relevant für eine facettierte Navigation sein könnten. Drei Bedingungen der Metadaten sollten vorweg für die Auswahl der Facetten berücksichtigt werden:

1. Jede Facette sollte für den überwiegenden Teil der Datensätze vorhanden sein, im Idealfall zu 100 Prozent.
2. Sie sollte je nach Facettentyp (Schlagworte, Erscheinungsdatum, Klassifikation etc.) nur von einem Standard unterstützt werden. Es sollten keine unterschiedlichen Klassifikationssysteme wie DDC, RVK oder SWD und freie Schlagworte vermischt werden.
3. Die Werte einer Facette, d.h. die Inhalte bzw. Benennungen, sollten nicht zu lang und außerdem allgemein verständlich sein. Reine Notationen aus Zahlen- und Buchstabenkombinationen werden nicht verstanden.

Wie die beiden genannten Beispiele der NCSU und des hbz zeigen, werden neben der inhaltlichen Erschließung durch Schlagworte und Klassifikationen vor allem Autor/Körperschaft, Publikationstyp, Erscheinungsjahr und Publikationssprache als Facetten angeboten. Ein Grund hierfür wird vor allem sein, dass diese Informationen bei der überwiegenden Zahl der Katalogsätze vorhanden sind. Machen diese Facetten aber auch bei einer inhaltlichen Suche Sinn? Sucht man nach einem Titelstichwort, werden die ersten fünf Autoren angezeigt, die am häufigsten publiziert haben. Die Menge der Publikationen ist jedoch – inhaltlich betrachtet – kaum relevant. Eine Autorin, die an 20. Stelle aufgeführt wird, weil sie nur einen Artikel geschrieben hat, kann für die Suche wichtiger sein, als der erst genannte Autor, was auch für Körperschaften gilt. Grundsätzlich ist zu unterscheiden, dass nach

¹⁹ Siehe <http://www.lib.ncsu.edu/searchcollection/> (Zugriff: 26.06.06)

²⁰ Siehe <http://www.hbz-nrw.de/>

²¹ Siehe <http://suchen.hbz-nrw.de/dreilaender/> (Zugriff: 26.06.06)

²² William Denton beschäftigt sich umfassend mit diesem schwierigen Thema, ebenda Denton (2003)

manchen Metadaten gezielt gesucht wird, wie Titel, Autor, Körperschaft. Bei inhaltlich bezogenen Rechercheansätzen wird eher nach Schlagworten oder Klassifikationen gebrowst.

Der Publikationstyp und die Publikationssprache können eine willkommene Einschränkungsoption sein, aber für eine inhaltliche Navigation bieten sie zunächst kaum eine Hilfe. Es ist außerdem nicht anwendungsfreundlich, für die inhaltliche Navigation alle relevanten sachlichen Erschließungsdaten in eine einzige Facette einfließen zu lassen wie beim Dreiländerkatalog. Die NCSU geht in dieser Hinsicht mit gutem Beispiel voran und betrachtet differenziert unterschiedliche Schlagworttypen in verschiedenen Facetten. An dieser Stelle kann nun auch die RSWK/SWD ansetzen. Sie unterscheidet – wie bereits genannt – verschiedene Schlagwortkategorien, wie Geographikum, Person (und zwar inhaltlich, nicht Autor), Sachschlagwort, Zeitschlagwort sowie Formschlagwort. Der Rückgriff auf die SWD macht dabei das Konzept „annähernd“ perfekt. Weitere Facetten wie Sprache oder Publikationsformat könnten jederzeit über eine individuelle Einstellungsmöglichkeit zu den voreingestellten RSWK-Facetten hinzugewählt werden.

