Tag-Cluster – Semantische Suche in Tag-Clouds

Kathrin Knautz

Auch in Bibliotheken sind mit Konzeptionen wie dem Katalog 2.0 Funktionalitäten von kollaborativen Diensten angekommen, so dass neue Retrievalformen in Online-Katalogen möglich werden. Thema dieser Studie ist die Entwicklung und Implementierung einer alternativen Retrieval-Schnittstelle innerhalb von Web2.0-Diensten in Form von Tag-Clustern. Durch die Berechnung der Ähnlichkeiten von Tags auf der Basis ihres gemeinsamen Auftretens bei der Indexierung von Ressourcen und anschließendem Clustering ist es möglich, den Defiziten aktueller Visualisierungsmethoden von Folksonomies entgegenzuwirken und ein effizienteres Retrieval zu ermöglichen.

Tag Cluster – Semantic search in Tag Clouds

By the use of concepts like the catalogue 2.0, features of collaborative services have also reached libraries so that new kinds of retrieval become possible in online catalogues. This study deals with the development and implementation of an alternative retrieval interface within web 2.0 services in terms of tag clusters. The calculation of similarities between tags based on their co-occurence during the indexing of resources as well as subsequent clustering enables us to countervail the deficiencies of current folksonomy visualisation methods. Furthermore, it allows a more efficient retrieval.

Les clusters de mots-clefs – recherche sémantique dans les nuages de mots-clefs

Dans les bibliothèques aussi, avec des conceptions comme le catalogue 2.0, caractères fonctionnels des services collaboratifs sont arrivés, de sorte que des nouvelles formes de recherche dans les catalogues en ligne deviennent possibles. Le thème de cette étude est le développement et la mise en œuvre d'une interface de recherche alternative en l'espace de web 2.0-services sous forme de clusters de mots-clefs. Par le calcul des similitudes des mots-clefs sur la base de son apparition commune lors de l'indexation des ressources et d'un clustering suivant, il est possible de contrarier les déficits des méthodes de visualisation actuelles des folksonomies et de permettre une recherche plus efficace.

1 Tag-Clouds als Visualisierung von Folksonomies

■ Die Interfaces von Folksonomies erlauben dem Nutzer den Zugang zum Content auf zwei verschiedene Weisen: über die manuelle Eingabe der Suchargumente in ein Suchfenster oder über Browsen durch das Vokabular, die Tags. Besonders bei konkreten und speziellen Suchanfragen, kann die erste Variante als geeigneter angesehen werden (1). Die zweite Variante, das Suchen von Ressourcen über Tags und das Verfolgen von verknüpften Links bis zum gewünschten aber wohl zufällig treffenden Dokument, die Serendipity, sieht Mathes als Stärke der Folksonomies an (2).

Hinsichtlich der Organisation und Repräsentation von Tags, haben sich als Such- und Navigationswerkzeuge verschiedene Methoden herausgebildet. Die wohl bekannteste und meist eingesetzte Methode ist die in Form von Tag-Clouds. Abbildung 1 zeigt die Tag-Cloud des zentralen Social Bookmark-Dienst Delicious.

Mit Hilfe des Dienstes Delicious haben Nutzer die Möglichkeit, ihre Bookmarks abzuspeichern, mit Tags inhaltlich zu erschließen und anderen zugänglich zu machen. Tags dürfen hierbei mehrfach vergeben werden, es handelt sich also um eine Broad Folksonomy.

Abbildung 1: Tag-Cloud des Bookmarkdienstes Delicious (http://delicious.com/tag)

Wie in Abbildung 1 erkennbar, visualisieren Tag-Clouds die Tags einer Folksonomy. In der Regel stellen sie die populärsten Tags in alphabetischer Reihenfolge dar, wobei die relative Größe und Gewichtung der Tags in Abhängigkeit ihrer relativen Nutzung stehen. Auf diese Weise bilden Tag-Clouds das wachsende Vokabular einer Folksonomy in ein Navigationstool um (3).

2 Eigenschaften und Usability von Tag-Clouds

Konkret erfüllen nach Rivadeneira et al. Tag-Clouds vier Funktionen: Mit Hilfe einer Tag-Cloud ist es dem Nutzer möglich, auch Tags mit in die Suche einzubeziehen, die das gesuchte Konzept abbilden und damit die Navigation erleichtern (*Search*). Zudem schaffen sie einen ersten Einstiegspunkt für die Nutzer bei unspezifischen Suchanfragen (*Browsing*). Als weiteres hilft der Blick auf die Tag-Cloud Nutzern einen Eindruck vom Content des Dienstes zu erhalten, Tag-Clouds besitzen damit eine Überblicksfunktion (*Impression Formation/Gisting*). Als letzten Faktor nennen Rivadeneira et al. die Möglichkeit durch eine Tag-Cloud zu erkennen, welche verschiedenen Aspekte (*Sets*) von Informationen abgebildet werden (*Recognition*) (4).

Hearst und Rosner merken an, dass diese Methode der Visualisierung von Folksonomies eine freundliche Atomsphäre schafft und den Einstieg in einen komple-

xen Dienst bzw. Website unterstützt. Auch sind Tag-Clouds nach Aussage der Verfasser in der Lage, durch einen zeitlichen Vergleich Trends im Tag-Gebrauch und damit im Content des Datenbestandes aufzuzeigen (5). Mit der Visualisierung von Tags über eine Zeitperiode beschäftigt sich beispielsweise Russel in seinem Tool Cloudalicious (6).

Inwieweit Tag-Clouds die Informationssuche unterstützen, untersuchten Sinclair und Cardew-Hall in einer Studie (1). Sie boten Nutzern zum einen eine Search Box und zum anderen eine Tag-Cloud zur Informationssuche. Als Ergebnis konnten sie festhalten, dass Tag-Clouds als visuelle Zusammenfassungen des Contents die von Rivadeneira genannten Funktionen erfüllen. Zudem stellten sie fest, dass das Scannen der Cloud und klicken auf die Tags den Nutzern leichter fällt als die Suchtermformulierung. Dies erleichtert nach Aussage der Verfasser auch den Zugang und das Retrieval in einer Fremdsprache (1).

Betrachtet man die Usability der Tag-Clouds in Studien genauer, sind interessante, aber auch zugleich unterschiedliche Ergebnisse zu beobachten. Rivadeneira et al. befassten sich mit verschiedenen Darstellungsmöglichkeiten der Tag-Clouds und testeten diese in Studien im Hinblick auf die Auswirkung verschiedener Parameter wie Größe, Farbe und Platzierung der Tags (4). Ihre Studien haben zum Ergebnis, dass Nutzer größer geschriebene Tags als signifikanter ansahen und ein vertikales nach der Tag-Größe gestaffeltes Listen-Layout zur Identifikation von thematischen Kategorien bevorzugten. Zu erwähnen ist zudem ein weiteres Ergebnis der Studie: werden in einer Tag-Cloud nur eine kleine Menge von Tags in einem räumlichen Layout visualisiert (Abb. 2), sollten kleiner geschriebene Tags im *„upper left"* (UL) Quadranten platziert werden, um die Aufmerksamkeit der Nutzer auf sich zu ziehen. Dementsprechend sind größere Tags in den anderen drei Quadranten zu platzieren (a). Eine Platzierung von Tags in der Nähe des Größten jedoch hat keinerlei Auswirkung (b).

Eine weitere Untersuchung zur adäquateren Aufbereitung mit Fokus auf die genannten Parameter wurde von Halvey und Keane durchgeführt (7). Ihre Tag-Cloud Varianten umfassten neben den alphabetischen Tag-Clouds, horizontalen und vertikalen Listen auch die ungeordneten Pendants. Hier kamen die Nutzer am Besten mit der (traditionellen) alphabetischen Reihenfolge innerhalb der Tag-Clouds zurecht, schienen diese aber eher zu scannen als wirklich zu lesen, wobei auch hier die Darstellung der Tags in verschiedenen Größen als sehr hilfreich eingestuft wurde.

Auch Bateman et al. testeten in ihrer Studie, welche Wirkung die Veränderung verschiedener visueller Eigenschaften innerhalb der Tag-Cloud nach sich zieht (8). Abbildung 3 bildet diese Parameter ab und setzt sie insofern in Relation zueinander, als dass die Richtung der Pfeile angibt, inwieweit andere Eigenschaften von der Veränderung der Ersteren betroffen sind. Auf dieser Basis konstruierten Bateman et al. zehn Cloud-Sets, in welchen jeweils zehn verschiedene Tag-Clouds mit unterschiedlichen Eigenschaften visualisiert wurden und die Probanden nutzten, um Suchanfragen zu bearbeiten.

Bateman et al. stellten fest, dass

> "Larger font size were chosen reliably and with the strongest effect among any visual features (8)."

Die Schriftgröße hat also auch in dieser Studie einen starken Einfluss auf das Retrieval. Ebenso konnte als Ergebnis festgehalten werden, dass die Veränderung der Schriftgewichtung und der Intensität eine gute Möglichkeit ist, die Aufmerksamkeit des Nutzers zu lenken (8).

Neben positiven Eigenschaften besitzen Tag-Clouds jedoch auch einige Nachteile. So erwähnen Hearst und Rosner beispielsweise, dass die Länge eines Wortes mit seiner Größe verschmolzen ist. Dadurch entsteht für Nutzer der Eindruck, dass Wörter mit mehr Zeichen populärer sind (5).

Zudem impliziert die Abbildung populärer Tags innerhalb der Tag-Cloud, dass die aufgezeigten Tags eine hohe semantische Dichte aufweisen, dabei besitzen diese auf Grund ihrer häufigen Verwendung einen eher schlechten Diskriminierungsgrad (9). So stellten Begelman et al. fest, dass auf diese Weise oftmals nur wenige Tags mit ihren Co-Tags eine ganze Tag-Cloud dominieren (10). Hieraus lässt sich ein weiterer Nachteil ableiten. Da Tag-Clouds zumeist den gesamten Datenbestand visualisieren, ist es sehr schwer für jüngere Tags und damit neue Bereiche in der Tag-Cloud überhaupt abgebildet zu werden. Trends nach Hearst und Rosner aufzuzeigen gelingt damit erst mit zeitlicher Verzögerung und nur über einen Vergleich mit beispielsweise dem Visualisierungstool Cloudalicious (6).

Als ein weiterer Nachteil kann festgehalten werden, dass Tag-Clouds fast keine Interaktivität hinsichtlich ihrer Darstellung erlauben. So können beispielsweise weder auf der Anbieter-Seite spezielle Aspekte personenbezogen hervorgehoben bzw. beworben werden noch besteht eine Möglichkeit auf der Nutzer-Seite, die Tag-Cloud individuell anzupassen.

Wohl das größte Problem innerhalb der Folksonomies und damit auch der Tag-Clouds ist jedoch das Fehlen jeglicher Semantik zwischen Tags (5, 9). Theoretische Ansätze für die strukturelle Erweiterung von Folksonomies werden beispielsweise unter den Stichworten „semantic enrichment" (11) und „emergent semantics" (12) diskutiert. Ein Versuch der wachsenden Semantik

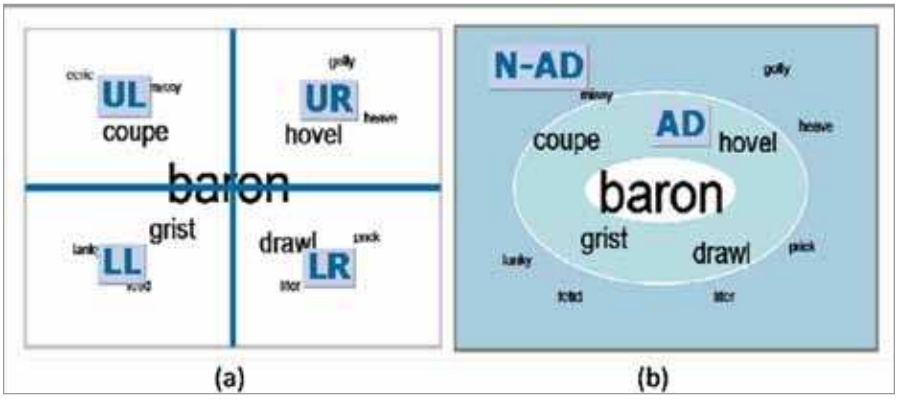


Abbildung 2: (a) In Quadranten eingeteilte räumliche Tag-Cloud (UL = upper-left, LL = lower-left, UR = upper-right, LR = lower-right); (b) Radiale Platzierung der Tags nach Größe in Abhängigkeit vom größten TAg (AD = adjacent, N-AD = non-adjacent) (4)

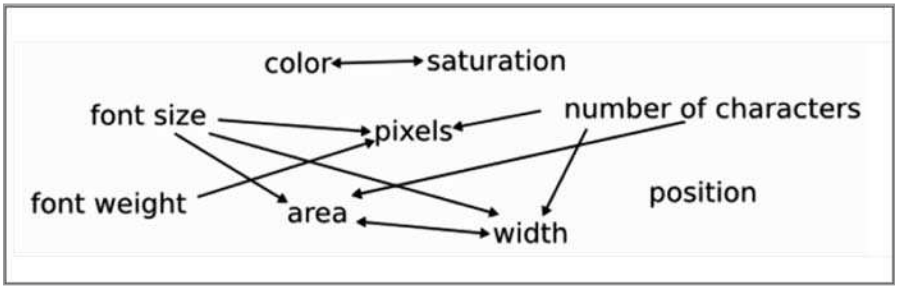


Abbildung 3: In der Studie betrachtete Merkmale und ihre Relation untereinander (8)

BibliotheksLeiterTag2010

ZUKUNFTSTHEMEN AUS DER PRAXIS FÜR DIE PRAXIS

5. Oktober 2010 in Frankfurt

Besuchen Sie uns auch vom 06.-10. Oktober auf der Frankfurter Buchmesse!
Halle 4.2 Stand P439

Auf dem Bibliotheksleitertag 2010 berichten hochkarätige Referenten in einem breiten Spektrum aus erfolgreichen Projekten über Bibliotheks-Kooperationen und - Vernetzung:

- **Gemeinsam stärker. Die Hauptbücherei Wien und ihre Kooperationspartner**
  Christian Jahl, Leiter der Hauptbücherei der Büchereien Wien
- **Die Bibliothek als kooperativer Bildungs- und Kulturpartner in Stadt und Region**
  Dr. Dirk Wissen, Direktor der Stadt- und Regionalbibliothek Frankfurt (Oder)
- **Ausgezeichnetes Informations- und Mediennetzwerk für Schüler/innen und Lehrer/innen**
  Simone Vetter, Leiterin Medienservice und IMeNS-Zentrale in Wetzlar
- **Lesestark! – Dresden blättert die Welt auf**
  Dr. Arend Flemming, Bibliotheksleiter der Städtischen Bibliotheken Dresden
- **Soziale Netzwerke - und wie Bibliotheken Zielgruppen neu für sich gewinnen können**
  Alexander G. Wehrmann, Managing Director LEWIS PR, Frankfurt (Main)

Mehr Informationen und Anmeldung unter:
www.bibliotheksleitertag.de, Telefon +49 (0)6324-9612-4211
oder E-Mail: bond@bond-online.de

Eine Veranstaltung von
Bibliothekssysteme bond

innerhalb von Folksonomies entgegenzukommen, ist die Berücksichtigung der Tag-Co-Occurrences bei einer Visualisierung.
Tag-Clouds weisen zwar einige Vorteile auf, aber es sollte Ziel sein, Nachteile zu beheben und vor allem den semantischen Anreicherungen durch die Beschreitung neuer Wege zu entsprechen.
Aus diesem Grund wird in dieser Studie ein alternativer Ansatz zur Verbesserung bzw. Ergänzung der Tag-Clouds vorgeschlagen und realisiert. Konkret thematisiert dieser Ansatz den Aufbau syntagmatischer Netze auf Basis von Tag-Co-Occurrencen und mit Hilfe von verschiedenen Cluster-Verfahren.

3 Alternative Layouts durch semantische Anreicherung von Tag-Clouds

Um den Defiziten von Tag-Clouds entgegen zu wirken, gibt es Ansätze, welche neue Wege hinsichtlich der Repräsentation des Vokabulars einer Folksonomy einschlagen und damit auf verschiedene Weisen versuchen, semantische Bezüge abzubilden. Da diese Ansätze eine wissenschaftliche Grundlage für die vorliegende Studie bilden, sollen einige an dieser Stelle vorgestellt werden.
Neben den aufgezeigten Studien zur Usability der Tag-Clouds in Kapitel 2, beschäftigten sich auch Kaser und Lemire mit der Gebrauchstauglichkeit von Tag-Clouds (13). In ihren Untersuchungen betrachteten sie verschiedene Algorithmen zur Optimierung der Usability und orientierten sich daran, dass semantisch ähnliche Tags in Bezug gesetzt werden sollten. Ähnlichkeit bedeutet an dieser Stelle jedoch nicht, dass zwei Tags das gleiche semantische Konzept abbilden, sondern vielmehr dadurch in einer semantischen Relation zueinander stehen, als dass sie innerhalb einer Ressource zu inhaltlichen Beschreibung verwendet wurden. Damit gingen Kaser und Lemire hinsichtlich Repräsentation des Vokabulars innerhalb von Folksonomies in eine andere Richtung. Sie realisierten ihre Variante der Tag-Cloud durch Verwendung des klassischen Knuth-Plass Algorithmus für die Textausrichtung (14) und einer Platzierungsanwendung nach Skiena (15).
Doch bereits 2005 wurden erste Ansätze unternommen, die Ähnlichkeiten von Tags, Usern und Links zur Visualisierung zu nutzen und Tag-Clouds als Graph auf einer „Map" (Abb. 4) abzubilden (16).
Begelman, Keller und Smadja verfolgten den Ansatz des automatischen Clusterings auf Basis von Tag-Co-Occurrences, um ein effektiveres Retrieval zu ermöglichen (10). Nach der Berechnung der Ähnlichkeiten von Tags, folgte ein anschließendes Spectral Clustering der Tags. Begelman et al. gelang es, verschiedene thematische Cluster zu generieren, in denen an den Kanten die errechneten Ähnlichkeitswerte abgetragen wurden und die Tags die Knotenpunkte bildeten. Abbildung 5 zeigt das Ergebnis des Verfahrens.

