

Suchmöglichkeiten in Primo auf dem Prüfstand

Dr. Philipp Zumstein
Universitätsbibliothek Trier
zumstein@uni-trier.de

04.11.11

Zusammenfassung

Primo ist ein Softwareprodukt der Firma Ex Libris und ermöglicht eine einheitliche Einschlitz-Suche im Bibliothekskatalog, Artikeldatenbanken und anderen Datenquellen wie etwa digitalen Repositorien.

An der Universitätsbibliothek Trier wird Primo eingeführt und die Artikeldatenbank Primo Central lizenziert mit dem primären Ziel, Zeitschriftenartikel und Bibliotheksbestand über eine Suchoberfläche zu erschließen. Im Folgenden wird insbesondere auf diese Implementierung eingegangen und einige Beobachtungen aus der Projektphase wiedergegeben. Die Arbeiten an Primo laufen bei uns momentan noch und eine öffentliche Version zur Ablösung des Online-Katalogs ist für 2012 geplant.

Nach einer kurzen [Einführung](#) und Beschreibung der [Funktionsweise von Primo](#) werde ich folgende Thesen weiter ausführen:

- Eine [einheitliche Einschlitz-Suche](#) wird aufgrund der Heterogenität der Metadatenqualität nur bedingt erreicht.
- Die Auswertung großer Treffermengen wird mit Primo insbesondere durch die [Facettierungen](#) in einem großen Masse unterstützt. Allerdings suggerieren die Facettenwerte eine Vollständigkeit, die in dieser Weise nicht gewährleistet wird. Ebenfalls können Facettenwerte abhängig von der Sortierreihenfolge verloren gehen.
- Die Darstellung der Daten kann man auf vielseitige Weise anpassen, so dass auch lokale Besonderheiten abgebildet werden können. Es wird als Beispiel gezeigt, wie wir die lokalen [Signaturen](#) in Primo zur Anzeige gebracht haben.

Am Ende folgt ein [Literaturverzeichnis](#) und eine [Liste mit Links](#).

Einführung

Es wird davon ausgegangen, dass der Leser, die Leserin Primo von der Benutzerperspektive her etwas kennt und auch schon einige Recherchen selbst einmal ausprobiert hat. Man kann dazu auch die [Link-Liste](#) am Ende dieses Dokuments benutzen, auf der auf einige Primo-Implementationen anderer Bibliotheken hingewiesen wird.

Ex Libris hat bei uns die Erstinstallation gemacht, Schulungen durchgeführt und bearbeitet Support-Anfragen. Des Weiteren liegt die Weiterentwicklung von Primo bei Ex Libris.

Unsere Primo-Installation wird vom kobv (Kooperativer Bibliotheksverbund Berlin-Brandenburg) gehostet, d.h. der kobv stellt die Hardware zur Verfügung, führt Software-Migrationen durch, beobachtet die Prozesse und greift gegebenenfalls ein. Damit befinden wir uns neben anderen Bibliotheken im kobv-Primo-Konsortium. Wie üblich werden in so einem Konsortium Absprachen getroffen und es findet ein verstärkter Erfahrungsaustausch statt. Ebenfalls kann man sich über EL Commons (Plattform von Ex Libris zum Austausch von Entwicklungen der Anwenderbibliotheken, siehe <http://www.exlibrisgroup.org/display/ElCommons/Home>) mit weiteren Primo-Anwendern austauschen.

Das Projektteam der UB Trier hat die Aufgabe, die Konfigurationen den lokalen Gegebenheiten und Wünschen anzupassen.

Technische Präambel

Suchergebnisse oder die Darstellung von einzelnen Treffern bilden jeweils eine Momentaufnahme und es ist nicht gewährleistet, dass diese ganz genau so wieder zu finden sind. Mit jedem neuen Buch können sich Suchergebnisse leicht verändern und eine Überarbeitung der Primo-Konfiguration kann zu leicht veränderten Darstellung von Treffern führen. Die beschriebenen Effekte sind aber für jeden leicht rekonstruierbar.

1 Funktionsweise von Primo

In diesem Abschnitt werden die Prozesse beschrieben, welche bei Primo intern ablaufen. Dabei wird zuerst auf die [Datenaufbereitung](#) eingegangen, die sich dem Thema widmet, welcher Vorlauf nötig ist, damit die Daten in Primo angezeigt und durchsucht werden können. Anschließend wird erklärt was bei einer [Suchanfrage](#) intern passiert und woher welche Daten stammen. Auf der nachfolgenden Seite wurde versucht diese Prozesse zu visualisieren, so dass man parallel beim Lesen des Textes die entsprechenden Sachverhalte auch graphisch verfolgen kann.