Unzulänglichkeiten der RSWK/SWD

Bewusst wird das Konzept „annähernd“ perfekt genannt. Zu den RSWK/SWD ist auch kritisches zu nennen, da sie in mancherlei Hinsicht problematisch sind. Mönnich meint: „Die RSWK sind der ehrgeizige Versuch alle Wissensgebiete zu bedienen. In der Praxis zeigt sich, dass die RSWK für die Geisteswissenschaften gut, für die Natur- und Ingenieurwissenschaften sowie für die Medizin nur bedingt geeignet sind, da in diesen Fächern die begriffliche Dynamik wesentlich höher ist. Hinzu kommt, dass die RSWK stark an den Bedürfnissen eines eindimensionalen Zettelkataloges ausgerichtet sind und für Online-Kataloge gravierende Nachteile aufweisen (Regeln zur Pleonasmusvermeidung, Epochen und Zeitangaben, Kettenbildung).“²³ Je nach Fachgebiet zeigt das normierte Vokabular der SWD Stärken und Schwächen. Vor allem die fehlende Standardisierung der Zeitangaben und Formschlagworte, die in RSWK ohne SWD-Identifizierung vergeben werden, ist für eine Umsetzung in einem Faceted Browsing schwierig. Dadurch erhöht sich unnötiger Weise die Anzeige von Begriffen in den beiden Facetten „behandelter Zeitraum“ und „Publikationsform“ um die nicht abgesprochenen Mehrfachnennungen gleicher Bedeutung. Nichtsdestotrotz sind diese fehlenden Standards bei Zeit und Publikationsart noch handhabbar. Zukünftig wird in der Anzeige der Facetten durch übergeordnete Algorithmen sicherlich eine übersichtliche und logische Zusammenführung ähnlicher Benennungen möglich sein.

Die fehlende Abbildung einzelner Teilbereiche der Wissenschaft in der SWD ist schon gravierender. Hier gilt es, die SWD fortan zu komplettieren und eventuell vorhandenes, auch fremdsprachiges, Standardvokabular zu übernehmen. Sicherlich trägt dazu bei, dass die Wissenschaftssprache einiger Fachgebiete englisch ist, warum diese nicht in der SWD vertreten sind, da die SWD vorwiegend deutsches Vokabular führt.

Mönnich weist mit Recht auf die Orientierung der RSWK an der Eindimensionalität eines Zettelkataloges hin. Kettenbildung, Permutationen und Synonyme haben zwangsläufig eine Pleonasmusbildung²⁴ zur Folge. Wird also der Zettelkatalog eins zu eins in einen Online-Katalog übertragen, hat das eine unnötige Überfrachtung der Bildschirmanzeige zur Folge. Dies kann zukünftig umgangen werden, indem bei der Darstellung einige Regeln der RSWK ignoriert und durch einfachere ersetzt werden. Überträgt man die RSWK/SWD auf die Mehrdimensionalität eines Suchmaschinen-Indexes mit Faceted Browsing, kann auf die Bildung mehrerer Ketten, die zum Teil Schlagworte wiederholen und sich nur in einem Schlagwort u.U. unterscheiden, gänzlich verzichtet werden. Innerhalb eines Indexes würden alle mehrfach aufgeführten Schlagworte auf jeweils eine Nennung reduziert werden. Danach ist nicht mehr die Reihenfolge innerhalb der Kette von Bedeutung, sondern ausschließlich die gemeinsame Erwähnung in einem Gesamt-Schlagwort-Gebilde, das in kürzester Form den Inhalt eines Werkes darstellt. Unumgänglich bleibt allerdings die Gliederung aller Schlagworte in die oben genannten Schlagwortkategorien, die durch einen Indikator gekennzeichnet werden. So werden die zur Schlagwort-Gruppe mutierten Schlagwortketten für die Facettenbildung differenziert. Die zuvor aufwändige Umsetzung der RSWK/SWD in der Eindimensionalität wird nun vereinfacht, wodurch letztlich eine Mehrdimensionalität bedient wird.

²³ Siehe M. Mönnich: Die Erschließung, 2000, S. 5. Elektronischer Volltext unter: <http://www.ubka.uni-karlsruhe.de/cgi-bin/psgunzip/2000/zentral/digibib/04-Moennich/04-Moennich.pdf> (Zugriff: 26.06.06)