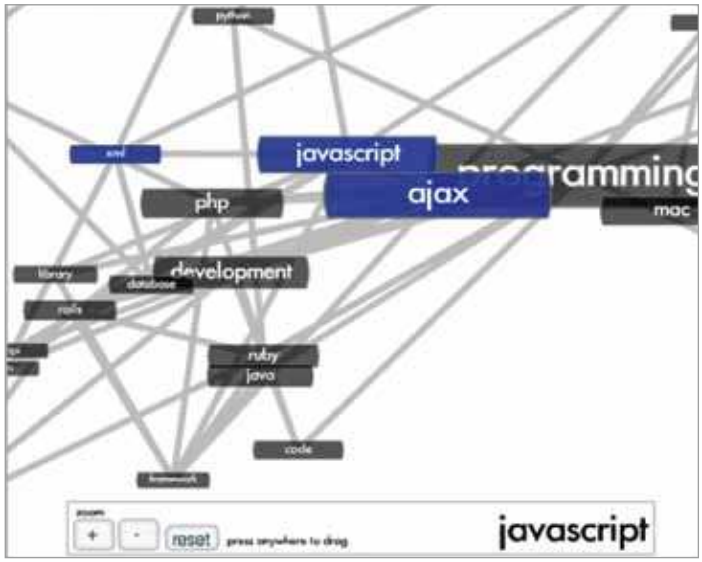


Abbildung 4: Abbildung einer Tag-Cloud als ungerichteter Graph auf Basis von Ähnlichkeitsalgorithmen (16)

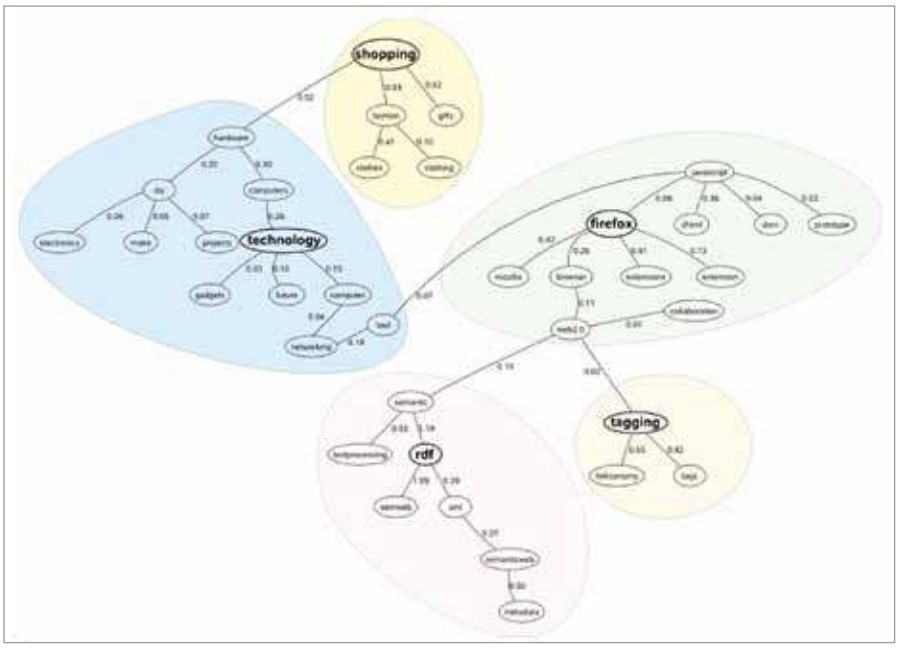


Abbildung 5: Ergebnis des Clusters (10)

Abbildung 6: Verbesserte Tag-Cloud auf Basis von Ähnlichkeitsberechnungen und Clustering (9)

Auch Hassan-Montero und Herrero-Solana führten Ähnlichkeitsberechnungen auf Basis der relativen Co-Occurences durch, um Tags miteinander in Beziehung zu setzen (9). Nach der Berechnung der Tag-Ähnlichkeit nach Jaccard, clusterten sie die Tags hierachisch nach dem k-means-Algorithmus (17).

Abbildung 6 zeigt eine nach dieser Vorgehensweise generierte Tag-Cloud, in welcher verwandte Tags in Bezug zueinander gesetzt wurden und damit den Nutzern das Retrieval nach verwandten Konzepten erleichtert werden soll.

Die verbesserte Tag-Cloud teilt mit der ursprünglichen Tag-Cloud 65 Tags. Viele, nach Aussage der Verfasser nicht signifikante Tags wie „toread" oder „cool" sind verschwunden und häufige Tags mit größerer Bedeutung und einem höheren Diskriminierungsgrad wie „philosophy", „food" oder „cinema" sind hervorgetreten.

Eine ähnliche Vorgehensweise nutzten auch Cattuto et al., um semantische Netze auf Basis von Tag-Co-Occurrences aufzubauen, jedoch mit dem Ziel, die Netzwerkstruktur von Folksonomies zu vergleichen (18). Diesen Ansatz führten sie 2009 weiter und analysierten Ähnlichkeiten zwischen Tags und Ressourcen (19).

Inwieweit ein effektives Browsing innerhalb einer großen Menge von Annotationen mit Hilfe von Clustering möglich ist, zeigten Li et al. (20). Sie ermöglichten zudem eine hierarchische und semantische Suche innerhalb der von ihrem System ELSABer generierten Tag-Cloud (Abb. 7). In dieser haben Nutzer über einen für sie relevanten Suchterm Zugang zu semantisch ähnlichen bzw. verwandten Tags und den entsprechnenden URLs.

Dass eine auf Basis von Tag-Co-Occurrences und Cluster Algorithmen generierte Tag-Cloud durchaus auch differenziertere Sichtweisen (personal, social, universal viewpoint) berücksichtigen kann, zeigten Huang et al. (21). Die Tag-Cloud passt sich hier je nach Wunsch beispielsweise den persönlichen oder den universellen Vorlieben an und erlaubt damit unterschiedliche Blickwinkel.

Eine Evaluation der Effekte von semantischer versus alphabetischer und zufälliger Anordnungen von Tags in Tag-Clouds führten Schrammel et al. durch (22). Sie konnten feststellen, dass bei einer spezifischen Suche die semantisch geclusterte Tag-Cloud mit zufälliger Anordnung der Tags eine Verbesserung bietet und zudem die Aufmerksamkeit der Nutzer auch auf kleiner geschriebene Tags lenkt.

Diese und andere Ansätze (12; 23) betrachten die Thematik der wachsenden semantischen Anreicherungen innerhalb von Folksonomies und versuchen, mögliche Lösungen anzubieten.

Zudem gibt es einige Demos zur Visualisierung von Tags (Abb. 8-12), welche im Web zugänglich sind. Exemplarisch seien an dieser Stelle fünf genannt.

Abbildung 7: Interface mit Social Browsing System (20)

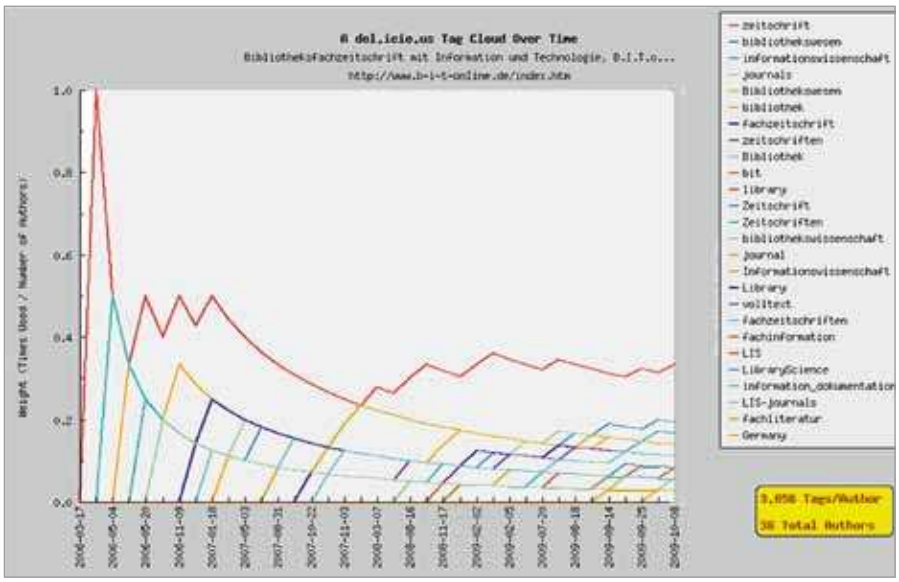


Abbildung 8: Screenshot Cloudalicious

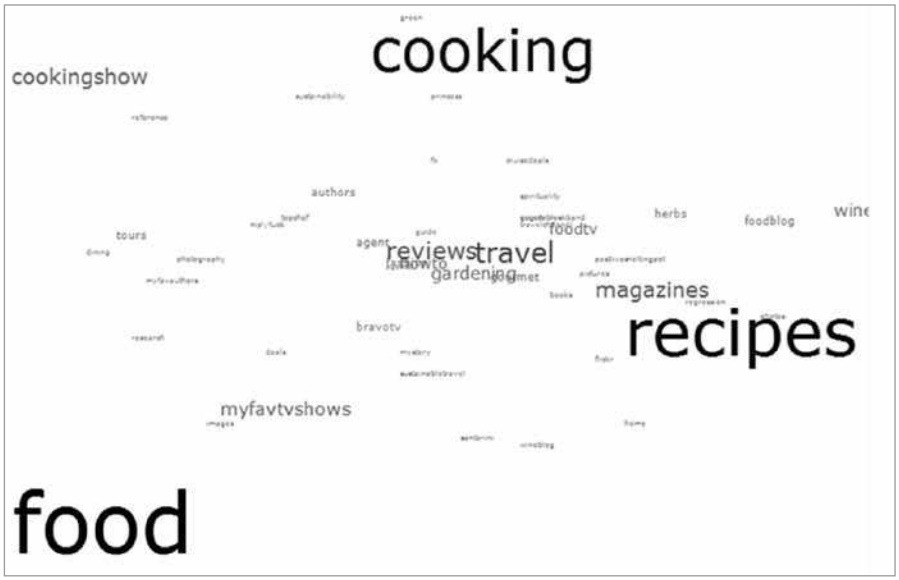


Abbildung 9: Screenshot Extispicious

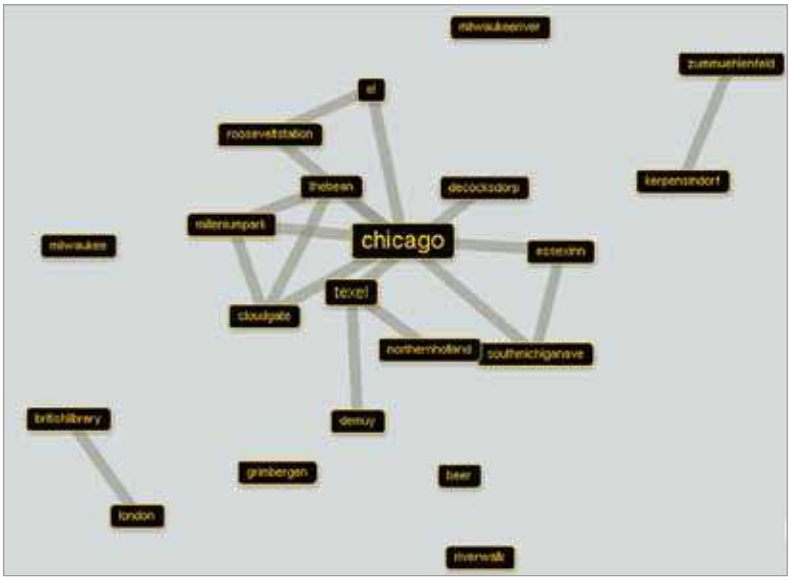


Abbildung 10: Screenshot Netr.it

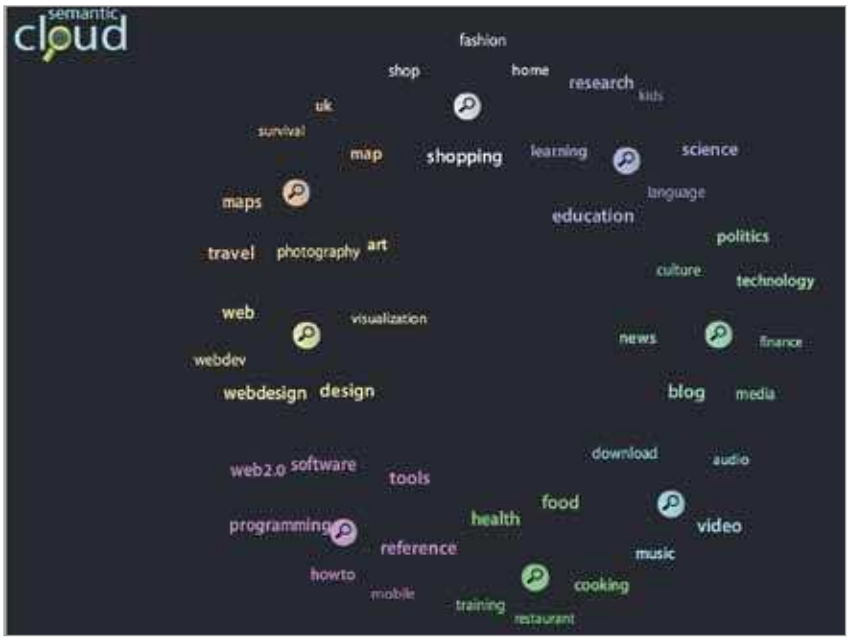


Abbildung 11: Screenshot Semantic Cloud

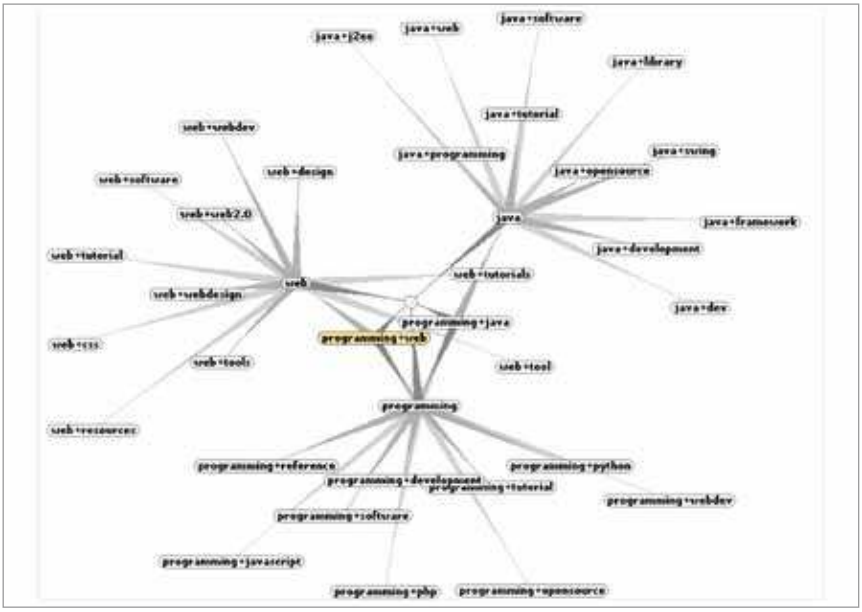


Abbildung 12: Schreenshot HubLog bei Eingabe der Tags „Programming", „Java" und „web"

- Cloudalicious[1] (6) generiert Graphen, in welchen die vergebenen Tags einer URL auf Delicious hinsichtlich ihrer quantitativen Nutzung über einen bestimmten Zeitraum angezeigt werden (Abb. 8).
- Extispicious[2] gibt eine zufällige textliche Streuung der persönlichen Nutzertags wieder (Abb. 9).
- Netr.it[3] visualisiert ein Co-Occurrences-Network auf Basis der persönlichen FlickR Tags, welches sich zudem manuell verändern bzw. erweitern lässt (Abb. 10).
- Semantic Cloud[4] generiert eine semantische Tag-Cloud auf Basis von Ähnlichkeiten zwischen Tags und anschließendes Clustering und emöglicht zudem eine hierarchisches Retrieval (Abb. 11).
- HubLog[5] ermöglicht ein visuelles Browsing zwischen verwandten Delicious Tags (Abb. 12).