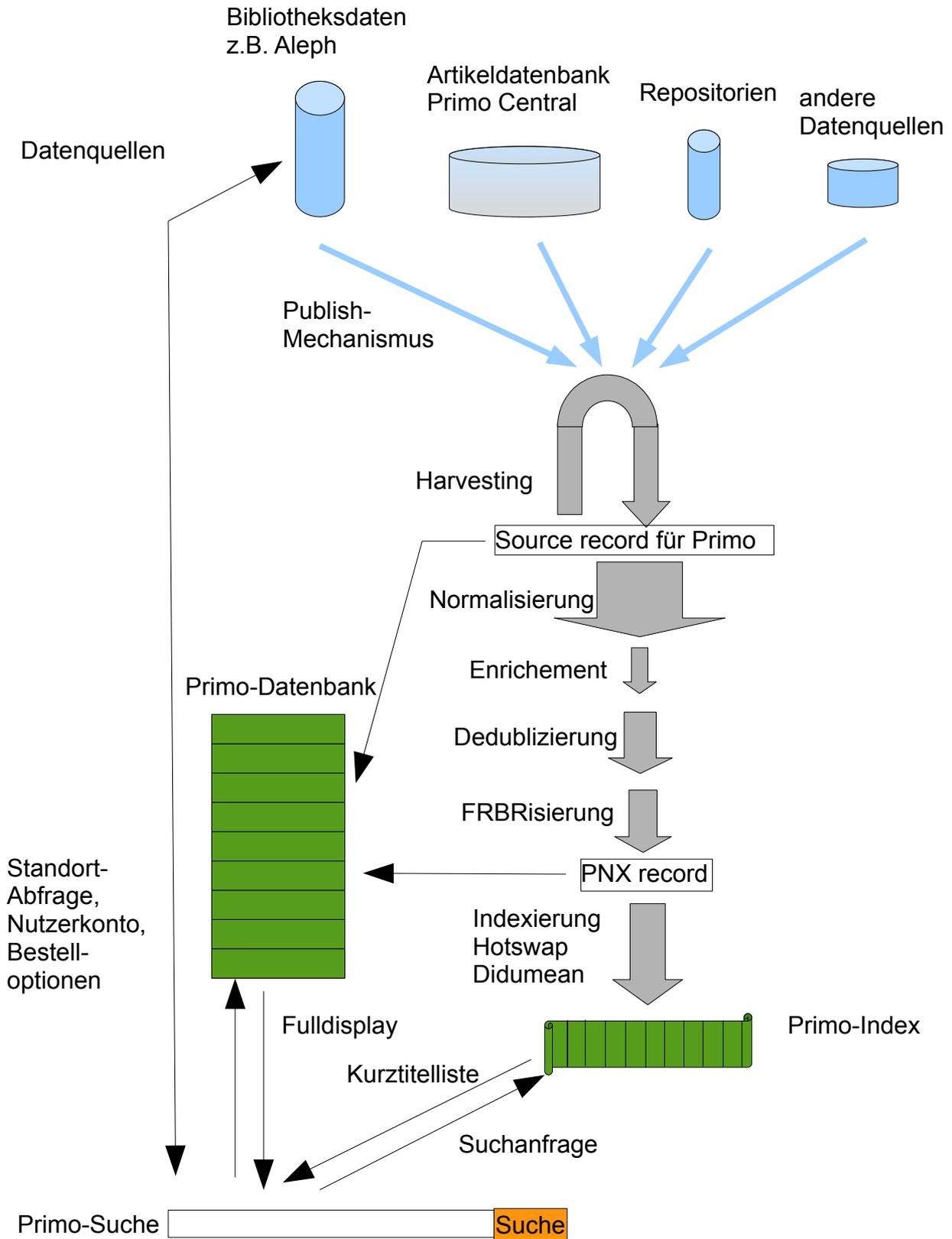
1.1 Datenaufbereitung

Unterschiedliche Datenquellen können in Primo eingebunden werden. Eine der Datenquelle besteht aus den Katalogdaten, die bei uns aus Aleph (Version 20) kommen. Diese Daten müssen zuerst einmal publiziert werden (publishing). Dieser Publishing-Mechanismus nimmt für jeden Titel die entsprechenden Daten und ergänzt diese zum Beispiel mit Normdaten, lokalen Notationen und Exemplardaten. Wir haben ebenfalls die Artikeldatenbank Primo Central als weitere Datenquelle, die wir von Ex Libris lizenziert haben. Andere Datenquellen, wie zum Beispiel Repositorien sind ebenfalls möglich. Diese müssen über entsprechende Schnittstellen verfügen.

Die Daten nach dem Publishing bilden die Quelldaten (source records) und werden für Primo in einem Verzeichnis abgelegt. Jede Änderung der Daten in Aleph (z.B. neue Titel, Korrekturen der Titelinformationen, Änderungen in Normdaten) wird täglich in Aktualisierungspaketen bereitgestellt. Primo holt sich diese Datenpakete jeweils durch einen Harvesting-Prozess. Der nächste Schritt heißt Normalisierung und führt die Daten in das PNX-Format (PNX = Primo normalized XML). Die Normalisierung wird über Normalisierungsregeln gesteuert und kann individuell angepasst werden. Im Prinzip handelt es hierbei um ein komplexes Datenmapping von den source records ins PNX-Format. Zum Beispiel kann man die Personen in den MAB-Feldern 100, 104, 108 etc. den entsprechenden PNX-Feldern display/autor und display/contributor zuordnen. Es ist auch möglich einen Enrichment-Prozess dazu zu schalten, um ergänzende externe Daten automatisch zu bekommen (wird momentan bei uns nicht genutzt).

Als nächstes folgt eine „Deduplizierung“, d.h. mögliche Dubletten werden erkannt und zusammengeführt. Da bei diesem Prozess Daten verloren gehen können, haben wir momentan darauf weitestgehend verzichtet. Danach folgt die „FRBRisierung“ (FRBR = Functional Requirements for Bibliographic Records), d.h. verschiedene Auflagen eines Werkes und unterschiedliche Erscheinungsformen (gedrucktes Buch und eBook) werden markiert. Diese Titel erscheinen bei einer Suche nur als ein Eintrag mit dem Vermerk, dass es mehrere Ausgaben gibt. Diese Ausgaben kann man sich dann mit einem weiteren Klick vollständig ansehen. Als Abschlussresultat erhält man den PNX record, welcher in der Primo-Datenbank gespeichert wird.

Grafik: Funktionsweise von Primo



Um die Daten suchbar zu machen, wird eine Indexierung benötigt. Dabei werden verschiedene Indexe aufgebaut und auch ein Gesamtindex (entspricht einer alle-Felder-Suche). Welche Daten indexiert werden kann man unabhängig von der Anzeige bei den Normalisierungsregeln festlegen. Zum Beispiel macht es Sinn die Verweisungsformen von Schlagwörtern suchbar zu machen, aber nur jeweils die Grundform anzuzeigen. Nach dem Indexieren muss ein Hotswap-Prozess eingeschaltet sein, der die alten Indexe durch die neuen ersetzt. Schlussendlich kann man einen Didumean-Prozess (did you mean) starten.

1.2 Bearbeiten einer Suchanfrage

Suchbegriffe werden in den Suchschlitz eingegeben und anschliessend eine Suche gestartet. Im Hintergrund greift die Primo-Suchmaschine auf den Index zu und gibt die entsprechenden Resultate als Kurztitelliste wieder. Wenn man die Suchresultate durch die Facetten weiter einschränkt, dann setzt dies eine neue Suche ab und greift wiederum auf den Primo-Index zu. Man muss sich in der Praxis insbesondere bewusst sein, dass beim Bearbeiten von Normalisierungsregeln für Facetten oder auch für die Suche, die Änderungen erst nach einem neuen Indexieren wirksam sind.

Bei einem Klick auf die Details gelangt man zur Vollanzeige, bei dem im Hintergrund die Daten aus der Primo-Datenbank abgefragt werden.

Die Verfügbarkeitsanzeige wird zum einen aus den gespeicherten Daten gezogen, aber anschließend mit den aktuellen Informationen aus Aleph abgeglichen. Dies sieht man daran, dass kurz die Meldung „wird geprüft“ erscheint. Dieses Vorgehen wird unter dem Namen Real-Time-Availability zusammengefasst. Es ist ebenfalls möglich nur nach momentan verfügbaren Titeln zu suchen. Aber dabei findet keine aktuelle Verfügbarkeitsabfrage mit Aleph statt. Deshalb kann es durchaus sein, dass bei einer Suche eingeschränkt auf verfügbare Medien auch Titel erscheinen, die nicht verfügbar sind, oder verfügbare Titel nicht gefunden werden. Der Datenabgleich von Aleph mit Primo läuft täglich, so dass diese Titel eher die Ausnahme sind. Ebenfalls erscheinen neu katalogisierte Medien (z.B. Bestellkatalogisate) somit erst am darauffolgenden Tag in Primo.