²⁴ Pleonasmus= unbrauchbare Sammlung gleichbedeutender Ausdrücke

Vorteile der SWD in einem Faceted Browsing

Bislang war es problematisch, NutzerInnen eine Suche nach der SWD anzubieten. Wer weiß schon, wie ein bestimmter Sachverhalt in der SWD abgebildet ist? Ein mit der SWD übereinstimmendes Wort in der Suche einzugeben, ist für NutzerInnen schwierig, meist gar nicht als möglicher Sucheinstieg bekannt und in der Suche eher zufällig. Der volle Nutzen eines standardisierten Vokabulars kommt erst in einem Faceted Browsing richtig zum Tragen. Das standardisierte Vokabular muss nicht bekannt sein, es wird ganz einfach zur Auswahl dynamisch angeboten. Für die Begriffsauswahl in der facettierten Navigation reicht jeweils der erste Begriff, also das Hauptschlagwort. Pleonasmen treten demnach für die NutzerInnen nicht mehr in Erscheinung. Dennoch gehen die Vorteile der Nennung aller Synonyme zum Hauptschlagwort nicht verloren. In der elektronischen Datenverarbeitung ist es mit einfachen Mitteln möglich, bei der konkreten Suche mit Schlagworten die zu den Hauptschlagworten zugehörigen Synonyme mit einzubinden.

Verwendung von Klassifikationen als Facette

Neben der RSWK/SWD werden in der Sacherschließung auch Notationen aus der Regensburger Verbundklassifikation (RVK)²⁵ oder der Dewey Decimal Classification (DDC)²⁶ vergeben. Letztere hat hierfür eine deutsche Übersetzung durch die Deutsche Nationalbibliothek erfahren.²⁷ Klassifikationen sind für die inhaltliche Erschließung jedoch nur eine eindimensionale Facette und können daher eher für eine hierarchische Navigation verwendet werden. Vor allem die RVK, die als Aufstellungssystematik konzipiert ist und hier ihre Stärken hat, ist für eine Navigation innerhalb eines Informationsportals nur bedingt einsetzbar.

Als Voraussetzung gilt, dass die Notationen in der Anzeige in ihre Benennungen aufgelöst werden müssen. Die wenigsten NutzerInnen (und auch BibliothekarInnen) kennen die Buchstaben-Zahlen-Folgen auswendig. Nehme man nun eine Notation aus der vierten Ebene, so müssten häufig die Benennungen der vorhergehenden Ebene(n) hinzugeholt werden, um den Sinn und das Thema der besagten Ebenen deutlich werden zu lassen.²⁸ Dadurch würden aber die Benennungen einer Klassifikation für eine Facette zu lang werden und damit unbrauchbar für ein Faceted Browsing. Ein großes Manko sind ohnehin die zahlreichen langen Benennungen. Damit verstößt die RVK gegen die dritte Bedingung benötigter Facetteneigenschaften von kurzen, knappen und verständlichen Benennungen.

Bei der DDC ist das Problem etwas anders gelagert. Fast jede Benennung der DDC kann für sich stehen, alle nachstehenden Ebenen können in diesem Dezimalsystem problemlos der jeweilig vorstehenden Ebenen zugeordnet werden. Ein weiterer Vorteil ist die große Verbreitung der DDC im internationalen und zunehmend auch nationalen Bibliothekswesen. So können die DDC-Klassifikationen über eine Fremddatenübernahme auch nachträglich bezogen werden. Damit wird die erste Bedingung für gute Facetteneigenschaften im Vergleich zu anderen Facetten noch am besten erfüllt. Allerdings ist die DDC nur bis zur dritten Ebene kostenfrei. Die übrigen Ebenen müssen von den Bibliotheken, die sie anwenden, lizenziert werden. Nur die ersten drei Ebenen zu verwenden, würde den Wirkungsgrad des Faceted Browsing nicht ausschöpfen. Einige Themenbereiche sind dabei sehr differenziert beschrieben wie die Religion, während andere völlig unterrepräsentiert sind.²⁹ Auch sind viele Benennungen sehr lang, womit wiederum die dritte Bedingung der Facetteneigenschaften nur unzulänglich unterstützt wird.

Für ein facetiertes Browsing mögen diese beiden Klassifikationen nicht optimal sein, jedoch kann mit ihnen sehr gut ein hierarchisches Browsing aufgesetzt werden, das Eigenschaften des Faceted Browsing beinhaltet. Eine erste Gestaltung eines hierarchischen Browsing auf Grundlage der RVK hat

²⁵ Siehe <http://www.bibliothek.uni-regensburg.de/Systematik/systemat.html> (Zugriff: 26.06.06)

²⁶ Siehe <http://www.oclc.org/dewey/>

²⁷ Siehe <http://www.ddc-deutsch.de/>

²⁸ Beispielsweise ist „IF 1700 - IF 3849 Literaturgeschichte“ (4. Ebene) ohne die vorhergehende Notation und Benennung (IF 1000 - IF 3849 Französische Literatur im 16. Jahrhundert) nicht verständlich.