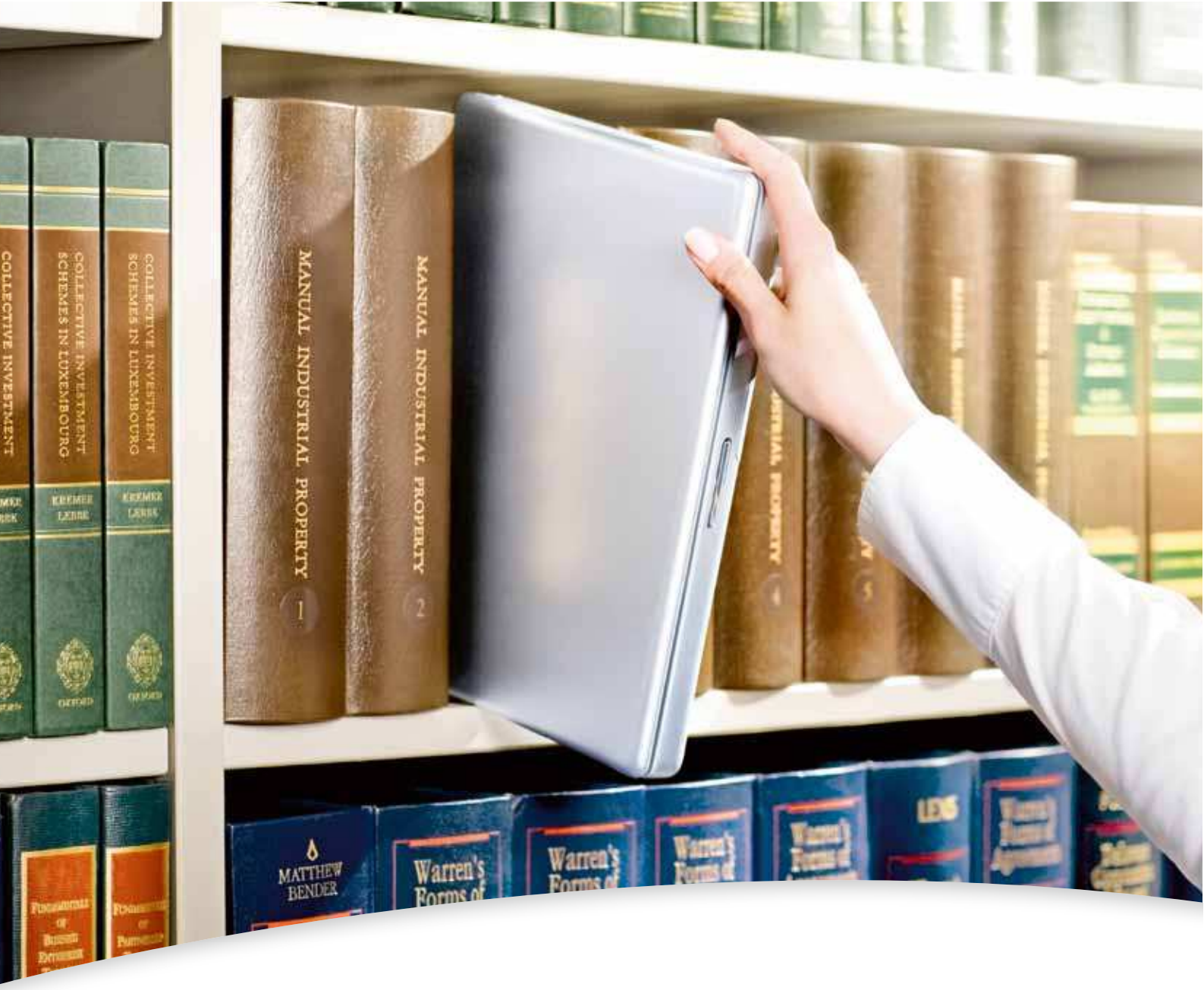
4 Realisierung von Tag-Clustern

Im Folgenden wird die Vorgehensweise zur Realisierung von Tag-Clustern aufgezeigt. Hierfür wird zunächst auf die grundlegende Struktur und deren Implementierung eingegangen (4.1). Konkret wird in einem ersten Schritt das Data Set vorgestellt, welches die Grundlage für die weitere Vorgehensweise bildet. Auf Basis der extrahierten Daten können anschließend die für das Clustering benötigten Ähnlichkeiten berechnet werden. Letzte Schritte bilden die Visualisierung, die Implementierung der Ranking-Algorithmen und die Einbindung in ein User Interface. Die Betrachtung der Ergebnisse der Vorgehensweise und Realisierung von Tag-Clustern mit Hilfe eines Applets bildet den zweiten Teil des Kapitels (4.2).

4.1 Materialien und Methode

An dieser Stelle soll das Konzept der Implementierung aufgezeigt werden. Hierzu wird auf das Model-View-Controller-Modell (MVC-Pattern-Modell) zurückgegriffen, um eventuelle spätere Änderungen oder Erweiterungen zu erleichtern. Die Umsetzung erfolgt in der Programmiersprache Java.
Wie in dem Klassendiagramm (Abb. 13) dargestellt, wurde die Applikation in vier Pakete unterteilt. Das Paket *cluster* ist dem Model, das Paket *parser* dem Control-Teil und das Paket *graph* dem View-Teil zuzuordnen. Die Klassen, welche sich direkt im Eltern-Paket *javatagclustering* befinden, zählen sowohl zum View- als auch zum Control-Teil.
Das *cluster*-Paket besteht aus den Klassen

1 http://cloudalicio.us/
2 http://kevan.org/extispicious
3 http://www.netr.it/
4 http://semanticcloud. rieskamp.info/
5 http://hublog.hubmed.org/archives/001049.html



Bei Schweitzer finden Sie Ihre Fachliteratur. Und nahezu 100.000 E-Books der Electronic Book Library.

Schweitzer Fachinformationen ist alleiniger Vertriebspartner der Electronic Book Library (EBL) in Deutschland. Und kann damit bereits jetzt annähernd 100.000 Titel aus weltweit 300 Verlagen digital anbieten. Ob Sie nun E-Books und E-Journals einzeln bestellen oder eine Komplettlösung für den E-Content Ihrer Bibliothek suchen: Mit Schweitzer Fachinformationen sind Sie immer gut beraten.

www.schweitzer-online.de

schweitzer
Fachinformationen

TagSimilarity, *Cluster*, *TagEntry* und *URLEntry*. Die Klasse *Cluster* ist dabei diejenige, über die auf das Modell zugegriffen werden kann. Mit ihr sind über eine Aggregations-Verbindung die Klassen *TagEntry* und (indirekt über *TagEntry*) *URLEntry* verbunden.
Das Paket *parser* besteht nur aus einer einzelnen Klasse *Parser*, welche über Methoden den Zugriff auf Delicious-Ergebnisse ermöglicht. Die Klasse ist über eine 1:1-Assoziation mit der Klasse *Cluster* verbunden. Das *graph*-Paket ist für die Darstellung verantwortlich und wird über die *ClusterGraph*-Klasse verwendet. Diese Klasse ist über Assoziationsrelationen mit *Cluster* und *TagCluster* verbunden, wobei letztere Klasse zur Darstellung des Graphen nötig ist. Das Paket *javatagclustering* ist das Eltern-Paket und bietet mit der Klasse *MainJFrame* eine grafische Benutzeroberfläche zum Zugriff auf das Modell an. Die Klasse ist über eine Assoziationsrelation mit der Klasse *ClusterGraph* verbunden.

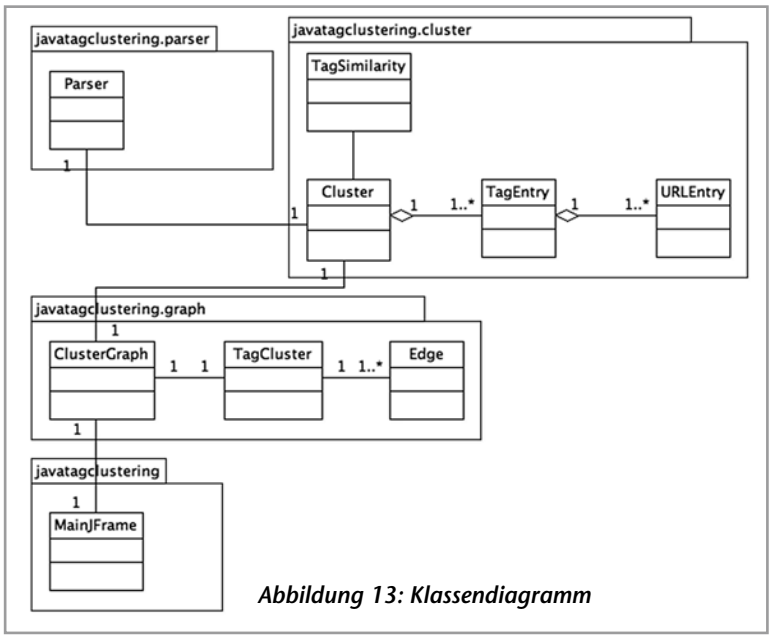


Abbildung 13: Klassendiagramm

4.1.1 Data Set

Damit Tag-Cluster realisiert werden können, bildet eine Menge von Bookmarks und die zugehörigen Tags eine Voraussetzung. Um die Themenunabhängigkeit zu demonstrieren, wurden exemplarisch zwei Mengen zu verschiedenen Themen aus dem zentralen Bookmark-Dienst Delicious extrahiert. Die Themen umfassen zum Einen die Bookmarks rund um den Tag *stemmer* und zum anderen Bookmarks im Bereich *recipes*, *cooking* und *seafood*.
Da die von Delicious zur Verfügung gestellte API nur den Zugriff auf die von einem spezifischen Benutzer vergebenen Tags bzw. getaggten Bookmarks erlaubt, musste ein auf den Kontext angepasster Parser geschrieben werden. Die Hauptaufgabe des Parsers besteht darin, bei einer vorgegebenen Suchanfrage, diese bei der Delicious-Tagsuche einzugeben und die Ergebnisse auszuwerten. Diese Ergebnisse setzen sich aus der URL des Bookmarks, dem Nutzernamen und der Anzahl, wie oft diese URL gebookmarkt wurde, zusammen.
Die URLs wurden extrahiert und gesammelt, wobei eine URL nur dann in den Datenpool aufgenommen wurde, wenn sie nicht bereits enthalten ist, um Dubletten zu vermeiden.
Mit Hilfe einer Bookmark-History ist es möglich, an die zu einem Bookmark zugehörigen Tags zu gelangen. Die History zeigt die Bookmarks inklusive der vergebenen Tags der letzten 2000 Nutzer. Damit es zu keinem Datenverlust kommt, wurde bei der Auswahl der Themen darauf geachtet, dass kein Bookmark von mehr als 2000 Nutzern getaggt wurde. Der Parser ist in der Lage diese History für alle gespeicherten URLs aufzurufen und die vergebenen Tags zu extrahieren, ihre Häufigkeiten zu zählen und die URL zu speichern, in welcher sie auftreten.

```
stemmer 325 http://snowball.tartarus.org/ 54 1396 251 http://nlp.ipipan.waw
http://www.perl.com/pub/a/2004/02/19/plucene.html 1 150 44 http://www.mpi-i
stemming 112 http://snowball.tartarus.org/ 191 1396 251 http://www.dirolf.c
nlp 88 http://snowball.tartarus.org/ 73 1396 251 http://cst.dk/online/lemma
search 80 http://snowball.tartarus.org/ 71 1396 251 http://www.compuglobalh
programming 77 http://snowball.tartarus.org/ 74 1396 251 http://tartarus.or
language 75 http://snowball.tartarus.org/ 48 1396 251 http://nlp.ipipan.waw
algorithm 67 http://snowball.tartarus.org/ 32 1396 251 http://tartarus.org/
porter 59 http://snowball.tartarus.org/ 11 1396 251 http://www.dirolf.com/p
linguistics 50 http://snowball.tartarus.org/ 74 1396 251 http://tartarus.or
bildverarbeitung 49 http://www.idw-online.de/pages/de/news299153 1 7 6 http
imaging 49 http://www.idw-online.de/pages/de/news299153 2 7 6 http://vision
algorithms 48 http://snowball.tartarus.org/ 76 1396 251 http://tartarus.org
stem 47 http://snowball.tartarus.org/ 5 1396 251 http://tartarus.org/~marti
tools 44 http://snowball.tartarus.org/ 31 1396 251 http://www.compuglobalhi
text 43 http://snowball.tartarus.org/ 33 1396 251 http://tartarus.org/~mart
php 43 http://snowball.tartarus.org/ 2 1396 251 http://www.compuglobalhiper
```

Abbildung 14: Liste der Tags nach Häufigkeit und der zugehörigen URLs

Im Folgenden wurden alle gefundenen Tags, die sich nur in der Groß- und Kleinschreibung unterscheiden, zusammenzufasst. Weitere Realisierungen in informationslinguistischer Hinsicht konnten in diesem Projekt noch nicht implementiert werden. Nach diesem Verfahren ergab die erste Suchanfrage mit dem Tag *stemmer* zunächst 599 Bookmarks, wobei abzüglich der Dubletten noch 327 unterschiedliche URLs übrig blieben. Der Parser lieferte nach der Untersuchung aller URLs 2743 Tags. Die zweite Suchanfrage mit den Tags *recipes*, *cooking* und *seafood* ergab vor der Eliminierung der Dubletten 684 Bookmarks, welche dann auf 518 schrumpfte und 2575 Tags lieferte (Tab. 1).

Tabelle 1: Zusammenfassung der Extraktionsergebnisse

Tag Menge	stemmer	recipes
Anzahl Bookmarks	599	684
Abzüglich Dubletten	327	518
Anzahl unterschiedlicher Tags	2743	2575

```
stemmer 342 10001 55 1396 251 10002 2 8 5 10003 1 4 4 10004 1 5 5 10005 1
stemming 137 10001 196 1396 251 10006 1 5 4 10007 12 166 43 10008 295 280
nlp 127 10001 80 1396 251 10004 1 5 5 10007 1 166 43 10008 57 2806 409 10
programming 109 10001 79 1396 251 10008 230 2806 409 10012 1 48 25 10014
search 109 10001 78 1396 251 10007 12 166 43 10008 95 2806 409 10014 78 1
language 92 10001 48 1396 251 10002 1 8 5 10008 135 2806 409 10012 2 48 2
porter 84 10001 13 1396 251 10006 1 5 4 10008 72 2806 409 10013 1 24 17 1
algorithm 81 10001 32 1396 251 10008 237 2806 409 10012 4 48 25 10014 32
algorithms 66 10001 77 1396 251 10008 188 2806 409 10012 3 48 25 10014 77
```

Abbildung 15: Liste der Tags mit eindeutiger URL-ID, um Speicherplatz zu sparen

Der endgültige Datensatz für jeden Tag enthält neben der Anzahl der URLs, in denen er als Tag enthalten war, eine Liste aller seiner zugehörigen URLs (vgl. Abb. 14). Diese Liste beinhaltet die URL, die Häufigkeit des Auftretens des Tags bei dieser URL, die Anzahl an insgesamt vergebenen Tags bei dieser URL und die Anzahl an verschiedenen Tags, die dieser URL hinzugefügt wurden. Für den Tag *stemmer* bedeutet dies, dass er 325mal in den URLs getaggt wurde, so z.B. bei *http://snowball.tartarus.org* 54mal. Insgesamt wurden bei diesem Bookmark 1396 Tags vergeben, worunter 251 verschiedene waren.
Um Speicherplatz zu sparen, wurden die URLs in diesen Listen durch eine eindeutige Identifikationsnummer ersetzt (Abb. 15). Damit später allerdings auch die dazugehörige URL angezeigt werden kann, enthält der Parser eine Methode, die bei Übergabe einer ID die URL zurückliefert.

4.1.2 Berechnung der Tag-Ähnlichkeiten

Die weitere Realisierung von Tag-Clustern geschieht auf der Basis von Tag-Co-Occurrences. An dieser Stelle werden die klassischen Ähnlichkeitsmaße Dice (1), Cosinus (2) und der Jaccard-Sneath-Koeffizient (3) eingesetzt werden, um die Koinzidenzwerte (φ) zwischen zwei Tags *A* und *B* zu berechnen.

$$(1)\ \varphi(A-B) = \frac{2g}{(a+b)}$$

$$(2)\ \varphi(A-B) = \frac{g}{\sqrt{ab}}$$

$$(3)\ \varphi(A-B) = \frac{g}{(a+b-g)}$$

Auf den Dienst Delicious und die ausgewählten thematischen Mengen angewandt bedeutet dies, dass es sich in der Formel bei *a* um diejenigen Bookmarks handelt, welche den Tag *A* beinhalten, und analog bei *b*, dass es sich um die Menge der Bookmarks handelt, in welchen der Tag *B* vorkommt. Bei *g* handelt es sich um die Bookmarks, in denen beide Tags gemeinsam vorkommen. Ähnlichkeit bedeutet hier jedoch nicht, dass die Tags identische Konzepte repräsentieren, sondern wir gehen davon aus, dass die Tags in irgendeiner Weise in semantischer Beziehung zueinander stehen, da sie gemeinsam zur Erschließung eines Bookmarks verwendet wurden. Eine exakte Identifikation der Relation kann an dieser Stelle bisher nur intellektuell erfolgen.
Technisch stellen sich die Anforderungen in einem ersten Schritt die drei Algorithmen zur Ähnlichkeitsberechnung zu implementieren und in einem weiteren die Ähnlichkeiten der Tags nach einem ausgewählten Algorithmus zu berechnen.
Hierzu wurden zwei Klassen realisiert: class *Cluster* und class *TagCluster*. Diese sind über eine 1:1-Aggregation verbunden. Als öffentliche Schnittstelle werden eine Funktion zur Auswahl des eingesetzten Ähnlichkeitsmaßes und ein öffentliches Feld zum Auslesen der Ähnlichkeiten zwischen den Tags zur Verfügung gestellt.
Die entwickelte Software bietet dem Nutzer die Möglichkeit an dieser Stelle auszuwählen, welches Ähnlichkeitsmaß zur Berechnung verwendet werden soll. Da die verschiedenen Berechnungsmethoden Einfluss auf die Cluster-Bildung und damit auf die Visualisierung haben, soll die Evaluation zeigen, welches Ähnlichkeitsmaß sich besser zur Visualisierung eignet bzw. von den Nutzern bevorzugt wird.