Die Standortangaben werden aus Aleph extrahiert und in Primo zur Anzeige gebracht. Ebenfalls das Benutzerkonto und die Bestelloptionen greifen auf Aleph zu.

2 Einheitliche Einschlitz-Suche

Bekannte und viel genutzte Webseiten wie etwa Google, Amazon, Wikipedia, Youtube stellen eine einheitliche und einfache Suche zur Verfügung. Man kann beliebige Begriffe in ein großes Suchfeld eingeben und die Suche starten. Falls man die gewünschten Informationen nicht gleich findet, kann man durch Hinzufügen neuer Suchbegriffe oder durch Einschränken auf bestimmte Kategorien, die Anfrage verfeinern. Diese Suchmöglichkeiten sind in ihren Anwendungen etabliert und auch den heutigen Bibliotheksnutzern bekannt.

Die Benutzer von Bibliothekskatalogen sind häufig überfordert von deren komplexen Suchmöglichkeiten und nutzen nur die einfachsten Funktionen. Umso komplexer die Suchmöglichkeiten sind, umso schwieriger ist es, diese als Bibliothek zu vermitteln. Erschwerend kommt hinzu, dass unterschiedliche Bibliotheken ganz unterschiedliche Web-OPACs besitzen. Ein Wechsel zu einer anderen Universität kann somit für Studierende auch bedeuten, einen neuen Bibliothekskatalog zu lernen.

Dies sind Gründe, um eine einfache und intuitive Suche zu präsentieren. Unser Sucheinstieg in Primo besteht aus einem einzigen Suchschlitz, in dem alle Suchfelder in beliebiger Kombination durchsucht werden können. Dazu hört man heutzutage häufig die Begriffe „Einschlitz-Suche“ oder auch im allgemeineren Zusammenhang „Google-like-Suche“. Wir verzichten dabei ganz bewusst auf weitere Auswahlmöglichkeiten.



The screenshot shows the top navigation bar of the UB Trier website. It includes the logo 'UB Trier' on the left and links for 'Gast', 'Mein Bereich', and 'Anmelden' on the right. Below the navigation bar, there are links for 'Neue Suche', 'E-Zeitschriften', 'Hilfe', and 'Sprache: Deutsch'. The main search area features a search bar with a 'Suche' button and a link for 'Erweiterte Suche'.

Neben solch einer einfachen Suche haben wir auch eine erweiterte Suche, bei der zum Beispiel spezifiziert werden kann, in welchen Suchfeldern gesucht werden soll. Damit kann man zum Beispiel Literatur von Goethe und Literatur über Goethe unterscheiden.



The screenshot shows the advanced search interface of the UB Trier website. It features a search bar with a 'Suche' button and a link for 'Einfache Suche'. The search criteria are organized into two columns. The left column contains three rows, each with a dropdown menu for the search field (e.g., 'Titel', 'Autor', 'Inhaltsverzeichnis'), a dropdown for the relationship (e.g., 'enthält'), and a text input field (e.g., 'lehrbuch'). The right column contains three rows, each with a label (e.g., 'Erscheinungsjahr:', 'Materialart:', 'Sprache:') and a dropdown menu (e.g., 'irgendein Jahr', 'Alle', 'Alle Sprachen').

Es gibt komplexere Abfragen, die auch die erweiterte Suche nicht anbietet. Das Blättern im Index, wie in Aleph-OPAC gewohnt, ist nicht mehr möglich in Primo. Dies ist aber zum Beispiel bei Autoren mit verschiedenen Schreibweisen nützlich. Ebenfalls ist es nicht (oder nicht ohne Weiteres) möglich, überall außer im Inhaltsverzeichnis zu suchen.

Mit Primo Central haben wir die Möglichkeit Aufsatzdaten zu recherchieren. Hier stellt sich die Frage, ob man diese Aufsatzdaten gemeinsam mit den Katalogdaten durchsuchbar macht oder ob man zwei getrennte Sucheinstiege anbietet. Wir haben uns für die zweite Variante entschieden und die beiden Sucheinstiege in unterschiedliche Tabs gesetzt: Katalog, Aufsatzdatenbank. Ebenfalls haben wir noch einen Tab Semesterapparate, auf den ich hier nicht näher eingehen möchte.

Eigentlich wäre eine gemeinsame Suche in Katalogdaten und Aufsatzdaten wünschenswert. Das Problem ist die Heterogenität der Daten. Die Katalogdaten sind qualitativ hochwertige Daten, die unter dem Regelwerk RAK-WB erstellt wurden und intellektuell bearbeitet und überprüft wurden. Die Aufsatzdaten in Primo Central sind von Verlagen und Aggregatoren mit mehr oder weniger Aufwand erstellt worden und werden durch Ex Libris automatisch eingespielt und angepasst. Dabei ist die Qualität der Metadaten sehr unterschiedlich. Insbesondere fehlt häufig die sachliche Erschließung oder diese ist nicht einheitlich.

Hier ein Beispiel für unsere Katalogdaten:

Mathematics of Ramsey theory
Nešetřil, Jaroslav [Hrsg.] 1990
Signatur(en): 55=HV.MATH/c9894
Verfügbar

Details Standorte Weitere Services Senden an ▾

Titel:	Mathematics of Ramsey theory
Sonstige Beteiligte:	Nešetřil, Jaroslav [Hrsg.]
Schlagwörter:	Ramsey-Theorie
Verknüpfte Titel:	Algorithms and combinatorics ; 5
Verlag:	Berlin [u.a.] : Springer
Erscheinungsjahr:	1990
Format:	XIV, 269 S. : graph. Darst.
Suche Reihe:	HT002876035 Algorithms and combinatorics
Identifikator:	ISBN 3-540-18191-1 : (Berlin ...); ISBN 0-387-18191-1 : (New York ...)
Notation:	55=HV Graphentheorie

Links

- > [Dieser Titel im Bibliothekskatalog](#)
- > [Inhaltsverzeichnis](#)
- > [Lesesaal E: 55= HV.MATH/c9894](#)

Als sachliche Erschließung dient hier ein Schlagwort, das aus der SWD (Schlagwort-Norm-Datei) stammt. Damit ist ein einheitliches Vokabular sichergestellt und auch Verweisungsformen sind suchbar. Die Person ist ebenfalls mit dem entsprechenden Normdatensatz verknüpft und somit sind auch andere Schreibweisen suchbar. Konkret sind für die Person im Beispiel oben folgende drei Ansetzungs- und Verweisungsformen im Normdatensatz gespeichert: Nešetřil, Jaroslav ; Nesetril, J. ; Nešetřil, Jarik. Eine Suche in unserem Bibliothekskatalog nach einer dieser Formen im Autorenfeld liefert immer genau 6 Treffer. Eine Suche in Primo Central nach diesen Formen liefert 21, 64 bzw 1 Treffer.