²⁹ Z.B. erscheint die Veterinärmedizin (636.089 Veterinärwissenschaften/Tiermedizin) erst in der vierten Ebene.

bereits das Bibliotheksservice Zentrum Baden-Württemberg (BSZ)³⁰ mit BibScout in Produktion gestellt.³¹ In der linken Navigationsleiste können sich NutzerInnen innerhalb der RVK bewegen, die sich in sich ausfächernden Pfaden einer hierarchischen Baumstruktur gliedert, und Titel zu den spezifischen Themen anzeigen lassen. Diese Vorgehensweise ist insofern eindimensional, weil man immer bis zu einer nicht mehr weiterführenden Ebene gelangt und von dieser wieder zurückkehren muss, um in andere Endpunkte der RVK-Pfade gelangen zu können. Es wird jeweils nur eine Eigenschaft eines Titels genutzt, Kombinationen mehrerer Eigenschaften sind nicht möglich. Allein deswegen kann man nicht von einem Faceted Browsing sprechen, sondern von einem hierarchischem Browsing. Außerdem werden alle RVK-Benennungen angezeigt, auch wenn sie keinen Titel enthalten. NutzerInnen können also Null Treffer in einem Pfadende vorfinden. Darüber hinaus fehlen die dynamischen Angaben der jeweils enthaltenen Titel in einer Notation.

Dennoch mag das hierarchische Browsing für manche Suchstrategien seine Daseinsberechtigung besitzen. Es wäre lediglich um zwei dynamische Eigenschaften des Faceted Browsing zu erweitern: Für jede Notation wird die aktuelle Titelanzahl genannt, während leere Notationen gar nicht erst angezeigt werden. Damit wäre eine Browsing-Struktur analog zum bereits erwähnten Beispiel eBay geschaffen.

Sacherschließungskonzept der Deutschen Nationalbibliothek

Das bislang ungelöste und schwerwiegendste Problem beim Einsatz der gewünschten Facetten RSWK/SWD ist die noch zu geringe Durchdringung dieser Erschließungsdaten in den Online-Katalogen. Die erste Bedingung für eine Facette ist die Erfüllung dieser Information in möglichst allen Katalogsätzen. Hundert Prozent sind eine Utopie, vor allem bei älterer Literatur. Aber auch bei der neueren Literatur ist keine vollständige, inhaltliche Erschließung zu erwarten, zumal die Deutsche Nationalbibliothek, die mit ihrer Sacherschließung eine zentrale Dienstleisterin für die deutschen Bibliotheken ist, ihr Sacherschließungskonzept mit der Einführung der DDC geändert hat:³² Nunmehr gilt:

- A (Monografien und Periodika des Verlagsbuchhandels): Titel der Reihe A werden neben der Vergabe von RSWK ab Bibliografiejahrgang 2007 auch vollständig nach DDC verbal erschlossen.
- B (Monografien und Periodika außerhalb des Verlagsbuchhandels): Titel der Reihe B werden ab Bibliografiejahrgang 2006 nur nach DDC inhaltlich erschlossen, dafür aber komplett jedes Fachgebiet. Zuvor wurden alle geisteswissenschaftlichen Titel mit Ausnahme der Religion nach RSWK erschlossen. Das wird wegfallen.
- H (Hochschulschriften): Titel der Reihe H werden ebenfalls ab Bibliografiejahrgang 2006 nur noch nach DDC inhaltlich erschlossen.