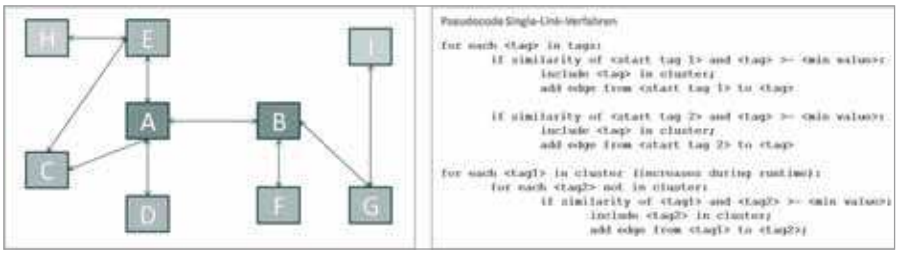


Abbildung 16: Clustering nach dem Single-Link-Verfahren und Pseudocode

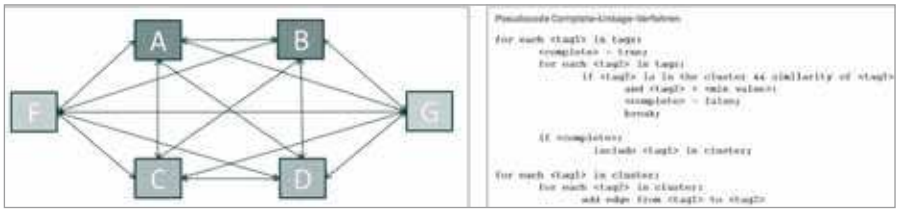


Abbildung 17: Clustering nach dem Complete-Link-Verfahren und Pseudocode

4.1.3 Clustering und Visualisierung

Nach der Berechnung der Koinzidenzwerte für alle Tag-Paare in der Treffermenge kann nun eine Klassifikation der Tags nach den errechneten Ähnlichkeiten vorgenommen werden. Für diese Clusteranalyse wurden drei verschiedene Verfahren ausgewählt: das Single-Link-, das Complete-Link- und das Group-Average-Verfahren (24).
Beim Single-Link-Verfahren (Abb. 16) wird von dem Tag-Paar (A, B) ausgegangen, welches die höchste Ähnlichkeit besitzt und damit das Ausgangspaar bildet. Dieses muss jedoch eine weitere Bedingung erfüllen: es muss in mindestens 50 Bookmarks vorkommen, um eine elaborierte Clusterbildung zu gewährleisten. Anschließend werden alle Tags hinzugefügt, die mit dem Tag A ähnlich sind (C, D und E). Hierbei kann bei der Clusterbildung ein Schwellenwert angegeben werden, der eine Ausuferung des Clusters verhindert. Ebenso wird mit dem Tag B (F, G) und den neu gewonnenen Tags verfahren (H, I).
Erfolgt die Cluster-Bildung durch das Complete-Link-Verfahren (Abb. 17), so bildet wieder das Tag-Paar (A, B) mit der höchsten Ähnlichkeit das Ausgangspaar, jedoch werden nur die Tags mit in den Cluster aufgenommen, welche mit A **und** B eine Ähnlichkeit über einen bestimmten Schwellenwert besitzen (C, D). Mit den neugewonnenen Tags wird ebenso verfahren (F und G).
Das Group-Average-Verfahren (Abb. 18) geht in einem ersten Schritt wie das Single-Link-Verfahren vor. Nachdem der Cluster auf die aufgezeigte Weise gebildet wurde, wird für diesen jedoch hinsichtlich der Ähnlichkeit ein arithmetischer Mittelwert berechnet. Der Mittelwert in Abbildung 19 beträgt 0,59. Im letzten Schritt werden alle Tags entfernt, die nicht eine Ähnlichkeit über diesen Schwellenwert mit ihren jeweiligen Ausgangstags besitzen. Auf diese Weise fallen die Tags H, G und I aus dem Tag-Cluster heraus.

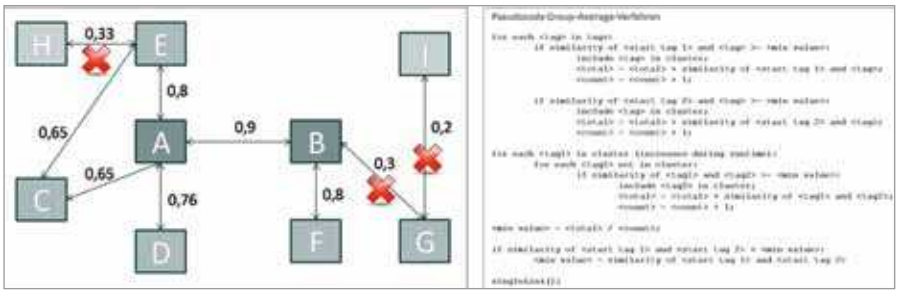


Abbildung 18: Clustering nach dem Group-Average-Verfahren (Mittel 0,59) und Pseudocode

Auf Basis der Koinzidenzwerte für alle Tag-Paare in der Bookmarkmenge lassen sich nun Tag-Cluster in Form von syntagmatischen Netzen realisieren. Um den komplexen Prozess der Visualisierung eines Graphen zu vereinfachen, wurde an dieser Stelle auf das „Java Universal Network/Graph Framework" (JUNG) zurückgegriffen, welches unter der BSD-Lizenz zur freien Verfügung steht. Somit besteht der Cluster aus ungerichteten Graphen, an welchen der errechnete Ähnlichkeitswert durch die Stärke der Graphen zwischen den Tags visualisiert wird. Aus Gründen der Übersichtlichkeit wurden die Ähnlichkeiten in zweier Schritten zusammengefasst, so dass die Ähnlichkeiten 0.1 und 0.2, 0.3 und 0.4 etc. über eine gemeinsame Strichstärke verfügen. Da die Nutzer in den zu Beginn erwähnten Studien besonders die unterschiedliche Größe der Tags innerhalb der Folksonomy-Realisierung als positiv betrachten, wurde dieser Aspekt auch in der Bildung der Tag-Cluster berücksichtigt. Die Größe der Tags repräsentiert demnach wie bei herkömmlichen Tag-Clouds die quantitative Nutzung dieser. In diesem Fall werden die Tags mit Hilfe der Minimum-Maximum-Normalisierung durch zehn verschiedene Größen im Cluster realisiert (Abb. 19).

4.1.4 Ranking-Algorithmen

Die Treffermenge wird den Nutzern (zu Evaluationszwecken) auf zwei Weisen dargeboten. In einer ersten Variante werden die Bookmarks nach der absoluten Häufigkeit der Tags gerankt. Kommt es durch Klicken auf die Linie zwischen zwei Tags zu einer UND-Verbindung, wird ebenso für beide Tags die absolute Häufigkeit berechnet und anschließend beide Werte addiert. Die Höhe der Summe bestimmt das Ranking.
Die zweite Variante erfolgt nach der Berechnung von WDF * ITF (25). Hierbei handelt sich um die Betrachtung der *Within Document Frequency* (WDF), also der Logarithmisierung der relativen Häufigkeit, und der *Inverse Tag Frequency* (ITF), einen textstatistischen Wert, welcher auf die Gesamtzahl der Tags in der Bookmark-Menge Bezug nimmt.

$$\text{WDF} * \text{ITF} = \left[\frac{ld\ freq(t,b)+1}{ld\ L}\right] * \left[\left(ld\ \frac{M}{m}\right) + 1\right]$$

Bei *freq(t,b)* handelt es sich um das Vorkommen eines Tags *(t)* in einem Bookmark *(b)*, d.h. wie oft wurde *(t)* vergeben. *L* ist Gesamttagzahl des Bookmarks, also alle Tags UND deren Häufigkeit eines Bookmark. Die Gesamtzahl aller Tags in der Treffermenge (alle Tags in der gesamten Menge und deren Häufigkeit addieren) wird als *M* daklariert. Bei *m* handelt es sich um die Anzahl des Tags *(t)* in der Menge.

```
Pseudocode Tag-Gewichtung

<relative occurence> = number of times the tag occurs - minimum number
                       of times any tag in the cluster occurs
<span> = maximum number of times any tag in the cluster occurs -
         minimum number of times any tag in the cluster occurs

<freq9> = <span> * 9/10
<freq8> = <span> * 8/10
...
<freq1> = <span> * 1/10

if <relative occurence> > <freq9>:
        set font size to 22
if <relative occurence> > <freq8>:
        set font size to 21
...
if <relative occurence> > <freq1>:
        set font size to 14
else:
        set font size to 13
```

```
Pseudocode Linien-Gewichtung

<relative sim> = (similarity - minimum similarity) / (1 - minimum
                 similarity);

if <relative sim> >= 0.9:
        return 4.5
if <relative sim> >= 0.8:
        return 4.0
if <relative sim> >= 0.7:
        return 3.5
...
if <relative sim> >= 0.1:
        return 0.5
if <relative sim> < 0.1:
        return 0.25
```

Abbildung 19: Pseudocode der Tag-Gewichtung und der Linien-Stärke

4.1.5 User Interface

Letzter Schritt ist das Design des Applets und damit des User Interfaces. In diesem wird darauf wertgelegt, dass die Retrieval-Schnittstelle möglichst übersichtlich und damit nutzerfreundlich ist. Im Applet werden alle implementierten Funktionen verknüpft und der Nutzer ist in der Lage, selbstständig alle Werte, welche zur Clusterbildung herangezogen werden können, auszuwählen. Abbildung 20 zeigt die Realisierung des User Interfaces.
So kann über einen Schieberegler der Schwellenwert (0 - 1.0) und über verschiedene Buttons das Ranking- sowie das Clusterverfahren und der Ähnlichkeitsalgorithmus eingestellt werden. Auch die beiden exemplarisch ausgewählten Testmengen (*Tag Sets*) können einzeln ausgewählt werden.

4.2 Ergebnisse

Die neue Retrieval-Schnittstelle in Form von Tag-Clustern ist in der Lage, nicht nur eine Übersicht über den Content einer Datenbank zu geben, sondern zudem nach einer initialen Suchanfrage eine Hilfestellung innerhalb der Fülle an Treffern zu geben. Hierfür werden die Tags innerhalb der Treffermenge nach dem aufgezeigten Verfahren geclustert und visualisiert.

Anstelle oder als Ergänzung zu einer herkömmlichen Tag-Cloud (Abb. 1), hat der Nutzer nun einen Cluster zur Verfügung, welchen er interaktiv anpassen kann. Abbildung 21 zeigt einen Screenshot eines Cluster-Aufbaus. Hier ist erkennbar, dass 2575 Tags zur Cluster-Bildung herangezogen werden und dass das Tag-Paar mit der höchsten Ähnlichkeit eine Koinzidenz von 0.995 besitzt.

```
Number of Tags:2575
max:0.995107136031466
tag1:1
tag2:2
Log: getGraph - Thickness:true
got GraphEvent type:VERTEX_ADDED for 0
got GraphEvent type:VERTEX_ADDED for 1
got GraphEvent type:VERTEX_ADDED for 2
got GraphEvent type:VERTEX_ADDED for 3
got GraphEvent type:VERTEX_ADDED for 4
got GraphEvent type:VERTEX_ADDED for 5
got GraphEvent type:VERTEX_ADDED for 6
got GraphEvent type:VERTEX_ADDED for 7
got GraphEvent type:VERTEX_ADDED for 8
got GraphEvent type:EDGE_ADDED for 0
got GraphEvent type:EDGE_ADDED for 1
got GraphEvent type:EDGE_ADDED for 2
got GraphEvent type:EDGE_ADDED for 3
got GraphEvent type:EDGE_ADDED for 4
```

Abbildung 21: Screenshot des Cluster-Aufbaus

Abbildungen 22 und 23 zeigen automatisch generierte Cluster aus einer initialen Treffermenge von rund 400 Bookmarks und den 2575 Tags. Zur Bildung wurden an dieser Stelle exemplarisch das Complete-Link-Verfahren und eine Ähnlichkeitsberechnung nach Cosinus ausgewählt. Abbildung

Abbildung 20: Tag-Cluster User Interface

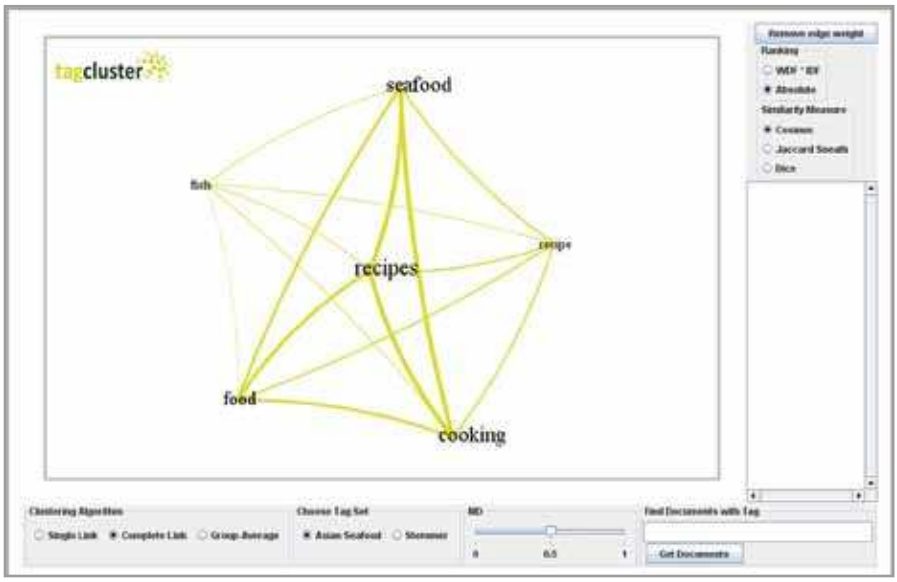


Abbildung 22: Automatisch generierter Complete-Link-Cluster mit einer Tag-Ähnlichkeit von 0.5; Ähnlichkeit berechnet nach Cosinus

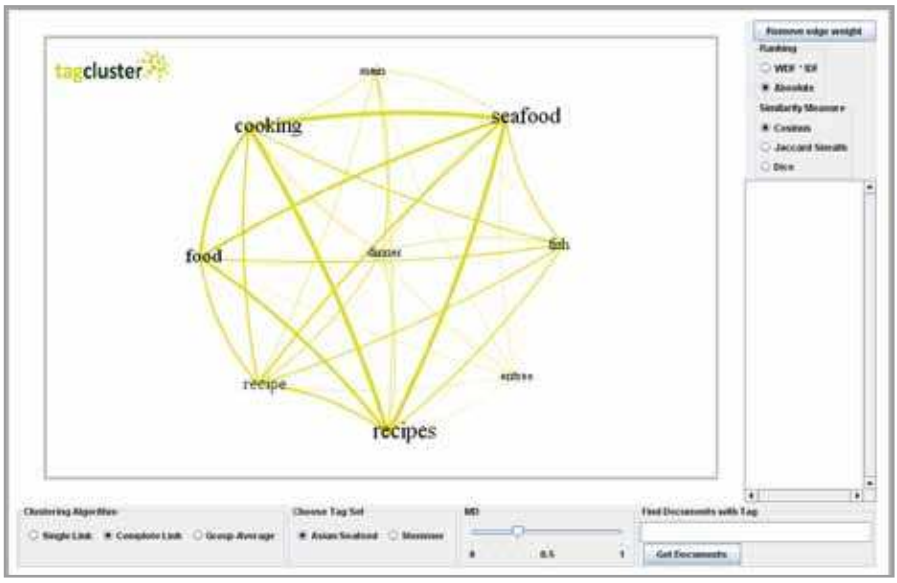


Abbildung 23: Automatisch generierter Complete-Link-Cluster mit einer Tag-Ähnlichkeit von 0.3; Ähnlichkeit berechnet nach Cosinus

22 zeigt einen Cluster mit einem eingestellten Schwellenwert von 0.5, in Abbildung 23 ist eine Cluster-Visualisierung mit einem Schwellenwert von 0.3 zu sehen. Durch die Einstellung der Vielzahl an Werten und Verfahren, ergeben sich eine große Anzahl an Cluster-Varianten. Die Senkung des Schwellenwerts von 0.5 auf 0.3 hat zur Folge, dass mehr Tags zur Clusterbildung zugelassen sind. Der Nutzer erhält hierdurch Zugang zu weiteren semantisch ähnlichen Tags, welche ihm evtl. vorher nicht bekannt oder bewusst waren und damit neue Aspekte beim Retrieval mit einbringen. Die Ähnlichkeit der Konzepte wird ihm durch die Linienstärke angezeigt. Die Größe der Tags gibt die Nutzung des Tags innerhalb der Treffermenge an. Durch Klicken auf einen Tag erhält der Nutzer die zugehörige Treffermenge. Eine UND-Verknüpfung und damit eine Einschränkung erhält er durch Klicken auf eine Verbindungslinie zwischen zwei Tags oder durch ein Anklicken von weiteren Tags. Das Ergebnis der Trefferliste passt sich dynamisch seiner Suchanfrage an und der Nutzer ist so in der der Lage, aktiv die Treffermenge zu beeinflussen und Subsets auszuwählen.

5 Evaluation

Dieses Kapitel thematisiert die Evaluation der Software und deren Ergebnisse. Hierzu wird zunächst auf die Methode und das Modell eingegangen, welches sich aus verschiedenen Bestandteilen, sowohl klassischen als auch neuen Elementen, zusammensetzt (5.1). In einem zweiten Teil werden die auf Grundlage des Modells erstellten Fragebögen ausgewertet und Ergebnisse präsentiert (5.2).

5.1 Methode

Um ein konkretes Web2.0 Service-Tool wie das Vorliegende zu evaluieren, ist es nicht sehr nützlich nur mit klassischen Qualitäts-Indikatoren wie beispielsweise Recall und Precision zu arbeiten. Das vorliegende IT-Evaluationsmodell Modell (Abb. 24) besteht aus drei Teilen: der Qualität des Informationssystems – inklusive des Retrieval Systems – (*IT system quality*), der Qualität der Wissensbasis (*knowledge quality*) und der Qualität des Services (*IT service quality*).

Eine etablierte Methode zur Evaluation von Diensten ist die Nutzung des SERVQUAL-Ansatzes (26). Dieser ist ein standardisiertes Verfahren zur Messung der Qualität von Dienstleistungen und stützt sich auf die Betrachtung von zwei Werten: zum Einen auf den Erwartungswert (*expectations E*) und zum Anderen auf den Beobachtungswert der Nutzer (*perceptions P*). Bei einer Evaluation hat der Nutzer bei jeder Frage – sowohl auf der Erwartungs- als auch bei der Beobachtungsseite – die Möglichkeit, auf Skalen von 1 (gar nicht) - 7 (absolut) seinen Standpunkt zu fixieren. Die Differenz zwischen Erwartung und Beobachtung (*P-E*) ist ein Ansatzpunkt hinsichtlich der Qualität (*Q*) des Systems. Parasuraman, Zeithaml und Berry definieren zudem fünf Dimensionen der Qualität, auf welchen das Verfahren basiert: *Reliability*, *Aussurance*, *Tangibles*, *Empathy* und *Responsiveness*.