Folgender Datensatz aus Primo Central soll vergleichsweise betrachtet werden:

Banach spaces and Ramsey Theory: some open problems; Espacios de Banach y Teoría de Ramsey: Ciertos problemas abiertos
Dodos, P. ; Lopez-Abad, J. ; Todorcevic, S.
Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas, 2010, Vol.104(2), p.435-450
Volltext verfügbar

Online Ressource Details Weitere Services

Senden an ▾

Titel:	Banach spaces and Ramsey Theory: some open problems; Espacios de Banach y Teoría de Ramsey: Ciertos problemas abiertos
Autor:	Dodos, P. ; Lopez-Abad, J. ; Todorcevic, S.
Schlagwörter:	Banach spaces ; Ramsey Theory ; open problems ; Mathematics Subject Classifications ; 03E02 ; 03E55 ; 05D10 ; 46B03 ; 46B07 ; 46B15 ; 46B26 ; 46B80 ; 47B38
Enthalten in:	Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas, 2010, Vol.104(2), p.435-450
Beschreibung:	We discuss some open problems in the Geometry of Banach spaces having Ramsey-theoretic flavor. The problems are exposed together with well known results related to them. Presentamos ciertos problemas en la geometría de los espacios de Banach que tienen conexiones con la teoría de Ramsey. Los problemas son expuestos juntamente con resultados bien conocidos relacionados con ellos.
Sprache:	Englisch ; Spanisch
Identifikator:	ISSN: 1578-7303 ; E-ISSN: 1579-1505 ; DOI: 10.5052/RACSAM.2010.27

Es fällt auf, dass die Vornamen in Primo Central nicht immer aufgelöst werden. Ebenfalls sind die Schlagwörter hier aus dem Englischen und aus der MSC-Klassifikation zusammengewürfelt. Aufgelöste Systematikbezeichnungen gibt es nicht und somit sind diese auch nicht recherchierbar. Zum Beispiel würde 05D10 bzw. auch „Ramsey Theory“ dem RSWK-Schlagwort „Ramsey-Theorie“ entsprechen, aber eine solche Zusammenführung/Vereinheitlichung des Vokabulars gibt es nicht.

Diese Heterogenität der verschiedenen Datenquellen erschwert eine einheitliche Suche. Wenn man die Daten von Primo Central mit den Katalogdaten mischt und gemeinsam durchsuchbar machen würde, dann hat dies insbesondere auch Auswirkungen auf die Möglichkeiten die Suchresultate weiter einzuschränken. Eine Einschränkung nach dem Schlagwort „Ramsey-Theorie“ würde mit einem Klick – und für den Nutzer unbekannt bzw. unbewusst – alle Primo Central-Treffer ausschließen. Ebenfalls machen Facetten wie lokale Systematik und Standort nur für Bibliotheksdaten Sinn. Aus diesen Gründen haben wir uns dafür entschieden, die Katalogdaten und Aufsatzdaten so einheitlich wie möglich aber in zwei verschiedenen Tabs durchsuchbar zu machen. Auf eine gemeinsame Recherche in beiden Datenpools wird dabei verzichtet.

3 Facettierungen

Facetten ermöglichen eine Suche weiter einzuschränken. Insbesondere stoßen Benutzer dabei auf weitere Such- bzw. Filtermöglichkeiten, an die man im Vorherein nicht gedacht hat. Würde man bei einer Suche das Stichwort „Deutsch“ eingeben, um nur Medien auf Deutsch zu finden? Wahrscheinlicher ist, dass man eine Suche nach einigen aussagekräftigen Stichwörtern absetzt und dann die Suchresultate mittels der Sprach-Facette weiter einschränkt. Ob man eine Facette benutzt, ist meist auch abhängig von der Anzahl der Treffer. Beispielsweise macht eine Facettierung bei 3 Treffern wenig Sinn. Andererseits ist eine Facettierung bei einigen Tausend Treffern fast schon unumgänglich.

Verschiedene Facettenwerte der gleichen Facette müssen sich keineswegs ausschließen. Eine Aufsatzsammlung kann sowohl Beiträge auf Deutsch wie auch auf Englisch enthalten. In diesem Fall würde die Aufsatzsammlung auch beide Facettenwerte besitzen und bei Suchen nach englischen wie auch nach deutschen Dokumenten erscheinen.

Die Anzahl der Facettenwerte, die angezeigt werden, ist standardmäßig auf maximal 25 beschränkt. Hier ist ein Beispiel für die Medientyp-Facette, bei der die ersten 5 Werte direkt sichtbar sind und die restlichen 5 Werte durch einen Klick sichtbar werden:



Die Information über den Medientyp wird aus dem kodierten MAB-Feldern 050, 051, 052 gewonnen. Im obigen Beispiel wurde nach dem Stichwort „Management“ im Bibliothekskatalog gesucht und insgesamt 27'035 Treffer erzielt (die exakten Zahlen variieren durch Bestandsveränderungen). Die Facettenwerte mit den meisten Treffern werden als erste aufgeführt und man sieht, dass die gesamte Treffermenge (bis auf einige wenige Ausnahmen) vollständig durch diese Facettenwerte abgedeckt werden.

Etwas anders sieht dies bei Facetten zu Schlagworten oder Autoren aus. Mit nur 25 potentiellen Autoren als Facettenwerten wird man nicht alle 27'035 Treffer abdecken. Somit gibt es sicherlich Autoren, die bei der Facette nicht auftreten. Dasselbe gilt für die Schlagworte. Es gibt keine Möglichkeit die restlichen Facettenwerte (zum Beispiel durch eine Blätter-Funktion) anzuzeigen. Somit hat man hier keine Vollständigkeit im eigentlichen Sinne. Wir sind der Frage nachgegangen, welche 25 Facettenwerte angezeigt werden. Man könnte vermuten, dass dies die Facettenwerte mit den meisten Treffern sind. Dies stimmt aber nur bedingt. Die Anzeige einer Facette kann man in zwei Schritte unterteilen: Bei einer Suche wertet Primo die ersten 400 Treffer aus und bestimmt daraus die Top-Facettenwerte und in einem zweiten Schritt berechnet Primo die Anzahl der Treffer

dieser Facettenwerte in der gesamten Treffermenge.