Auch wenn die Reihe A den größten Anteil aller Titel pro Jahr umfasst, so fällt dennoch die Sacherschließung der RSWK/SWD für die Reihen B und H komplett weg. Wenn sich an diesem Sacherschließungskonzept nichts dahingehend ändern sollte, auch die fehlenden Reihen nach den RSWK/SWD inhaltlich zu erschließen, sollten automatisierbare Lösungen geprüft werden. Denkbar wäre zumindest für die elektronischen Volltexte, die Reihen B und H über eine automatische Klassifizierung und Verschlagwortung nach RSWK/SWD inhaltlich zu erschließen. Eine hierfür noch zu projektionierende Entwicklung wäre für die inhaltlich-bezogene Recherche sehr hilfreich und wirkungsvoll. Der Inhalt könnte eventuell mit Hilfe von Textwerkzeugen analysiert, die Schlagworte müssten je nach Häufigkeit sowie Bedeutung entnommen und anschließend mit der SWD abgeglichen werden. Auf Grundlage der spezifizierten Schlagwortketten wäre im Anschluss als Nebenprodukt eine zielgerichtete Zuordnung in die RVK und DDC möglich.

³⁰ Siehe <http://www2.bsz-bw.de/cms/>

³¹ Siehe <http://bibscout.bsz-bw.de/bibscout/> (Zugriff: 26.06.06)

³² Ebenda Heiner-Freiling (2005), S. 7f.

Der KOBV und Faceted Browsing

In den Bibliotheken des KOBV ist die Vergabe von RSWK/SWD uneinheitlich.³³ Außerdem ist der Anteil an Katalogsätzen mit RSWK/SWD im KOBV-Index noch nicht umfassend genug. Andere Verbünde stehen diesbezüglich besser da. Um das Ergebnis im KOBV-Index zu verbessern, ist eine nachträgliche Datenanreicherung über Abgleiche mit anderen Verbundkatalogen angedacht. Bis dahin wird bewusst in Kauf genommen, dass das Faceted Browsing nicht den kompletten Bestand abdecken wird. Der Nutzungsgewinn in Bezug auf die nach RSWK/SWD erschlossenen Titel wird höher eingeschätzt, als die Tatsache, dass manche Titel dabei nicht wahrgenommen werden.

Für ein KOBV-eigenes User-Interface mit Faceted Browsing auf Grundlage der RSWK/SWD ist die technische Voraussetzung ein Suchmaschinen-Index. Hierfür wird aktuell in der KOBV-Zentrale an der Umsetzung des KOBV-Indexes mit der Suchmaschinentechologie Lucene gearbeitet. Lucene ist eine Open-Source vertriebene Volltext-Suchmaschine auf Java-Basis.³⁴ Eine erste Beta-Version des Suchmaschinen-Indexes wird voraussichtlich zum vierten Quartal 2006 online zugänglich sein, wobei eine Implementierung des Faceted Browsing erst wenig später erfolgen wird. Die KOBV-Zentrale betrachtet diese neue Navigationsweise als eine der zentralen Verbesserungen im Information Retrieval, neben der Einbindung von – wenn vorhanden – Volltext-Indices und von Ranking-Algorithmen für die Darstellung der Treffer.

Zusammenfassung

Abschließend kann festgehalten werden, dass die RSWK/SWD für sich betrachtet in der bislang eingesetzten Form nicht ihre volle Wirkung im Information Retrieval erreicht. Das Faceted Browsing, für das bisher alle möglichen und unmöglichen Metadaten verwendet werden, ist zwar ein geeignetes Feature, liefert aber immer noch nicht den spürbaren Nutzungsgewinn. Die SWD, die leider noch nicht alle Wissenschaftsbereiche umfasst, und die unvollständige Erschließung mit RSWK/SWD ist zum jetzigen Zeitpunkt nicht vollends befriedigend. Doch ein Arrangement mit den Unzulänglichkeiten bis zur kontinuierlich verlaufenden Optimierung dieser Situation lohnt sich. Die RSWK/SWD und das Faceted Browsing spielen erst gemeinsam ihre Stärken richtig aus. Standardisiertes Vokabular (SWD) in einen Sinnzusammenhang gestellt (RSWK) bildet alle relevanten Eigenschaften eines Textes in kurzer, maschinenlesbarer Form (Faceted Browsing) ab. Auf diese Weise erhalten wir eine inhaltlich-intuitive Navigation über Bibliotheksbestände und weitere Literatur.

³³ Über den Kooperativen Bibliotheksverbund Berlin-Brandenburg (KOBV) siehe <http://www.kobv.de>

³⁴ Siehe <http://lucene.apache.org/java/docs/index.html> (Zugriff: 26.6.06)