Zentral jedoch ist der Wert, welcher sich durch die Differenz von Erwartung und Beobachtung ergibt (27). Der SERVQUAL-Ansatz kann auch zur Bewertung der Qualität bzw. Effektivität von Informationssystemen adaptiert werden (28). Jedoch kommt es in diesem Fall zu Kontroversen auf Grund der ausschließlichen Betrachtung der Differenz (29) und der zuvor definierten fünf Dimensionen der Qualität (30). Um den Defiziten des Ansatzes entgegenzuwirken, wird daher in der Forschung ein anderer Ansatz diskutiert: SERVPERF. Hier liegt der Fokus der Qualitätsmessung bei der Wahrnehmungskomponente bzw. bei der Beobachtung von Seiten der Nutzer. Das Qualitätsniveau, bestimmt durch die einzelnen Nutzer-Bewertungen, kann entweder als ergänzender Wert bei einer Evaluation herangezogen werden (31) oder als der einzige ausschlaggebender Faktor einer Evaluation dienen (32).

In der vorliegenden Evaluation wurden beide Varianten berücksichtigt, so dass der SERVQUAL-Ansatz eine solide Basis bildet, die um Aspekte des SERVPERF-Ansatzes ergänzt wurde und damit eine nützliche

Methode ist, um die Qualität von Software zu messen (33; 34). Neben der Berechnung der Differenz, werden somit auch die Erwartungs- und Beobachtungswerte an sich zur Evaluation herangezogen. Auf die vordefinierten fünf Dimensionen wurde jedoch nicht zurückgegriffen, da eigene, besser passende, Dimensionen entwickelt wurden.

Zusätzlich wird die Qualität der Software auf Basis klassischer Indikatoren für das Research in IT-Systemen gemessen: der wahrgenommenen Einfachheit der Bedienung (*perceived ease of use*) und der wahrgenommenen Nützlichkeit (*perceived usefulness*) (35). Mit Vertrauen (*trust*) wird in der Literatur ein weiterer Qualitäts-Faktor diskutiert (36), welcher ebenfalls mit in die Evaluation aufgenommen wird. Als relativ neuer, aber durchaus interessanter Aspekt, wird der empfundene Spaß (perceived playfulness (37) bzw. perceived enjoyment (38), den die Testpersonen bei der Benutzung des Tools haben, berücksichtigt. Dies ist insbesondere daher wichtig, als das gegrade dieser Unterhaltungsfaktor in Web2.0 Umgebungen im Vordergrund steht und maßgeblich über den Erfolg einer Applikation entscheiden kann.

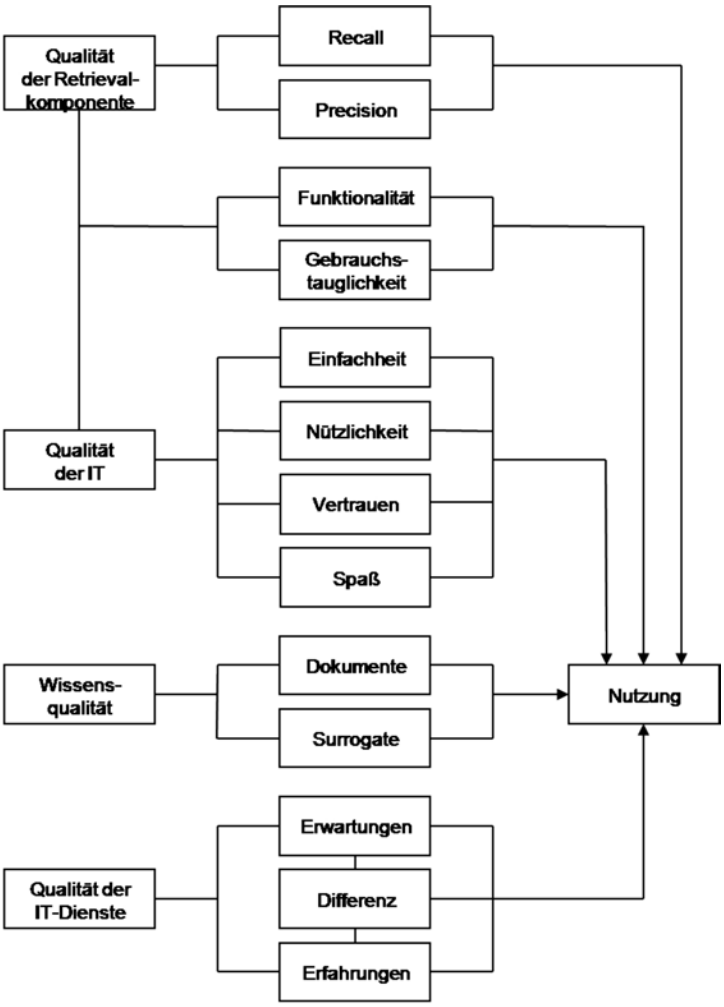


Abbildung 24: Tag-Cluster Evaluationsmodell

In dem vorliegenden Modell (Abb. 24) werden mit der Qualität der Zugangspunkte zu den Dokumenten und der Qualität derer (*knowledge quality*) zusätzliche Qualitätsindikatoren bzw. -dimensionen berücksichtigt. Diese können jedoch ebenso wie die klassischen Parameter Precision und Recall als Indikatoren des Information Retrieval Subsystems zur Qualitätsbewertung eines Informationssystems nicht berücksichtigt werden, da sie abhängig von dem Informationssystem (*Service*) selbst und nicht von den Applikationen wie Tag-Clouds oder Tag-Clustern sind. Zur Bewertung des Retrieval-Systems müssen differenzierte und vielfältige Anfragen gestellt werden, was auf Grund der geringen Datenbasis der Prototyp-Applikation nicht möglich ist und daher in zukünftigen Untersuchungen thematisiert werden sollen.

Das vorliegende IT-Evaluationsmodell ist damit eine modifizierte Version des DeLone und McLean Modells (39; 40) und des Jennex und Olfman Modells (41).

5.2 Ergebnisse

Um einen Vergleich zwischen herkömmlichen Tag-Clouds und den neu generierten Tag-Clustern anzustellen, wurde ein Fragebogen erstellt, welcher die im Methodenteil genannten Parameter und Indikatoren berücksichtigt.

Ein Pre-Test im Mai 2009 mit Information Professionals der Abteilung Informationswissenschaft am Institut für Sprache und Information der Heinrich-Heine-Universität (N=6) führte zur teilweisen Ergänzung bzw. Umformulierung der Fragen und damit zu einem Update des Evaluationsbogens. Die Endfassung des Evaluationsbogens umfasst 10 übergeordnete Fragen, welche sich auf insgesamt 26 Teilfragen aufteilen. Der Test zielt bei den ersten 6 Fragen darauf ab, einen Korrelation zwischen Erwartung und Beobachtung zu ziehen. Die letzten 4 Fragen zielen auf die genannten klassischen Parameter ab. Die Evaluation von Tag-Clouds (Delicious) und Tag-Clustern wurde durch 28 Studenten des 2. Semesters im Rahmen des Tutoriums zur Vorlesung „Wissensrepräsentation" und zwei Mitgliedern des Fachbereichs Informationswissenschaft der Heinrich-Heine-Universität im Mai 2009 nach einer thematischen Einführung durchgeführt.

Tabelle 2 zeigt die Ergebnisse der IT Service Qualität von Tag-Clouds und Tag-Clustern. Hieraus ist erkenntlich, dass die Testpersonen Tag-Cluster als Visualisierungsmethode von Folksonomies bevorzugten. Hinsichtlich der verschiedenen Varianten der Ähnlichkeitsberechnungen konnte keine Präferenz von Seiten der Nutzer festgestellt werden, nur eine leichte Tendenz zur Bevorzugung der Berechnung nach Cosinus. In Bezug auf die Clustering-Algorithmen eignete sich nach Evaluation durch die Nutzer das Group-Average-Verfahren, gefolgt vom Complete-Link-Verfahren, am besten. Als weiteres Ergebnis konnte festgehalten werden, dass bezüglich der zwei angebotenen Ranking-Algorithmen das WDF*ITF-Verfahren leicht besser bewertet wurde als das Ranking nach absoluter Häufigkeit. Der Erwartungswert von Nutzerseite war in Bezug auf das Einstellen des Schwellenwertes mit durchschnittlich 6.07 von 7.00 Punkten sehr hoch. Zwar konnte die Beobachtung diesen Wert nicht erreichen, war aber mit einem Wert von 5.60 Punkten ebenfalls hoch. Gleiches ist bei dem Erwartungswert bezüglich der Darstellung der Ähnlichkeit über die Linienstärke zu beobachten. Dieser Wert lag mit 6.37 Punkten von 7 noch höher und konnte mit einem hohen Beobachtungswert von 5.55 Punkten bestätigt werden.

Da der Fragebogen auch ein freies Anmerkungsfeld zur Verfügung stellte, nutzten die Evaluierenden dieses, um Verbesserungswünsche in Bezug auf die visualisierte Ähnlichkeit zu äußern. Vorschläge waren beispielsweise die Anbringung der Ähnlichkeitswerte an den Linien (Fragebogen 8), die Visualisierung von ähnlichen Tags in geringerer Distanz zueinander (Fragebogen 12) oder eine farbliche Abstufung der Linien je nach Ähnlichkeit (Fragebogen 10).

Tabelle 2: Ergebnisse der Evaluation hinsichtlich der IT service qualität

IT service quality indicator	Erwartungswert (SD)	Beobachtungswert (SD)	Differenz (SD)
1. Visualisierung von Folksonomies			
Tag-Clouds	5.69 (0.76)	3.52 (1.12)	-2.17
Tag-Cluster		5.39 (0.83)	-0.30
2. Berechnung der Ähnlichkeiten			
Dice	5.59 (1.05)	4.87 (1.36)	-0,72
Jaccard		4.73 (1.31)	-0.85
Cosine		5.10 (1.06)	-0.49
3. Cluster-Algorithmen			
Single-Link Verfahren	5.83 (1.00)	3.87 (1.50)	-1.96
Complete-Link Verfahren		4.76 (1.30)	-1.07
Group-Average Verfahren		5.10 (1.09)	-0.73
4. Relevance Ranking			
Absolute Häufigkeit	5.46 (1.29)	4.39 (1.60)	-1,07
WDF*ITF		4.86 (1.43)	-0.61
5. Tag-Cluster Charakteristiken			
Schwellenwert	6.07 (0.88)	5.60 (1.28)	-0.47
Linien-Stärke	6.37 (0.85)	5.55 (1.30)	-0.82

N: zwischen 28 und 30; SD: Standard-Abweichung; Skala: 1 (stimme nicht zu) bis 7 (stimme absolut zu)

In Bezug auf die Qualität des IT-Systems (Tab.3), konnte mit Hilfe der Evaluation festgestellt werden, dass sich sehr signifikante statistische Unterschiede zwischen Tag-Clouds und Tag-Clustern in Bezug auf die wahrgenommene Nützlichkeit (*perceived usefulness*), Vertrauen (*trust*) und Spaß (*perceived enjoyment*) ergeben. Nutzer präferierten in den genannten Punkten teilweise sehr deutlich Tag-Cluster. Tag-Cluster wurden nicht nur als nützlicher wahrgenommen, sondern die Benutzung brachte den Nutzern auch mehr Spaß und der visualisierte Content wurde von ihnen als vertrauenswürdiger eingestuft.

Tabelle 3: Evaluationsergebnisse hinsichtlich der Qualität des IT-System

Qualitäts-Indikator des IT-Systems	Durchschnitt Tag-Clouds (SD)	Durchschnitt Tag-Cluster (SD)	Signifikanz-Unterschied?
Einfachheit der Bedienung	5.17 (1.34)	5.57 (1.22)	Keine Signifikanz
Wahrgenommene Nützlichkeit	4.10 (1.35)	5.66 (1.23)	$\alpha < 0.001$
Vertrauen	2.90 (1.01)	4.43 (1.17)	$\alpha < 0.001$
Spaß	4.10 (1.37)	5.45 (1.35)	$\alpha < 0.001$

N: zwischen 28 und 30; SD: Standard-Abweichung; Skala: 1 (gar nicht) to 7 (absolut)

Seit 1872
pleuser etiketten

Spezialetiketten für Außen- und Innenbeschriftung von Büchern und Archivalien

- weichmacherfestes Papier und weichmacherfreier Kleber
- pH-neutral und alterungsbeständig
- Format und Gestaltung nach Wunsch
- verschiedene Papieroberflächen für jedes Beschriftungssystem
- garantierte Haftung auf allen Bucheinbänden und vielen anderen Materialien
- geprüft vom Institut für Papierfabrikation und Mechanische Verfahrenstechnik der TU Darmstadt
- Programm zur Optimierung der Signaturbeschriftung in Zusammenarbeit mit der Verbundzentrale des GBV

Bernhard Pleuser GmbH
Otto-Hahn-Str. 16
D-61381 Friedrichsdorf
Telefon +49 (0) 6175 . 79 82 727
Fax +49 (0) 6175 . 79 82 729
info@pleuser.de www.pleuser.de

Tabelle 4: Korrelation (Pearson, beidseitig) der IT-Qualitäts-Indikatoren

	Bedienung	Nützlichkeit	Vertrauen	Spaß
Bedienung	1			
Nützlichkeit	+0.42	1		
Vertrauen	+0.32	- 0.08	1	
Spaß	+0.31	+0.40	+0.19	1

Zudem konnte eine positive Korrelation zwischen der Bedienungsfreundlichkeit (*ease of use*) und allen anderen Indikatoren sowie mit Spaß identifiziert werden (Tab. 4). Die wahrgenommene Nützlichkeit korreliert stark mit der Bedienungsfreundlichkeit (+0.42) und mit Spaß (+0.40). Insgesamt scheint es, dass der Indikator Spaß eine große Bedeutung bei der Benutzung von Tag-Clustern besitzt, was die Akzeptanz einer solchen Retrieval-Schnittstelle durch die Nutzer erleichtern kann. Daher wurden im Folgenden auch die Korrelationen von Spaß mit allen anderen Indikatoren betrachtet. Eine hohe positive Korrelation (+0.43) ergab sich zwischen Spaß und den Erwartungen in Bezug auf die Einstellung des Schwellenwertes zur Clusterbildung mittels des Schiebereglers (Tab. 5). Je größer die Erwartungen der Testpersonen hinsichtlich der Justierung des Schiebereglers waren, umso stärker war der Spaß-Effekt bei der Clusterbildung (und umgekehrt) zu beobachten.

Tabelle 5: Korrelationen (Pearson, beidseitig) zwischen der Erwartung der Nützlichkeit der Schwellenwerteinstellung und der IT-Qualitäts-Indikatoren

Korrelation der Erwartung der Nützlichkeit der Schwellenwerteinstellung	
Bedienung	- 0.01
Nützlichkeit	+ 0.16
Vertrauen	+ 0.08
Spaß	+ 0.43

Offensichtlich ist das Einstellen der Clusterauflösung via Schieberegler und das daraus resultierende Hinzufügen von Tags durch einen geringeren Schwellenwert bzw. eine Abnahme von Tags durch einen höheren Schwellenwert mit Spaß von Seiten der Nutzer verbunden. Das Resultat dieser Option ist je nach Einstellung entweder eine differenzierteres Bild der Treffermenge oder die Visualisierung der Basis-Strukturen (Themen) in der Treffermenge. Unterstützt wird diese Aussage auch durch Kommentare im freien Anmerkungsfeld des Evaluationsbogens. So macht es Spaß „[…], die Schwellenwerte zu verändern und den Einfluss auf die Bookmarks zu sehen […]" (Fragebogen 8) und „die Interaktivität durch Einstellung des Schwellenwertes ist ein erheblicher Spaßfaktor, der Tag-Cluster als Visualisierung attraktiv macht" (Fragebogen 5).
Zusammenfassend lässt sich sagen, dass Tag-Cluster sehr gut von den Nutzern angenommen wurden und einen klaren Vorzug gegenüber Tag-Clouds erhielten.
Bei einer Implementierung sind jedoch einige Dinge zu beachten. So sollte auf Grund der Ergebnisse der Evaluation eine Bildung der Tag-Cluster über das Group-Average Verfahren erfolgen. Hinsichtlich der Berechnung der Tag-Ähnlichkeiten und der Ranking-Variante konnte keine deutliche Präferenz von Seiten der Nutzer festgestellt werden, so dass dies über eine Erprobung der Applikation erfolgen sollte. Da der Spaß bei der Clusterbildung erheblich durch die Einstellung des Schwellenwertes erfolgte, sollte diese Möglichkeit der Interaktivität, beispielsweise über einen Schieberegler, auch bei einer Implementierung zur Verfügung gestellt werden.

6 Abschlussbemerkung und Ausblick

Der in Web2.0-Diensten von Nutzern generierte Content steigt stetig an und bedarf neben der Strukturierung auch der Verbesserung des Zugangs, um ein effizientes Retrieval für die Nutzer zu gewährleisten. Daher wurde in dieser Arbeit mit Tag-Cluster ein alternativer Ansatz vorgestellt, der den derzeitigen Defiziten herkömmlicher Visualisierungsmethoden von Folksonomies versucht entgegen zu wirken und den Forderungen den semantischen Anreicherungen zu entsprechen nachkommt.
Die Nutzung syntagmatischer Relationen auf Basis der Tag-Co-Occurrences und das anschließende Clustering ermöglicht eine Reihe von Möglichkeiten und Verbesserungen hinsichtlich der visuellen Realisierung von Folksonomies. Zusammenfassend lässt sich festhalten:

- das Clustering ermöglicht eine größere kohärente visuelle Verteilung als die alphabetische Anordnung;
- den Forderungen der Einbeziehung semantischer Bezüge wird durch diesen Ansatz entsprochen;
- „Gleichberechtigung" aller Tags bei der Clusterbildung ermöglicht, dass auch neue Tags und Bereiche die Chance auf die Visualisierung haben; somit wird die eingangs erwähnte fehlende zeitliche Abgrenzung einer Tag-Cloud kompensiert;
- nach einer initialen Suche über Tag-Clouds kann die Fülle an Treffern durch Visualisierung der Tags in der Treffermenge in Form von Clustern ein weiterer/zusätzlicher thematischer Überblick geboten werden;
- durch die aktive Beeinflussung der Darstellung wird die Interaktivität für Nutzer und Anbieter realisiert;
- neue Tags im Cluster durch einen niedrigeren Schwellenwert geben Nutzern einen differenzierten Einblick, eine verminderte Anzahl von Tags auf Grund eines höheren Schwellenwerts lassen den Themenschwerpunkt erkennen;
- durch die Nutzung von Tag-Clustern sind User in der Lage, ihre Suchanfrage zu reformulieren und aktiv die Treffermenge anzupassen.