Welches die ersten 400 Treffer sind hängt wiederum von der Sortierung ab, z.B. Relevanz-Sortierung. Es ist möglich die Einstellung 400 zu erhöhen, wobei sich dies sehr schnell auf die Performanz niederschlägt (und 400 ist schon ein erhöhter Wert, standardmäßig ist 200 in Primo eingestellt). Ein Facettenwert, der nicht in den ersten 400 Treffern vorkommt, wird nicht angezeigt unabhängig davon wie oft er insgesamt vorkommt!



Als Beispiel für die Abhängigkeit von der Sortierung sieht man hier die Facettenwerte für die Alle-Felder-Suche nach „Management“ bei verschiedenen Sortierungen:

Sortierung nach Relevanz

Schlagwort
 Management (2.949)
 Deutschland (1.354)
 Unternehmen (1.096)
 Strategisches Management (1.009)
 USA (521)
 Risikomanagement (441)
 Innovationsmanagement (328)
 Projektmanagement (323)
 Verwaltungsreform (312)
 Controlling (283)
 Internationales Management (172)
 Kundenmanagement (162)
 Schweiz (156)
 Wettbewerbsstrategie (153)
 Unternehmensentwicklung (131)
 Interkulturelles Management (122)
 Netzwerkverwaltung (105)
 Lean Management (88)
 Supply Chain Management (61)
 Management Buyout (49)
 Diversity Management (31)
 Facility-Management (25)
 Case Management (24)
 Industrieforschung (22)
 Management Auditing (16)
 Ausblenden der letzten 20 ▲

Sortierung nach Jahr

Schlagwort
 Management (2.949)
 Deutschland (1.354)
 Unternehmen (1.096)
 Strategisches Management (1.009)
 Datenbanksystem (569)
 Qualitätsmanagement (535)
 Klein- und Mittelbetrieb (437)
 Innovationsmanagement (328)
 Verwaltungsreform (312)
 Multinationales Unternehmen (306)
 Europäische Union (242)
 Wissensmanagement (231)
 Internationales Management (172)
 Europa (159)
 Schweiz (156)
 Krankenhaus (150)
 Kommunikation (106)
 Evaluation (91)
 Nachhaltigkeit (89)
 Wirtschaftlichkeit (27)
 Unterricht (26)
 Neue Medien (24)
 Change Management (23)
 Lehrer (23)
 Ratgeber (11)
 Ausblenden der letzten 20 ▲

Sortierung nach Bekanntheit

Schlagwort
 Management (2.949)
 Deutschland (1.354)
 Unternehmen (1.096)
 Strategisches Management (1.009)
 USA (521)
 Risikomanagement (441)
 Projektmanagement (323)
 Verwaltungsreform (312)
 Personalpolitik (303)
 Controlling (283)
 Wissensmanagement (231)
 Unternehmensplanung (182)
 Internationales Management (172)
 Schweiz (156)
 Wettbewerbsstrategie (153)
 Unternehmensentwicklung (131)
 Interkulturelles Management (122)
 Netzwerkverwaltung (105)
 Verteiltes System (93)
 Lean Management (88)
 Supply Chain Management (61)
 Management Buyout (49)
 Facility-Management (25)
 Case Management (24)
 Vielfalt (20)
 Ausblenden der letzten 20 ▲

Der Top-Wert ist jeweils das Schlagwort „Management“ mit insgesamt 2949 Treffern gefolgt von „Deutschland“, „Unternehmen“ und „Strategisches Management“. Aber bereits dann gibt es Unterschiede: Bei einer Sortierung nach Jahr sieht man das Schlagwort „Datenbanksystem“, welches bei den anderen beiden Sortierungen unterschlagen wird. Scheinbar sind alle diese Dokumente bei der Sortierung nach Relevanz und Bekanntheit an Positionen nach 400 und darum erscheinen die in deren Auswahl nicht.

Um dieses Phänomen noch etwas eingehender zu untersuchen, habe ich versucht die Facettenwerte mit den meisten Treffern (unabhängig von jeglicher Sortier-Reihenfolge) zu ermitteln. Nachfolgend sieht man eine Tabelle der gefundenen Top-25-Facettenwerte, bei der die in der Relevanz-Sortierung enthaltenen grün markiert sind und die nicht enthaltenen rot und zusätzlich kursiv:

1	Management	2949
2	Deutschland	1354
3	Unternehmen	1096
4	Strategisches Management	1009
5	<i>Datenbanksystem</i>	569
6	<i>Qualitätsmanagement</i>	535
7	USA	521
8	<i>Operations Research</i>	502
9	<i>Marketing</i>	470
10	Risikomanagement	441
11	<i>Klein- und Mittelbetriebe</i>	437
12	<i>Organisation</i>	385
13	<i>Umweltbezogenes Management</i>	348
14	Innovationsmanagement	328
15	Projektmanagement	323
16	<i>Führung</i>	321
17	Verwaltungsreform	312
18	<i>Multinationales Unternehmen</i>	306
19	<i>Personalpolitik</i>	303
20	<i>Informationsmanagement</i>	297
21	<i>Großbritannien</i>	291
22	Controlling	283
23	<i>Personalentwicklung</i>	278
24	<i>Führungskraft</i>	259
25	<i>Zeitschrift</i>	255

Die ersten vier Treffer wurden alle auch von Primo gefunden wurden, aber die nächsten zwei nicht. Um es nochmals deutlich zu machen: Das Schlagwort „Datenbanksystem“ hat zwar 569 Treffer aber keiner davon ist unter den ersten 400 bei einer Sortierung nach Relevanz und somit wird dies bei der Schlagwortfacette nicht angeboten. 10 der 25 Facettenwerte, die Primo bei einer Sortierung nach Relevanz anzeigt, sind auch hier unter den ersten 25. Es gibt 4 weitere Facettenwerte unter den ersten 50 und dazu 6 weitere unter den ersten 100. Schließlich haben 5 der 25 Facettenwerte, die Primo anzeigt, einen Rang über 100. Insgesamt verteilen sich die von Primo gefundenen 25 Werte wie folgt:

Rang	1 – 25	26 – 50	51 – 75	76 – 100	101 –
Anzahl	10	4	3	3	5

Ein anderes Beispiel können wir mit der Autoren-Facette machen. Nehmen wir einmal an, ein Benutzer sucht nach „Gedichten“ in allen Feldern. Bei der FU Berlin in Primo findet man 8919 Treffer (wir lassen diese Facette in Trier nicht mehr anzeigen). In der Autoren-Facette würde man dann anschließend vermuten, dass bekannte Autoren von Gedichten bzw. solche die viele Gedichtsbände geschrieben haben erscheinen. Im Folgenden sieht man die Facettenwerte nach einer Suche nach „Gedichten“:

- Autor**
- Dedecius, K (25)
 - Tieck, L (23)
 - Meister, E (22)
 - Matthisson, F (12)
 - Huchel, P (10)
 - Geibel, E (8)
 - Handke, P (7)
 - Nietzsche, F (7)
 - Stolberg-Stolberg, C (6)
 - König, J (3)
 - Herbert, Z (3)
 - Eminescu, M (3)
 - Bertram, E (3)
 - Nick, D (3)
 - Fauser, J (3)
 - Holtei, K (3)
 - Koller, B (2)
 - Kaufmann, A (2)
 - Prutz, R (1)
 - Ludwig <Bayern, König I.> (1)
 - Follain, J (1)
 - Röding, J (1)
 - Bach, J (1)
 - Jean <de Condé> (1)
 - Conz, K (1)
 - Ausblenden der letzten 20 ▲

Bekannte Autoren von Gedichten wie etwa Goethe, Schiller oder Hölderlin werden nicht angezeigt, da sie in der Sortierung zu weit hinten kommen. Bei einer kombinierten Suche nach „Gedichten“ und dem jeweiligen Autor kann man die exakten Facettenwert-Größe ablesen:

Goethe	98
Hölderlin	44
Schiller	42

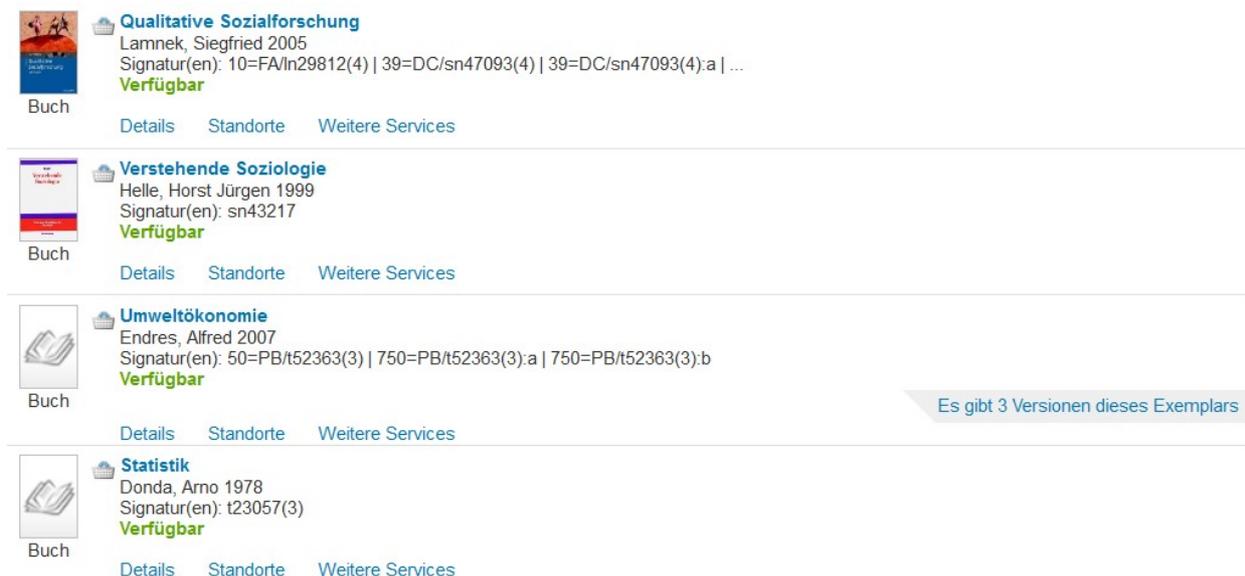
Somit beinhalten die angezeigten Facettenwerte von Prim nicht einmal eine der drei Top-Facettenwerte, sondern diese gehen aufgrund der Sortierung verloren.

Zusammenfassend kann man sagen, dass die Facetten die Suche erleichtern können und insbesondere neue Such- und Filtermöglichkeiten in Abhängigkeit der Treffermenge anbieten können. Bei Facetten mit nur wenigen Werten gibt es wohl kaum Probleme. Zusätzlich kann man in diesem Fall die Facette auch statisch machen, d.h. alle möglichen Facettenwerte werden im Vornherein gespeichert und Schritt 1 bei der Erzeugung der Facette entfällt. Für diese Liste aller möglichen Facettenwerte werden dann die entsprechenden Zahlen in der Treffermenge berechnet. Somit hat man bei statischen Facetten die richtigen Top-25-Facettenwerte angezeigt. Bei großen Facetten erhält man aber keine Vollständigkeit über die Facettenwerten und wichtige Facettenwerte können durch die Sortierung verloren gehen, womit die entsprechenden Titel nicht über diese Facette auffindbar sind.

4 Darstellung lokaler Besonderheiten am Beispiel von Signaturen

4.1 Anzeige von Signaturen

Die Darstellung der Daten in Primo kann auf vielfältige Weise beeinflusst und individualisiert werden. Alle lokalen Daten muss man selbst zur Anzeige in Primo bringen bzw. suchbar machen. Hier möchte ich erläutern, wie wir die Signaturen in Primo zur Anzeige bringen und suchbar machen. Zum einem ist die Signatur ein eindeutiger Identifikator eines Mediums und es muss möglich sein, nach Signaturen zu suchen. Zum anderen enthält die Signatur eines Exemplars bei uns auch Standortinformationen. Um die gesuchte Literatur in der Bibliothek zu finden, kann es nützlich sein, sich eine Kurztitelliste mit Signaturen auszudrucken. Dazu müssen die Signaturen in Primo in der Kurztitelliste angezeigt werden. Also verfolgten wir zwei Ziele: die Signaturen suchbar zu machen und in der Kurztitelliste zur Anzeige zu bringen. Dabei gilt es zu beachten, dass einige Bücher in Mehrfachexemplaren vorhanden sind und somit auch mehrere Signaturen besitzen. Ein Buch kann an verschiedenen Standorten vorhanden sein oder aber auch in einer großen Anzahl zum Beispiel in der Lehrbuchsammlung vorhanden sein. Daher haben wir uns entschieden, bis zu drei verschiedene Signaturen zur Anzeige zu bringen (aber natürlich alle suchbar zu machen). Nachfolgend ein Beispiel für die Anzeige von Signaturen:



The screenshot displays four book entries in a Primo library catalog. Each entry includes a book cover icon, the title, author, year, and a list of signatures. The status 'Verfügbar' (Available) is shown in green. Navigation links for 'Details', 'Standorte' (Locations), and 'Weitere Services' (Further Services) are provided for each entry. A callout box for the third entry indicates 'Es gibt 3 Versionen dieses Exemplars' (There are 3 versions of this exemplar).