Neben den zahlreichen Vorteilen, gibt es auch einige Verbesserungen der Retrieval-Software, welche in Zukunft Gegenstand der Betrachtung sein könnten. So ist der Cluster-Aufbau in zeitlicher Hinsicht etwas ineffizient und muss für einen Gebrauch der neu entwickelten Retrieval-Schnittstelle innerhalb eines Web2.0-Dienstes optimiert werden. Mögliche zukünftige Betrachtungen sind zudem die informationslinguistische Tag-Bearbeitung wie beispielsweise die relativ einfache Generalisierung der Numeri oder die Identifizierung von Synonymen durch das Clustering. Auch die Bildung von Sub-Clustern ist ein interessanter Forschungspunkt, welcher ein differenzierteres hierarchisches Clustering und damit ein verbessertes Retrieval ermöglicht.
Auf Grund der Ergebnisse dieser und anderer Arbeiten kann festgehalten werden, dass die Betrachtung semantischer Strukturen innerhalb von Folksonomies einen wichtigen Ansatzpunkt bildet, um den Zugang zu den stetig wachsenden Ressourcen in Web2.0-Diensten zu gewährleisten und zu verbessern. Mit Tag-Cluster wurde eine visuelle Retrieval-Schnittstelle geschaffen, die diesen Aspekt aufgenommen hat und das Potential dieses Ansatzes erkennen lässt.
Der vorliegende Artikel ist eine deutsche, erweiterte Version eines Fachbeitrages der Hawaii International Conference on System Sciences (HICSS 2010).

ELSEVIER

Books on ScienceDirect

Die Elsevier Reference Works:
50% Sonderrabatt auf Reference Works
bis zum 31.12.2010

A broader perspective

- Agricultural and Biological Sciences
- Biochemistry, Genetics and Molecular Biology
- Chemistry
- Computer Science
- Earth and Planetary Sciences
- Energy
- Engineering
- Environmental Science
- Forensic
- Immunology and Microbiology
- Material Sciences
- Medicine and Dentistry
- Neuroscience
- Pharmacology, Toxicology and Pharmaceutical Science
- Physics and Astronomy
- Psychology
- Social Sciences

Weitere Informationen unter
http://info.sciencedirect.com/content/books/bookformats/Reference-works/
Bitte fordern Sie Ihr individuelles Angebot an unter: DACH@elsevier.com

Authoritative content
Fully linked with journals
Complete the picture in your subject area
www.sciencedirect.com

ScienceDirect™
makes sense.

Literaturverzeichnis

(1) Sinclair, J., & Cardew-Hall, M. (2008). The Folksonomy Tag Cloud: When Is It Useful? *Journal of Information Science, 34 (1)* , S. 15-29.

(2) Mathes, A. (2004). *Folksonomies - Cooperative Classification and Communication Through Shared Metadata.* Abgerufen am 25. Mai 2009 von adammathes.com: http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

(3) Dieberger, A., Dourish, P., Höök, K., Resnick, P., & Wexelblat, A. (2000). Social Navigation – Techniques for Building More Usable Systems. *Interactions 7* , S. 36-45.

(4) Rivadeneira, A. W., Gruen, D. M., Muller, M. J., & Millen, D. R. (2007). Getting Our Head In The Clouds: Toward Evaluation Studies of Tagclouds. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (S. 995–998). New York: ACM.

(5) Hearst, M. A., & Rosner, D. (2008). Tag clouds: Data analysis tool or social signaller? In *Proceedings of 41st Hawaii International Conference on System Sciences (HICSS 2008), Social Spaces minitrack.*

(6) Russell, T. (2006). Cloudalicious: Folksonomy Over Time. In *Proceedings of the 6th ACM/IEEEC-CS Joint Conference on Digital Libraries* (S. 364). New York: ACM.

(7) Halvey, M., & Keane, M. (2007). An Assessment of Tag Presentation Techniques. In *Proceedings of the 16th International Conference on World Wide Web* (S. 1313–1314).

(8) Bateman, S., Gutwin, C., & Nacenta, M. (2008). Seeing Things in the Clouds: the Effect of Visual Features on Tag Cloud Selections. In *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia, Pittsburgh, PA, USA* (S. 193–202). New York: ACM.

(9) Hassan-Montero, Y., & Herrero-Solana, V. (2006). Improving Tag-Clouds as Visual Information Retrieval Interfaces. In *Proceedings of International Conference on Multidisciplinary Information Sciences and Technologies (InSciT2006).* Mérida, Spanien.

(10) Begelman, G., Keller, P., & Smadja, F. (2006). Automated Tag Clustering: Improving Search and Exploration in the Tag Space. In *Proceedings of the WWW 2006 Collaborative Web Tagging Workshop at WWW2006.* Edinburgh, Scotland.

(11) Angeletou, S., Sabou, M., Specia, L., & Motta, E. (2007). Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report. In *Workshop: Bridging the Gap between Semantic Web and Web 2.0, European Semantic Web Conference* (S. 30–43).

(12) Zhang, L., Wu, X., & Yu, Y. (2006). Emergent Semantics from Folksonomies: A Quantitative Study. In *Journal on Data Semantics VI (Lecture Notes in Computer Science)* (S. 168–186). New York: Springer.

(13) Kaser, O., & Lemire, D. (2007). Tag-Cloud Drawing: Algorithms. In *Proc. WWW'07 Workshop on Taggings and Metadata for Social Information Organization.*

(14) Knuth, D. E., & Plass, M. F. (1981). Breaking Paragraphs Into Lines. *Software – Practice and Experience 11* , S. 1119–1184.

(15) Skiena, S. S. (1997). *The Algorithm Design Manual.* Berlin: Springer.

(16) Shaw, B. (December 2005). Utilizing Folksonomy: Similarity Metadata from the Del.icio.us System. Project Proposal. Abgerufen am 9. 5 2009 von http://www.metablake.com/webfolk/web-project.pdf

(17) Steinbach, M., Karypis, G., & Kumar, V. (2000). A Comparison of Document Clustering. In *Proceedings of KDD Workshop on Text Mining.*

(18) Cattuto, C., Schmitz, C., Baldassarri, A., Servedio, V. D., Loreto, V., Hotho, A., et al. (2007). Network Properties of Folksonomies. *AI Communications, 20 (4)*, S. 245–262.

(19) Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., & Stumme, G. (2009). Evaluating Similarity Measures for Emergent Semantics of Social Tagging. In *Proceedings of the 18th International Conference on World Wide Web.* New York: ACM.

(20) Li, R., Bao, S., Yu, Y., Fei, B., & Su, Z. (2007). Towards Effective Browsing of Large Scale Social Annotations. In *Proceedings of the 16th International Conference on World Wide Web* . New York: ACM.

(21) Huang, Y.-C., Hung, C.-C., & Hsu, J. Y. (2008). You Are What You Tag. In *Proceedings of AAAI 2008 Spring Symposium Series on Social Information Processing.*

(22) Schrammel, J., Leitner, M., & Tscheligi, M. (2009). Semantically Structured Tag Clouds: An Empirical Evaluation of Clustered Presentation Approaches . In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (S. 2037–2040).

(23) Capocci, A., & Caldarelli, G. (2008). Folksonomies and Clustering in the Collaborative System CiteULike. *Journal of Physics A-Mathematical and Theoretical* , S. 1–7.

(24) Rasmussen, E. (1992). Clustering Algorithms. In W. Frakes, & R. Baeza-Yates, *Information Retrieval: Data Structures and Algorithms* (S. 419–442). Upper Saddle River, NJ, USA : Prentice-Hall, Inc.

(25) Peters, I. (2009). Folksonomies: Indexing and Retrieval on the Web 2.0. In *Knowledge and Information. Studies in Information Science, Vol. 1.* München: Saur.

(26) Parasuraman, A., Zeithaml, V., & Berry, L. (1988). SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing, 64(1)*, S. 12–40.

(27) Buttle, F. (1996). SERVQUAL: Review, Critique, Research Agenda. *European Journal of Marketing 30(1)*, S. 8–32.

(28) Pitt, L., Watson, R., & Kavan, C. (1995). Service Quality: A Measure of Information Systems Effectiveness. *MIS Quarterly, 19(2)*, S. 173–187.

(29) Van Dyke, T. P., Kappelman, L., & Prybutok, V. (1997). Measuring Information Systems Service Quality: Concerns on the Ose of the SERVQUAL Questionnaire. *MIS Quarterly, 21(2)* , S. 195–208.

(30) Pitt, L., Watson, R., & Kavan, C. (1997). Measuring Information Systems Quality: Concerns for a Complete Canvas. *MIS Quarterly, 21(2)*, S. 209–221.

(31) Kettinger, W., & Lee, C. (1997). Pragmatic Perspectives on the Measurement of Information Systems Service Quality. *MIS Quarterly, 21(2)*, S. 223–240.

(32) Van Dyke, T. P., Prybutok, V. R., & Kappelman, L. A. (1999). Cautions on the Use of the SERQUAL Measure to Assess the Quality of Information Systems Services. *Decision Sciences, 30 (3)*, S. 1–15.

(33) Kettinger, W., & Lee, C. (1999). Replication of Measures in Information Systems Research: The Case of SERVQUAL. *Decision Sciences, 30(3)*, S. 893–899.

(34) Jiang, J., Klein, G., & Carr, C. (2002). Measuring Information System Service Quality: SERVQUAL from the Other Side. *MIS Quarterly, 26(2)*, S. 145–166.

(35) Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13* , S. 319–340.

(36) Kim, B., & Han, I. (2009). The Role of Trust Belief and its Antecedents in a Community-driven Knowledge Environment. *Journal of the American Society for Information Science and Technology, 60(5)*, S. 1012–1026.

(37) Lin, C., Wu, S., & Tsai, R. (2005). Integrating Perceived Playfulness Into Expectation-Conformation Model for Web Portal Context. *Information and Management, 42(5)*, S. 683–693.

(38) Van der Heyden, H. (2004). User Acceptance of Hedonic Information Systems. *MIS Quarterly, 28(4)*, S. 695–703.

(39) DeLone, W., & McLean, E. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research, 3* , S. 60–95.

(40) DeLone, W., & McLean, E. (2003). The DeLone and McLean Model of Information Systems Success: A ten-year Update. *Journal of Management Information Systems, 19(4)*, S. 9–30.

(41) Jennex, M., & Olfman, L. (2006). A Model of Knowledge Management Success. *International Journal of Knowledge Management, 2(3)*, S. 51–68.

AUTORIN

Kathrin Knautz, B.A.
(Informationswissenschaft & Sprachtechnologie)
Wissenschaftliche Hilfskraft
(und Studentin im Master-Studium Informationswissenschaft & Sprachtechnologie)
Abteilung Informationswissenschaft
Institut für Sprache und Information
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1
40225 Düsseldorf
Kathrin.Knautz@uni-duesseldorf.de

Wohin mit Ihren aussortierten Altbeständen?

Wir bieten ein Vermarktungskonzept mit „Rundumservice“: Bundesweite Abholung, Lagerung, Verkauf und Versand.

Etat erhöhen – ohne Aufwand und Kosten für Sie.

getbooks.de GmbH, Edgar Hardt, Max-Planck-Str. 21a, 65520 Bad Camberg, Fon: 06434-905390, ehardt@getbooks.de

Der „Check full text"– Service der Bibliothek der Medizinischen Fakultät Mannheim

Mit „ReDI>>Links" zu einem verbesserten LinkResolver

Nina Waldkirch

Um die tägliche Recherche im Informationsdschungel zu erleichtern, bietet die „Bibliothek der Medizinischen Fakultät Mannheim der Universität Heidelberg" ihren Nutzern seit 2005 einen LinkResolver an. Aufgrund zunehmender Probleme mit dem bis dahin eingesetzten Produkt erfolgte Anfang des Jahres 2010 der Umstieg auf „ReDI>>Links" von ReDI. Der vorliegende Artikel beschreibt die Funktionsweise von ReDIs LinkResolver und zeigt dessen technische Möglichkeiten auf. Dabei werden Vergleiche zum ehemals eingesetzten Produkt gezogen. Schließlich wird die konkrete Umsetzung an der „Bibliothek der Medizinischen Fakultät Mannheim" erläutert und aufgezeigt, wie „ReDI>>Links" an unserer Einrichtung zur Verbesserung der LinkResolver-Akzeptanz und Einsparung von Arbeitsaufwand, Zeit und Kosten beiträgt.

The „Check full text"-service of the Library of the Medical Faculty of Mannheim – With „ReDI>>Links" to an improved LinkResolver

To facilitate the daily research in the „information jungle", the „Library of the Medical Faculty of Mannheim of the University of Heidelberg" has been offering a linkresolver to its users since 2005. Increasing problems with the heretofore utilized product led us to switch to „ReDI>>Links" of ReDI at the beginning of the year 2010. The present article describes the functionality of ReDI's linkresolver, demonstrates its technical options and draws comparisons to the product formerly used. Finally the article gives an overview of the implementation at the „Library of the Medical Faculty of Mannheim" and demonstrates how „ReDI>>Links" contributes to a greater acceptance of linkresolver and to the reduction of labour input, time and costs at our institution.

Le « Check full text »-service de la Bibliothèque de la Faculté de Médecine de Mannheim – Avec des « ReDI>>Links » pour un service amélioré de notre LinkResolver

Pour rendre plus facile la recherche quotidienne dans la jungle de l'information la Bibliothèque de la Faculté de Médecine Mannheim de l'Université Heidelberg offre à ses utilisateurs depuis l'année 2005 le service « link resolver ».
A cause de plusieurs difficultés avec le programme que nous avons utilisé avant, nous avons changé le produit – maintenant (depuis 2010) nous utilisons « ReDI>>Links » de ReDI.
Cet article caractérise le fonctionnement du produit « ReDI>>Links » et explique ses possibilités techniques en faisant la comparaison avec le produit initial.
Finalement la réalisation concrète à la Bibliothèque de la Faculté de Médecine Mannheim de l'Université Heidelberg sera expliquée et nous montrerons, comment « ReDI>>Links » contribue à l'amélioration de l'acceptation du LinkResolver et comment cela aide à économiser du temps, du travail et à épargner de l'argent.

■ Im multimedialen Zeitalter springen Bibliotheksnutzer auf der Suche nach Artikeln, Fachzeitschriften und anderen Medien zwischen verschiedenen Datenbanken, lokalen Bestandskatalogen und Bestellformularen diverser Online-Literaturlieferdienste hin und her. Um die tägliche Recherche im Informationsdschungel zu erleichtern, bietet die Bibliothek der Medizinischen Fakultät Mannheim der Universität Heidelberg ihren Nutzern bereits seit Januar 2005 einen LinkResolver an. Der Service ermöglicht die Verknüpfung von Rechercheergebnissen mit den gedruckten und elektronischen Zeitschriftenbeständen und Dienstleistungen der Bibliothek.

Zum besseren Verständnis seien zunächst ein paar Worte zur Organisations- und EDV-Situation unserer Bibliothek vorangestellt: Die Bibliothek der Medizinischen Fakultät Mannheim ist Teil des Bibliothekssystems der Universität Heidelberg, ist jedoch auf dem Campus der Universitätsmedizin Mannheim (UMM) lokalisiert. Da sich auf dem Campus sowohl die Medizinische Fakultät Mannheim als auch das Universitätsklinikum befinden, existieren dort auch die voneinander getrennten IT-Netzwerke beider Einrichtungen. Die Bibliothek der Medizinischen Fakultät Mannheim befindet sich im Fakultätsnetz, was bedeutet, dass auch unsere Online-Services ausschließlich im Fakultäts- und nicht im Klinikumsnetz nutzbar sind. Den externen Zugriff erleichtern wir unseren Benutzern mittels eines so genannten VPN-Clienten, dessen Software ein Einwählen (mittels der individuellen Bibliothekskennung) in das Fakultätsnetz ermöglicht. Da die Bibliothek der Medizinischen Fakultät Mannheim als eigenständiger Standort – unabhängig von der Universitätsbibliothek Heidelberg – fungiert, gibt es für Mannheimer und für Heidelberger Nutzer auch unterschiedliche Bibliotheksausweise. Das Login für den VPN-Clienten erfolgt daher über den Mannheimer Benutzungsausweis. Auf diese Weise erfolgt auch der externe Zugriff auf unseren LinkResolver.

Seit 2005 war das Produkt „LinkSource" der Firma EBSCO im Einsatz, ab Anfang 2007 wurde dieses jedoch durch Ovids „LinkSolver" abgelöst. Doch die alltägliche Erfahrung, jüngste Umfragen und Rückmeldungen an der Auskunft zeigten, dass die Funktionsweise des „LinkSolver" bei uns fehlerhaft, eingeschränkt und damit nicht zufriedenstellend war. Aus diesem Grund wurde Ende des Jahres 2009 der Entschluss gefasst, auf ein anderes System umzustellen: „ReDI>>Links" von ReDI[1] dient nun seit Anfang 2010 als neuer LinkResolver der Bibliothek der Medizinischen Fakultät Mannheim und ist an unserer Einrichtung unter dem Namen „Check full text"-Service bekannt.

1 Regionale Datenbank-Information Baden-Württemberg (http://www-fr.redi-bw.de/)

Funktionsweise des „Check full text"-Services

Der „Check full text"-Service ist in den Datenbanken PubMed, MedPilot, Web of Science, der Cochrane Library[2], bei den E-Journals des Aggregators EBSCOHost, sowie im Literaturverwaltungsprogramm EndNote integriert.
Nach einer Recherche in PubMed erscheint in der Abstract-Ansicht rechts oben, nach einer Suchanfrage in MedPilot in der Kurztrefferliste unterhalb des Artikels unser „Check full text"-Button[3]:

Nach dessen Betätigung werden die Metadaten des gewünschten Aufsatzes an den LinkResolver übertragen und gleichzeitig unsere lokalen Bestandsinformationen abgefragt. Für Letzteres dienen sowohl der Bestand der Elektronischen Zeitschriftenbibliothek Heidelberg als auch der Katalog des Bibliothekssystems der Universität Heidelberg als Datenbasis. In einem neuen Browserfenster wird dem Nutzer das Ergebnis mitgeteilt:
Sofern eine Lizenz vorhanden ist (Abb. 1), wird unter dem Menüpunkt *Volltext* ein Link angeboten, der in der Regel direkt zum elektronischen Volltext des Aufsatzes führt. Bei mehreren Volltext-Links findet ein Ranking statt: lizenzierte Titel erscheinen vor frei verfügbaren, unter gleichwertigen Links findet ein alphabetisches Ranking statt, Links zu Titeln in Aggregator-Datenbanken (z.B. Academic Search Premier via EBSCOHost) werden generell niedriger eingestuft, da hier in der Regel Embargos den Lizenzzeitraum einschränken. So wählt der Nutzer mit dem ersten Link auch stets den besten.
Außerdem wird unter dem Menüpunkt *Journal Citation Reports* ein weiterer Link angezeigt, der zum Impact Factor der Zeitschrift führt. Dabei wird auf die „Journal Citation Reports" im „ISI Web of Knowledge" zugegriffen, mit deren Hilfe sich Impact Faktoren über einen Zeitraum von bis zu fünf Jahren berechnen lassen.

Abbildung 1: „ReDI>>Links"-Fenster Volltext vorhanden

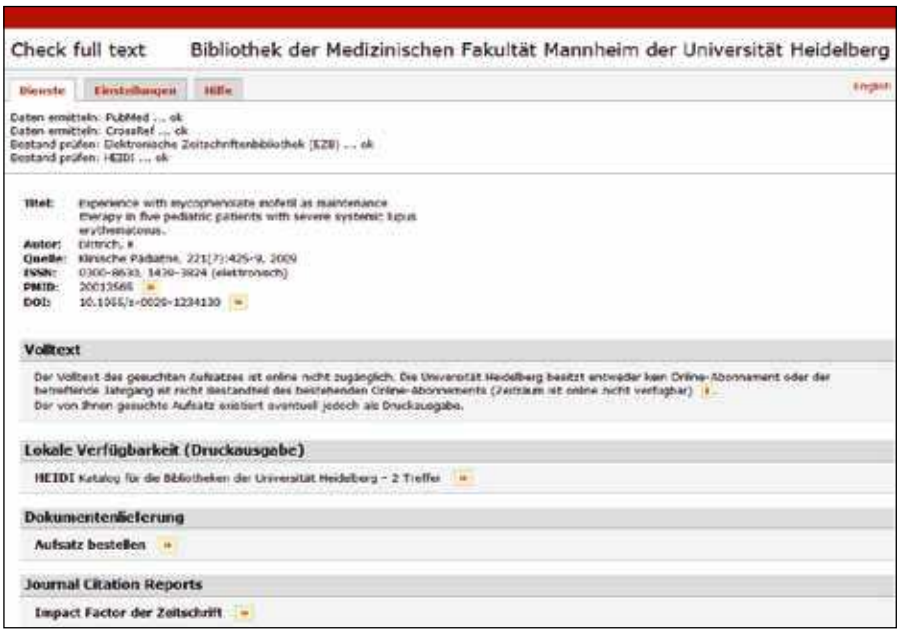


Abbildung 2: „ReDI>>Links"-Fenster Volltext nicht vorhanden

Ist keine Lizenz für den Volltext vorhanden (Abb. 2), wird durch Anklicken des Menüpunktes *Lokale Verfügbarkeit (Druckausgabe)* im lokalen Bestandskatalog des Bibliothekssystems der Universität Heidelberg „HEIDI" eine automatische Suche nach der Druckausgabe der Zeitschrift durchgeführt. Diese kann dann gegebenenfalls vor Ort eingesehen und kopiert werden.
Hat die Recherche in „HEIDI" keine Treffer ergeben (bereits im LinkResolver-Fenster wird der Nutzer über das Rechercheergebnis informiert), oder ist unsere Bibliothek nicht im Besitz der Zeitschrift bzw. des gesuchten Jahrgangs, ermöglicht der Link *Aufsatz bestellen unter Dokumentlieferung* die automatische Übernahme der bibliographischen Daten in das Bestellformular unseres Literatur-Liefer-Dienstes. Zum Absenden der Bestellung müssen lediglich noch Bibliotheksausweisnummer und Passwort eingetragen und eine Kostenobergrenze festgelegt werden.
Zusätzlich wird dem Nutzer auch hier wieder der Link zum Impact Factor der Zeitschrift angeboten.

Layoutgestaltung und Datenbasis

Sowohl der Button als auch das LinkResolver-Fenster können individuell angepasst werden:

- Bei der Layoutgestaltung des Seitenkopfes setzt ReDI keine Grenzen, für Ovids „LinkSolver" hatten wir diesbezüglich nur eingeschränkte Möglichkeiten.
- Texte, angebotene Dienste bzw. Menüpunkte und die jeweiligen Links können beliebig betitelt und sortiert werden. Dies war uns mit dem bisherigen Produkt nicht möglich.

2 Der Service ist in den Cochrane-Datenbanken „Database of Abstracts of Reviews of Effects", „Cochrane Central Register of Controlled Trials", „Cochrane Methodology Register" und „NHS Economic Evaluation Database" verlinkt. In den Datenbanken „Cochrane Database of Systematic Reviews", „Health Technology Assessment Database" und „About The Cochrane Collaboration" ist die Einbindung eines LinkResolvers nicht notwendig, da hier die Volltexte unmittelbar zur Verfügung stehen.

3 Da die anderen oben erwähnten Datenbanken nicht zwischen der Universitätsbibliothek Heidelberg und der zum Bibliothekssystem der Universität Heidelberg gehörenden Bibliothek der Medizinischen Fakultät Mannheim unterscheiden können, ist es in diesen Fällen nur möglich, einen einzigen LinkResolver-Button zu hinterlegen. Angezeigt wird daher dort der Button der UB Heidelberg, der nach Anklicken jedoch ebenfalls auf unseren „Check full text"-Service verlinkt.
In den genannten Datenbanken der Cochrane Library ist ein eigens von Wiley erstellter LinkResolver integriert, der jedoch durch Einbindung des OpenURL-Links „Check full text MEDMA" auch in diesem Fall auf unseren „Check full text"-Service verweist.

OBID® – RFID by FEIG ELECTRONIC

OBID®
Welcome to RFID

RFID-Gate-Antennen für Bibliotheken

Perfektion in Design, Funktion und Service.

reddot design award
winner 2010

Designpreis Deutschland 2011
NOMINIERT

FEIG ELECTRONIC GmbH
Lange Straße 4 · D-35781 Weilburg
Phone: +49 6471 3109-0
Fax: +49 6471 3109-99 · www.feig.de

FEIG ELECTRONIC

- Wie auch für Ovids „LinkSolver" können Regeln für die Anzeige der Links festgelegt werden.
- Eine Sprachauswahl ermöglicht unseren Nutzern ab sofort die Entscheidung, den Service in deutscher oder in englischer Sprache in Anspruch zu nehmen. Bei Ovids „LinkSolver" mussten wir uns für eine Sprache entscheiden.
- Unter dem Reiter *Hilfe* sind allgemeine Hinweise zum „Check full text"-Service und mögliche Fehlermeldungen zur Erläuterung verlinkt. Die Einbindung einer Hilfefunktion war uns mit dem bisherigen Produkt nicht möglich.
- Darüber hinaus werden im LinkResolver-Fenster mehr bibliographische Angaben angezeigt: hier werden sowohl P-ISSN, als auch E-ISSN, zusätzlich die DOI und – im Falle einer PubMed-Recherche – die PMID angezeigt. Der „LinkSolver" übertrug lediglich eine ISSN.

Zum Vergleich siehe auch Abbildungen 3 und 4.

Einer unserer Hauptgründe für den Umstieg des Ovid-LinkResolvers waren jedoch Verlinkungsprobleme und fehlerhafte Metadatenübertragungen. Die Datenbasis des „LinkSolver" bildet eine zentrale Datenbank bei Ovid, die von den dortigen Support-Mitarbeitern gepflegt wird. Jeder Ovid-Kunde verfügt zwar über ein Administrations-Tool, in dem die lokalen Lizenzinformationen in Form von ISSN-Listen – eingebunden in diversen Anbietergruppen – hinterlegt werden, die zugehörigen Metadaten eines jeden Artikels müssen jedoch in der zentralen Ovid-Datenbank nachgewiesen sein. Andernfalls erhält der „LinkSolver" keine Ergebnisse. Speziell bei Epubs ahead of Print und aktuellen Artikeln – und gerade diese sind im Medizinbereich diejenigen Publikationen von größtem Interesse – führten (noch) fehlende Metadaten dazu, dass in der zentralen Ovid-Datenbank keine Treffer gefunden wurden und somit – trotz vorhandener Lizenz – kein Volltext-Link angeboten wurde. Auf diese Weise erhielten unsere Nutzer falsche Bestandsinformationen, wodurch die zuverlässige Funktionsweise des LinkResolver-Services angezweifelt wurde und der Mehrwert des Dienstes verloren ging: „Ich benutze den LinkSolver nicht, weil er sowieso meistens nicht funktioniert.", war eine häufige Rückmeldung an der Auskunft. Fehlerhafte Übertragungen von Metadaten führten nicht selten auch zu Komplikationen für unser Dokumentlieferteam. So gingen Bestellungen für vermeintlich nicht online zugängliche Zeitschriftenartikel ein, obwohl eine Zugriffsberechtigung vorlag oder die eingehenden Bestellungen enthielten lückenhafte bibliographische Angaben, wodurch das Dokumentlieferteam die Daten erneut recherchieren musste.

Diese Probleme konnten mit dem Umstieg auf „ReDI>>Links" glücklicherweise behoben werden, da hier bei der Bestandabfrage über den „Check full text"-Service weitaus weniger Metadaten abgefragt werden: So reicht in der Datenbank PubMed lediglich die PMID zur vollständigen Übertragung der Artikeldaten aus, wodurch auch ohne Volume-, Issue-Angaben die Volltexte zu Epubs ahead of Print und zu aktuellen Artikeln gefunden werden. Sollte es dennoch vorkommen, dass die vom Anbieter hinterlegten Metadaten des gesuchten Artikels unzureichend sind, erhält der Benutzer einen entsprechenden Hinweis.

Da als Datenbasis der E-Journals-Bestand der Elektronischen Zeitschriftenbibliothek (EZB) Heidelberg – im Vergleich zu einer eigenen Admin-Oberfläche zur Verwaltung der Bestände bei Ovid – dient und daher das regelmäßige Pflegen und Generieren von Bestandslisten entfällt, besteht bei „ReDI>>Links" weniger Arbeitsaufwand. Die Zahl der Problemfälle konnte erheblich reduziert werden, wodurch erneut Arbeitszeit gewonnen werden konnte. Zudem konnten Kosten eingespart werden, da ReDIs LinkResolver vollkommen gebührenfrei und ohne Abschluss eines Lizenzvertrages genutzt werden kann.

Selbstverständlich bietet ein kommerzieller, datenbankbasierter LinkResolver auch einige Vorteile, wie z.B. das Angebot einer Statistikfunktion, häufigere artikelgenaue Verlinkung[4] durch eine größere Titelabdeckung über CrossRef[5] und präzisere Standortunter-

Abbildung 3: „LinkSolver"-Fenster Volltext vorhanden

Abbildung 4: „LinkSolver"-Fenster Volltext nicht vorhanden

4 Artikelgenaue Verlinkung heißt, dass man nach Anklicken des Volltext-Links im LinkResolver direkt zur HTML- oder PDF-Version des gewünschten Artikels weitergeleitet wird. Ist keine artikelgenaue Verlinkung gegeben, wird man zunächst auf die Startseite der jeweiligen Zeitschrift geführt, von wo aus man den gesuchten Aufsatz erneut im Online-Archiv suchen muss.

5 CrossRef (http://www.crossref.org/) ist eine gemeinnützige Verlinkungsvereinigung, die Digital Object Identifier (DOIs) für wissenschaftliche Veröffentlichungen vergibt, um deren Metadaten im Web besser zu lokalisieren. Jeder Verlag, bzw. Anbieter, der Mitglied von CrossRef ist (für eine aktuelle Übersicht siehe: http://www.crossref.org/01company/17crossref_members.html), hinterlegt dort die Metadaten seiner Publikationen und hat danach automatisch Zugriff auf alle im CrossRef-Datenpool hinterlegten DOIs, welche eine artikelgenaue Verlinkung gewährleisten.

scheidung der frei geschalteten Lizenzen.[6] Diese Einschränkung nahmen wir jedoch gerne in Kauf, da insgesamt die Vorteile von „ReDI>>Links“ deutlich überwiegen.

Umsetzung

Bei der Umsetzung stand uns der zuständige ReDI-Mitarbeiter an der UB Freiburg via Telefon und E-Mail stets mit Rat und Tat zur Seite. Der Umstieg nahm nur wenige Wochen in Anspruch. Hierfür wurde für den Mannheimer Standort zunächst eine Testsicht generiert, mit deren Hilfe wir verschiedene Szenarien ausprobieren, uns von der Funktionsweise überzeugen und direkte Vergleiche zu unserem – zum damaligen Zeitpunkt – noch aktiven LinkSolver ziehen konnten. Layoutvorschläge wurden dem Supporter mitgeteilt und zu unserer vollsten Zufriedenheit umgesetzt. Alternativ hat man als ReDI-Kunde auch die Möglichkeit, aufwendigere Gestaltungswünsche in einer eigenen css-Datei zu hinterlegen, dies war bei uns allerdings nicht notwendig.

Der Service ist innerhalb des Fakultätsnetzes nutzbar. Obwohl die Bibliothek der Medizinischen Fakultät Mannheim zum Bibliothekssystem der Universität Heidelberg gehört, konnte eine eigene Mannheimer Sicht – unabhängig von Heidelberg – realisiert werden. Dies war daher sinnvoll, da sich einige unserer Layout-Festlegungen von denen des Heidelberger LinkResolvers unterscheiden und darüber hinaus teilweise unterschiedliche Lizenzabschlüsse bestehen. Aus diesem Grund erstellte uns der ReDI-Support-Mitarbeiter für den Echtbetrieb eine eigene Basis-URL[7] und eine URL für den Mannheimer „Check full text“-Button[8], welche von uns in die jeweiligen Sources eingebunden wurden. Dazu mussten wir diese lediglich bei den unterschiedlichen Datenbanken den jeweiligen Ansprechpartnern per E-Mail mitteilen.

Bei PubMed erfolgt der Aufruf der Datenbank über eine spezielle URL; das von der National Library of Medicine (NLM) standardisierte Testverfahren zum Nachweis qualifizierter Ressourcen wurde von uns bereits 2007 bei der Einführung des „LinkSolver“ absolviert und musste daher für den Umstieg auf ReDI nicht wiederholt werden. In der „Find Full Text“-Maske des EndNote-Programms wird die neue Base-URL von jedem Nutzer manuell eingetragen.

Die Regelung der externen Authentifizierung für den Service gestaltet sich aufgrund getrennter Netzwerke auf dem Campus und getrennter Bibliotheksausweise innerhalb des Bibliothekssystems Heidelberg als besondere Herausforderung. Für Ovids „LinkSolver“ wurde die Sachlage zusätzlich verkompliziert, da im Falle des externen Zugriffs eine automatisierte Seite die Nutzer zum Ovid-Login aufforderte, was häufig zu Verwirrungen führte (siehe Link oben rechts in Abb. 3 und 4), da eine solche Identifikation nicht vorgesehen war und unseren Nutzern daher auch keine entsprechenden Login-Daten zur Verfügung standen. „ReDI>>Links“ bietet uns nun jedoch die Möglichkeit, bei externem Zugriff eine individuelle Informationsseite anzeigen zu lassen, die auf den VPN-Client als Zugangsweg verweist und Verlinkungen zu den entsprechenden Informationen auf unseren Webseiten anbietet.

Öffentlichkeitsarbeit

Der neue „Check full text“-Service wurde auf unseren Webseiten, über unseren Newsblog und Newsletter, über einen Flyer und über eine Power-Point-Präsentation, die auf einem zentralen Wandbildschirm im Benutzungsbereich über die Bibliothek informiert, beworben. Zusätzlich wurde ein Film[9] zur Nutzung des „Check full text“-Service kreiert, welcher ebenfalls auf unserer Website verlinkt ist.

Fazit

Mit dem neuen LinkResolver von ReDI konnten wir durch den Wegfall des regelmäßigen Generierens und Pflegens von Bestandslisten und die deutliche Reduzierung von Reklamationen – sowohl auf Seiten des Dokumentlieferteams, als auch auf Seiten der LinkResolver-Sachbearbeiterin – Aufwand, Zeit und durch das kostenlose Angebot von „ReDI>>Links“ schließlich auch Geld einsparen. Gleichzeitig konnte die Funktionsweise des „Check full text“-Service nachweislich verbessert und der Mehrwert des LinkResolver-Dienstes wiederhergestellt werden. Der Umstieg auf das neue Produkt verlief für uns also durchweg positiv. Aufschlüsse über die Akzeptanz des neuen „Check full text“-Services werden wir im Rahmen der nächsten Benutzerumfrage erhalten.