Title	Author	Year	Signature(s)	Status
Qualitative Sozialforschung	Lamnek, Siegfried	2005	10=FA/In29812(4) 39=DC/sn47093(4) 39=DC/sn47093(4):a ...	Verfügbar
Verstehende Soziologie	Helle, Horst Jürgen	1999	sn43217	Verfügbar
Umweltökonomie	Endres, Alfred	2007	50=PB/t52363(3) 750=PB/t52363(3):a 750=PB/t52363(3):b	Verfügbar
Statistik	Donda, Arno	1978	t23057(3)	Verfügbar

Das erste Buch steht bei Pädagogik (10=...) und bei Soziologie (39=...) und weitere Signaturen sind mit den drei Punkten angedeutet. Das zweite Buch sn43217 steht im Magazin. Mit Hilfe des Leitsystem (Hinweistafeln, Beschilderung der Regale) in der Bibliothek oder der Standortvisualisierung BibMap finden die Benutzer die Bücher am Standort leicht auf.

4.2 Implementierung Signaturanzeige

In einem ersten Schritt muss man sicherstellen, dass die entsprechenden Daten (d.h. Signaturen der Exemplare) in Primo vorhanden sind. Die Informationen müssen beim Publishing mitgeliefert werden. Das Publishing wird mit der `tab_ger` konfiguriert, welche wiederum einzelne Routinen ansteuert. Die Informationen aus den einzelnen Exemplardatensätzen werden mit einer Routine anhand der Konfigurationsdatei `expand_doc_bib_z30_primo` herausgeholt. Darin mussten wir die folgenden Zeilen hinzufügen:

#	z30-collection	2	space	N
#	z30-call-no	3	space	X

Damit werden für jeden Datensatz die entsprechenden Exemplardaten geöffnet und unter anderem die einzelnen Signaturen herausgeschrieben. Die Primo source records haben dann Einträge der Art:

```
<datafield tag="LOC" ind1=" " ind2=" " >
  <subfield code="2">10</subfield>
  <subfield code="3">FA/ln29812(4)</subfield>
  <subfield code="9">ln</subfield>
  <subfield code="q">20050905</subfield>
  <subfield code="f">49</subfield>
</datafield>
```

Man sieht, dass die Signatur sich aus den Unterfeldern 2 und 3 des Feldes LOC zusammensetzen lässt. Zuerst nehmen wir alle diese Informationen für die Suche und speichern diese in ein eigenes Suchfeld `Isr10` im `PNX`. Anhand der Normalisierungsregeln kann man steuern, welche Informationen von Primo source record wie und wann verwendet werden sollen. Hier ist die Normalisierungsregel die dies macht:

The screenshot shows the configuration for a rule group named "search_Isr10".

- Rule group:** search_Isr10
- Source:** MAB (dropdown), Field: LOC, Ind1: [empty], Ind2: [empty], Last: [empty], Step: 0, Subfield: Include (dropdown), 2,3 (input), [checked] Enabled.
- Conditions logic:** Conditions: False (dropdown).
- Transformations:** Transformation: Define subfield delimiter (dropdown), Parameter: [=] (input). Behavior: MERGE (dropdown), First delimiter: | (input), First delimiter Space: Both (dropdown), Repeat Number: [empty], Remaining delimiters: | (input), Remaining delimiters spaces: Both (dropdown).

Das Wesentliche ist hier die Zeile „source“, also woher wird die Information genommen. Wir möchten die Informationen vom Feld LOC die Unterfelder 2 und 3 nehmen. Die Transformationen steuern die Darstellung der Daten.

Das Resultat des obigen Beispiels sieht dann wie folgt aus:

```
<lsrc10>
  10=FA/ln29812(4) | 39=DC/sn47093(4) | 39=DC/sn47093(4):a |
  710=FA/ln29812(4):a | 710=FA/ln29812(4):b | 710=FA/ln29812(4):c |
  710=FA/ln29812(4):d | ln29812(4):e
</lsrc10>
```

Die senkrechten Striche sind hier nicht wichtig (sie sind nur für das weitere Vorgehen mit lsrc10 praktisch). Für die Suche nach Signaturen in Primo muss man die Normalisierungsregeln ändern und dann diese Normalisierungsregeln wieder auf alle Datensätze anwenden, d.h. eine Update-Pipe muss gestartet werden. Anschließend sind diese neuen Informationen gespeichert, aber noch nicht indexiert. Darum muss man das Ganze mit einer Neuindexierung und Hotswapping abschließen. Jede der acht Signaturen ist für sich einzeln suchbar, wobei sich eine exakte Suche anbietet.

Um die Signaturen übersichtlich anzuzeigen, speichern wir den entsprechenden Teil in der display-Sektion des PNX in ein eigenes Feld lsrc10. Die größte Schwierigkeit bestand darin, nur höchstens drei Signaturen anzeigen zu lassen. Da wir in Primo keine direkte Möglichkeit zum zählen oder abrechnen nach dreimal gesehen haben, sind wir auf reguläre Ausdrücke zurückgegangen.

The screenshot shows the configuration interface for a rule group named 'display_lsrc10'. The interface is divided into two sections, labeled 1 and 2.

Section 1: This section shows the configuration for the rule group. The 'Rule group' is 'display_lsrc10'. The 'Source' is 'PNX' and the 'Value' is 'search/lsrc10'. The 'Conditions logic' is 'True'. The 'Condition 1 - Logic' is 'True'. The 'Condition 1 - Source' is 'PNX' and the 'Value' is 'search/lsrc10'. The 'Match Any' dropdown is set to 'Match Any'. The 'Condition 1 - Routines' section shows a 'General Parameter' field, a 'Routine' dropdown set to 'Validate (use reg. exp.)', and a 'Parameter' field containing the regular expression '.*\.\.\.\.\.*'. The 'Action' is 'ADD'.

Section 2: This section shows the configuration for the display section. The 'Rule group' is 'display_lsrc10'. The 'Source' is 'PNX' and the 'Value' is 'search/lsrc10'. The 'Conditions' are 'No conditions Specified'. The 'Transformations' section shows three transformations: 'Take string (use reg. exp.)' with parameter '^([\.\.\.\.\.])', 'Add to end of string' with parameter '^...', and 'Add to beginning of string' with parameter 'Signatur(en):^'. The 'Behavior' is 'OR'.

Als Quelle verwenden wir hier das eben beschriebene Feld Isr10. Zwei Normalisierungsregeln steuern das Feld Ids10. Die erste Regel kommt zum Zug, falls es mehr als drei Signaturen gibt, was in der Bedingung (condition) geprüft wird. Falls dies der Fall ist, dann wird nach der dritten Signatur „abgeschnitten“ und drei Punkte hinzugefügt. Dazu verwenden wir den regulären Ausdruck

```
^[^\|]*\[^\|]*\[^\|]*\|
```

den man etwa wie folgt lesen muss:

^ gehe zum Beginn der Zeile
(der Ausdruck ab hier wird zurückgeliefert
[^\|]* irgendein Zeichen nur nicht |, beliebig oft hintereinander (1. Signatur)
\| das Zeichen |
[^\|]* irgendein Zeichen nur nicht |, beliebig oft hintereinander (2. Signatur)
\| das Zeichen |
[^\|]* irgendein Zeichen nur nicht |, beliebig oft hintereinander (3. Signatur)
) der Ausdruck bis hier wird zurückgeliefert
\| das Zeichen | (d.h. es folgen noch weitere Signaturen)

Die zweite Normalisierungsregel kommt in allen anderen Fällen zum Zug. Wichtig ist hier die „action: or“, was bedeutet, dass die zweite Normalisierungsregel nicht ausgeführt wird, falls schon davor eine andere ausgeführt wurde.

Nachdem diese Normalisierungsregeln durch eine Update-Pipe auf alle Daten angewandt wurde, sieht unser Beispiel wie folgt aus:

```
<Ids10>  
Signatur(en): 10=FA/ln29812(4) | 39=DC/sn47093(4) | 39=DC/sn47093(4):a | ...  
</Ids10>
```

Dies wird durch eine Einstellung in den Views dann auch entsprechend zur Anzeige in der Kurztitelliste gebracht.

4.3 Zusammenfassung

Zusammenfassend kann man sagen, dass die Möglichkeiten von Primo sehr vielfältig sind aber auch nicht immer ganz einfach. Besonders erschwerend kommt hinzu, dass man um eine Information hinzuzufügen, an vielen verschiedenen Stellen etwas ändern muss. Die Update-Pipe und das Indexieren kann einige Stunden dauern und somit sind solche Änderungen zumindest im laufenden System gut zu planen und müssen eventuell mit anderen Kooperationspartnern abgesprochen werden.

Literaturverzeichnis

Diplom- und Studienarbeiten:

Möglichkeiten und Grenzen von Primo bei der Einführung in deutschsprachigen Bibliotheken und Bibliotheksverbänden / Raicher, Elisabeth M. – Diplomarbeit, FH Eisenstadt, 2010.

<http://hdl.handle.net/10760/14865>

Vom Schlagwort zur Facette : Sacherschliessungsinformationen in Primo am Beispiel von „u:search“ der UB Wien / Boos, Nadine ; Donven, Michel ; Scholz, Stephanie ; Voß, Viola – unveröffentlichte Studienarbeit, FH Köln, 2010

Mit Primo besser finden? / Boos, Nadine ; Donven, Michel ; Scholz, Stephanie ; Voß, Viola – unveröffentlichte Studienarbeit, FH Köln, 2010

Artikel:

Prima Primo?! / Raicher, Elisabeth M. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare*. 2010. S. 85 - 89.

<http://hdl.handle.net/10760/15803>

Sachliche Erschließung und die neue Suchmaschine des Österreichischen Bibliothekenverbundes / Knitel, Markus ; Labner, Josef ; Oberhauser, Otto In: *B.I.T.online*, 13 (2010) Nr. 3, S. 275 - 285

http://www.obvsg.at/fileadmin/files/obvsg/publ/Knitel_et_al_2010_Sachliche.pdf

Architektur und Parametrisierung von Primo im österreichischen Bibliotheksverbund : Ein Überblick / Knitel, Markus In: *Mitteilungen Österreichischer Bibliothekarinnen und Bibliothekare*, 63 (2010) 1/2, S. 56 - 68

<http://fiz1.fh-potsdam.de/volltext/voeb/10491.pdf>

Primo an der Universitätsbibliothek Wien / Putz, Michaela In: *The ne(x)t generation : Das Angebot der Bibliotheken ; 30. Österreichischer Bibliothekartag*, Graz, 2009, S. 227 - 232

Präsentationen:

Ist Primo prima? Das neue Discovery-System der SLUB Dresden / Bonte, Achim

http://www.slideshare.net/Achim_Bonte/ist-primo-prima-das-neue-discoverysystem-der-slub-dresden

Vom Opac zur Suchmaschine : Mehr lokale Funktionen durch Nutzung konsortialer Synergien / Babitchev, Viktor ; Brandauer, Johann

<http://www.opus-bayern.de/bib-info/volltexte//2011/1085/>

Abbildung von E-Books in Primo – kurz & gut / Hänger, Christian

<http://www.opus-bayern.de/bib-info/volltexte/2011/1098/>

Management integrierter, konsortialer Portallösungen / Lohrum, Stefan

<http://www.opus-bayern.de/bib-info/volltexte//2011/1146/>

Primo Mobile: UB der TU Berlin - mobil mit Primo / Kuberek, Monika ; Dietsch, Matthias

<http://www.opus-bayern.de/bib-info/volltexte/2011/1103/>

Optimiertes Discovery und Delivery Beobachtungen zum Nutzungsverhalten und zur Nutzung von E-Ressourcen nach Einführung von Primo an der FU Berlin / Kowalak, Mario ; Sabisch, Andreas

<http://www.opus-bayern.de/bib-info/volltexte//2011/1157/>

Umdenken mit Primo / Kowalak, Mario

http://www.kobv.de/fileadmin/download/kobv_forum/08_kowalak_kobv_forum_2010.pdf

OPAC Plus der Universitätsbibliothek Mannheim und konsortiale Nutzung von Primo im KOBV / Hänger, Christian ; Lohrum, Stefan

<http://www.opus-bayern.de/bib-info/volltexte/2010/899/>

Katalog 2.0 : Lösungsfindung an sächsischen Hochschulbibliotheken / Kluge, Andreas

<http://www.opus-bayern.de/bib-info/volltexte/2010/869/>

Verbundintegrationsplattform Primo : Synergien in einem homogenen Verbund / Hamedinger, Wolfgang

<http://www.opus-bayern.de/bib-info/volltexte/2010/932/>

Links

FU Berlin

<http://primo.fu-berlin.de/>

TU Berlin

<http://portal.ub.tu-berlin.de/>

HU Berlin

<http://primus.ub.hu-berlin.de/>

UB Mannheim

http://vs30.kobv.de/primo_library/libweb/action/search.do?&vid=MAN_UB

SLUB Dresden

<http://www.slub-dresden.de/recherche/>

ULB Münster

<http://www.ulb.uni-muenster.de/ULB/disco/>

Uni Wien

<http://usearch.univie.ac.at/>

UB Innsbruck

<http://search.obvsg.at/UIB>

ETH Zürich

<http://www.library.ethz.ch/>

New York University

<http://www.bobcat.nyu.edu/>

University of Iowa

<http://www.lib.uiowa.edu/>