■ AUTORIN

Dipl.-Bibl. Nina Waldkirch
Online-Services und Schulungen
Bibliothek der Medizinischen Fakultät Mannheim der Universität Heidelberg
Theodor-Kutzer-Ufer 1-3
68167 Mannheim
nina.waldkirch@medma.uni-heidelberg.de

6 Durch die Verwaltung unserer lokalen Bestände über die Admin-Oberfläche von Ovid, gab es dort technisch keinerlei Probleme, unsere Lizenzen als die einer eigenständigen Einrichtung – losgelöst von Heidelberg – zu erkennen. Daher fand die Unterscheidung der von Mannheim und der von Heidelberg abgeschlossenen E-Journals-Verträge hier auf Volume-, Issue-Ebene statt, d.h., unterschiedliche Lizenzzeiträume von Mannheim und Heidelberg bei ein und derselben Zeitschrift wurden vom LinkResolver korrekt erkannt. Bei ReDI kann die Standortunterscheidung aufgrund der gemeinsamen Datenbasis EZB Heidelberg nur auf Zeitschriftenebene stattfinden, d.h., der LinkResolver kann zwar Lizenzabschlüsse für unterschiedliche Zeitschriften, nicht aber für verschiedene Lizenzzeiträume einer Zeitschrift trennen.

7 http://www.redi-bw.de/links/unihdma

8 http://www.redi-bw.de/links/gifs/unihdma_icon.gif

9 http://www.ma.uni-heidelberg.de/bibl/schulungen/checkfulltext/checkifulltext.html

www.b-i-t-online.de

B.I.T.online präsentiert aus der Praxis für die Praxis. Ein Buch zur Qualitätsentwicklung und Qualitätssicherung für eine ausgezeichnete Bibliothek

Gut ist uns nie gut genug!

Herausgegeben von Tom Becker und Cornelia Vonhof

NEU

Das Streben nach Perfektion – nach einer ‚ausgezeichneten' Bibliothek – ist der Leitfaden dieses Buches. *„Gut ist uns nie gut genug!"* ist der Anspruch, der uns mit Blick auf die Praxis dazu verführen soll, nie stehen zu bleiben und uns als kundenorientierter Dienstleistungsbetrieb immer wieder aufs Neue einem zielgerichteten kontinuierlichen Verbesserungsprozess freiwillig zu unterwerfen.
In den Aufsätzen werden aus unterschiedlichen Perspektiven Ideen, Konzepte, Methoden und Instrumente für ein passgenaues Qualitätsmanagement in Bibliotheken aufgezeigt.

Ein kleiner Auszug der Themen des Buches …
Die ausgezeichnete Bibliothek ▪ Erfolgreiches Innovationsmanagement ▪ Ideen fallen nicht vom Himmel ▪ Evaluation und Qualitätssicherung ▪ Leinen los! ▪ Das Trouble-Ticketsystem ▪ Wandel im Quadrat ▪ Die Portfolio-Analyse zur Profilierung ▪ Lebensstilanalyse ▪ Profilbildung und Zielgruppenanalyse ▪ Informationslogistische Rollen

… und der mitwirkenden Autoren
Tom Becker & Andrea Born, Tom Becker, Ute Engelkenmeier, Ursula Georgy, Anette Hagenau, Petra Häuslbauer, Jens Ilg, Katharina Lück, Frauke Schade, Martin Szlatki, Cornelia Vonhof

ISBN 978-3-934997-33-2, 2010
Brosch., 354 Seiten, € 29,50 zzgl. Versandkosten
Dinges & Frick Verlag Wiesbaden
Bestellung auf www.b-i-t-online.de oder einfach per Mail an dfverlag@dinges-frick.de

Linked Data – und warum wir uns im hbz-Verbund damit beschäftigen!

Adrian Pohl und Felix Ostrowski

Der Themenkomplex „Linked Open Data" findet zunehmend Einzug in die bibliothekarische Sphäre. Während „Open Data" sich auf die rechtliche Situtation von Daten aller Couleur bezieht, ist es das Ziel von „Linked Data", die Interoperabilität von Daten durch einheitliche (Web-) Standards zu verbessern. In diesem Artikel wird zunächst erläutert, warum „Linked Data" auch für die Bibliothekswelt relevant ist. Daraufhin werden in einer knappen Einführung die wichtigsten Linked Data-Standards vorgestellt sowie einige Gedanken zur Migration von bibliographischen Daten in eben diese Standards skizziert. Zuletzt werden die derzeitigen diesbezüglichen Aktivitäten im hbz-Verbund dargestellt.

Linked Data at the hbz

The topic „Linked Open Data" is pushing more and more into the library sphere. While „Open Data" is all about the legal status of any kind of data, „Linked Data" aims at augmenting the interoperability of data by consistently applying (web-)standards. This article begins with an illustration of why „Linked Data" is relevant to the library world. This is followed by a short introduction to the most important Linked Data standards and some thoughts on how to migrate bibliographic data to these standards. Finally the current activities of the hbz library network concerning this matter are presented.

« Linked Data » – et pourquoi nous nous penchons sur le réseau hbz !

Le sujet des « Linked Open Data » fait de plus en plus son chemin dans la sphère des bibliothèques. Alors que l'expression « Open Data » a trait au statut juridique de tout type de données, « Linked Data » a pour but d'augmenter l'interopérabilité des données en appliquant en permanence les standards (du Web). Cet article commence par illustrer pourquoi « Linked Data » est aussi pertinent pour le monde des bibliothèques. Suivent une courte introduction des plus importants standards de Linked Data et quelques réflexions sur la manière de faire passer les données bibliographiques à ces standards. Enfin, les activités actuelles, à cet égard, du réseau de bibliothèques hbz, sont présentées.

Vorwort

Mit dem Internet und – als dessen wichtigstem Bestandteil – dem World Wide Web formt sich seit einigen Jahrzehnten eine umfassende Publikations- und Kommunikationsplattform aus, auf der zukünftig der Großteil aller Publikation und Kommunikation stattfinden wird. Als eine Erweiterung des bestehenden Webs lässt sich Linked Open Data verstehen. Mit Linked Open Data werden zwei Standards bezeichnet, die die Funktionalität eines Netzes von Daten sichern sollen, indem sie die rechtliche und technische Kompatibilität von Daten im Web garantieren:

- *Open-Data-Standards* sorgen für die *rechtliche* Basis der Nutzung und Kombination verteilter Daten im Netz.
- *Linked-Data-Standards* sorgen für die *technische* Kompatibilität zwischen verteilt vorliegenden Daten.

In einer dreiteiligen Artikelreihe über Linked-Open-Data-Aktivitäten am Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) sollen die rechtliche wie technische Dimension von Linked Open Data erläutert werden und die Notwendigkeit, die Ziele und der Nutzen von Linked Open *Bibliographic* Data[1] dargelegt werden. Im ersten Teil dieser Reihe über das Was, Warum und Wie von Linked-Open-Data-Aktivitäten am hbz sollen einige Fragen zu *Open* Data geklärt werden. Er erscheint gedruckt in ProLibris 3/2010. Der zweite Teil – gemeinsam verfasst von Felix Ostrowski und Adrian Pohl – mit dem Schwerpunkt *Linked* Data erscheint gedruckt in B.I.T. online 3/2010 und der dritte, in dem sich ebenfalls Felix Ostrowski und Adrian Pohl mit der Konzeption und Implementierung einer Open-Data-Plattform befassen, wird gedruckt im Tagungsband der DGI-Konferenz *Semantic Web & Linked Data Elemente zukünftiger Informationsinfrastrukturen* publiziert.

Alle Texte werden darüber hinaus unter einer CC-BY-Lizenz im Web publiziert, siehe etwa unter http://www.hbz-nrw.de/dokumentencenter/produkte/lod/

1 Einleitung

Im März 2010 haben Kölner Bibliotheken und das Landesbibliothekszentrum Rheinland-Pfalz in Kooperation mit dem Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) ihre Katalogrohdaten der Öffentlichkeit als Ganzes unter einer Public-Domain-Lizenz zur freien Verfügung gestellt.[2] Um die Motivation und Ziele hinter diesem Schritt sowie dessen Umsetzung zu beleuchten, fand im ersten Teil dieser Reihe eine vorwiegend (lizenz-)rechtliche Betrachtung von Open Data statt.[3] Da es erklärtes Ziel ist, die Daten mittelfristig in Linked Data zu überführen soll in diesem Beitrag nun die technische Dimension von Linked Open Data, eben das Was, Warum und Wie von *Linked* Data erläutert werden.

1 Die Ausdrücke „Freie Katalogdaten" und „Open Bibliographic Data" werden in diesem Kontext synonym verwendet.

2 Siehe die Pressemitteilung Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (2010): Freigabe der Katalogdaten: Kölner Bibliotheken leisten Pionierarbeit. Einsehbar unter: http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe

3 Pohl, Adrian (2010): Open Data im hbz-Verbund. Erscheint in ProLibris 3/2010. Preprint u.a. einsehbar unter http://www.hbz-nrw.de/dokumentencenter/produkte/lod/

2 Semantic Web und Linked Data

Was steckt nun hinter ‚Linked Data'?[4] Warum beschäftigen sich Bibliothekarinnen und Bibliothekare, Bibliotheken und Verbünde damit?

Linked Data ist der erste Schritt zum Semantic Web. Die Entwicklung des Semantic Web ist ein langfristiges Projekt, vorangetrieben vom Erfinder des World Wide Web Tim Berners-Lee[5] und dem W3C (World Wide Web Consortium). Während ein großer Teil des World Wide Web aus Texten besteht, die sich an *Menschen* richten, ist das Semantic Web der Versuch, Informationen so im Web bereitzustellen, dass sie von *Maschinen* leicht verarbeitet werden können, um nützliche Dinge zu tun. Mit anderen Worten: Das Semantic Web ist die Erweiterung des World Wide Webs als eines Netzes von Dokumenten zu einem Netz von Daten, von maschinenlesbaren, einzelnen Aussagen. Durch maschinelles logisches Schlussfolgern soll aus diesen Daten zusätzliches Wissen generiert werden können.

Auf den Seiten des W3C findet sich folgende Definition des Semantic Web:

"The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF).[6]"

Die Vision des Semantic Web könnte so formuliert werden: Im Web werden immer mehr veröffentlichte Daten in RDF (s.u.) vorliegen, d.h. sie folgen denselben Standards und können daher frei miteinander kombiniert und verlinkt werden. So entsteht das Semantic Web als eine verteilte Datenbank, die komplexe Abfragen und Schlussfolgerungen über dezentral vorliegende Daten ermöglicht.

Linked Data ist der erste Schritt und das Herzstück des Semantic Webs. Auf den Webseiten des W3C heißt es:

"[T]o make the Web of Data a reality, it is important to have the huge amount of data on the Web available in a standard format, reachable and managable by Semantic Web tools. Furthermore, not only does the Semantic Web need access to data, but *relationships among data* should be made available, too, to create a *Web* of Data (as opposed to a shere collection of datasets). This collection of interrelated datasets on the Web can also be referred to as Linked Data.[7]"

Linked Data legt also das Augenmerk auf die ersten Schritte, auf das Veröffentlichen und Verlinken von Daten, während im Semantic Web Anwendungen, die auf maschinell durchführbaren logischen Schlussfolgerungen basieren, die entscheidende Rolle spielen. Es ist aber zu konstatieren, dass im gegenwärtigen Sprachgebrauch ‚Semantic Web' und ‚Linked Data' häufig synonym verwendet werden.

3 Warum Linked Library Data?

Warum sollen Bibliotheken ihre Daten als Linked Data publizieren? Im Prinzip verlinken Bibliothekare Daten spätestens seit der Etablierung des Kartenkatalogs, indem sie etwa Verweise von Schlagwörtern und Personen auf Titeldaten dokumentieren. Bibliotheken produzieren also schon lange verlinkte Daten. Außerdem sind sie vertrauenswürdige, öffentliche Institutionen, die tief in unserer Kultur verwurzelt sind und Vertrauenswürdigkeit ist eine wichtige Eigenschaft, die Linked Open Data haben sollte. Bibliotheken sind also geradezu prädestiniert dazu, das Linked-Data-Netz mit aufzubauen, zu pflegen und seine Vertrauenswürdigkeit und Stabilität zu sichern.

Ein weiterer Grund, Bibliotheksdaten in das Semantic Web zu migrieren, sind die stattfindenden Veränderungen in der Produktion, Publikation und Verbreitung wissenschaftlicher Ergebnisse, seien dies Daten oder Texte. Semantisches Markup und Verlinkung werden eine wichtige Rolle in der Weiterentwicklung der Wissensproduktion spielen und damit großen Einfluss auch auf Wissenschaftliche Bibliotheken haben.[8] Dies macht deutlich, dass Semantic-Web-Tech-

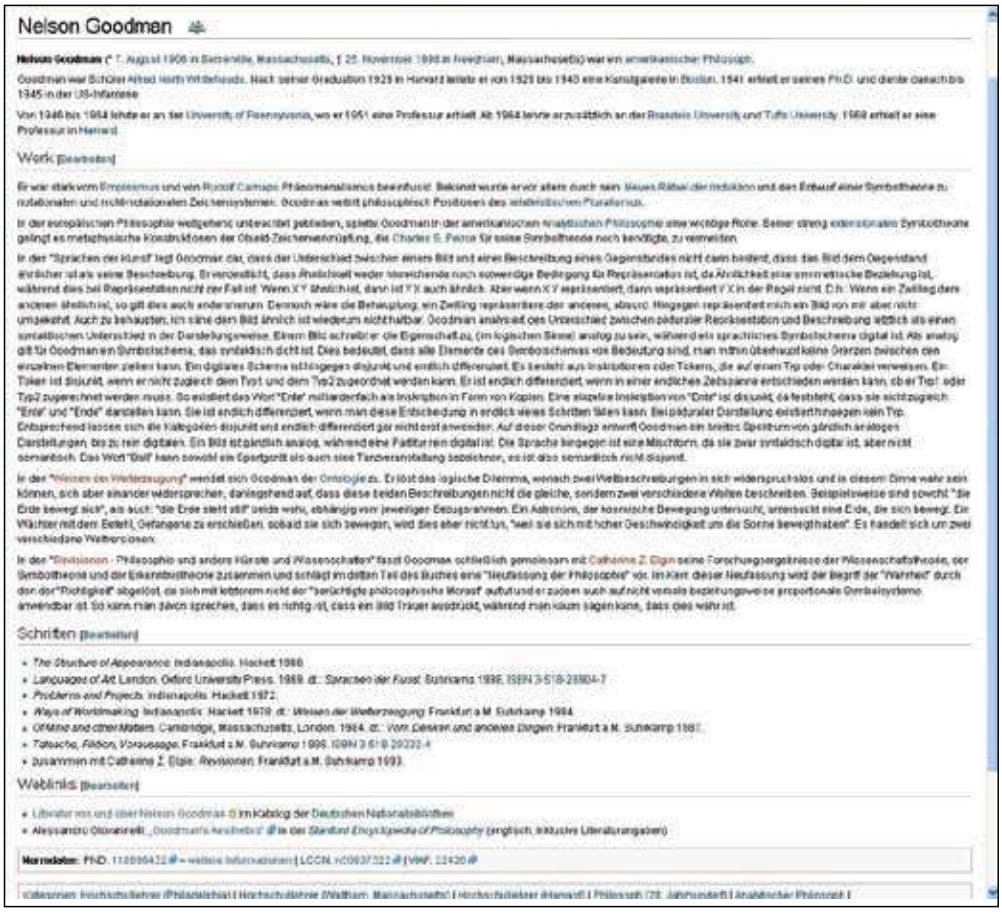


Abbildung 1: Wikipedia-Eintrag zu Nelson Goodman

4 Verwiesen sei hier zur Beantwortung dieser Frage auch auf die sehr gute, knappe einführende Übersicht mit weiterführenden Links Campbell, Lorna M. / MacNeill, Sheila (2010): The Semantic Web, Linked and Open Data. A Briefing Paper. Einsehbar unter http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf

5 Eine der ersten breit rezipierten Publikationen zum Semantic Web ist Berners-Lee, Tim / Hendler, James / Lassila, Ora (2001): The Semantic Web, Scientific American 284, no. 5: 34. Frei zugänglich unter http://www.scientificamerican.com/article.cfm?id=the-semantic-web

6 http://www.w3.org/2001/sw/

7 http://www.w3.org/standards/semanticweb/data

8 Für einen Einblick in Experimente und Entwicklungen, die im Bereich des sogenannten *Semantic Publishing* stattfinden, vgl. Shotton, David (2009): Semantic publishing: the coming revolution in scientific journal publishing, Learned Publishing 22 (April 2009): 85-94. DOI: 10.1087/2009202.

Semantic Web in Bibliotheken

SWIB 10

29.–30. November 2010
Köln

gemeinsam veranstaltet vom

Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen

und der

Deutschen Zentralbibliothek für Wirtschaftswissenschaften – Leibniz Informationszentrum Wirtschaft

http://swib.org/swib10/

hbz
Wissen. Information. Innovation.

Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen

Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen · Jülicher Straße 6 · 50674 Köln · Telefon: 0221 40075-0 · Telefax: 0221 40075-180 · www.hbz-nrw.de

nologien in Zukunft ohnehin eine große Rolle im Bibliothekswesen spielen werden. Darüber hinaus gibt es eine Menge weiterer Gründe, Linked Library Data zu produzieren, von denen hier einige genannt seien:

- Um endlich im Web anzukommen[9]
- Um die eigene Sichtbarkeit im Web zu maximieren und die eigenen Bestände leichter auffindbar zu machen.
- Um die Nachnutzbarkeit bibliothekarischer Daten durch andere – insbesondere nichtbibliothekarische Akteure – sicherzustellen
- Um die Informationen aus Bibliotheks- und Verbunddatenbanken durch die Verknüpfung mit anderen Informationen anzureichern
- Um die Recherchemöglichkeiten über die Daten zu optimieren

Illustriert sei das hier Geschilderte an einem Beispiel: Eine interessierte Kölnerin hat von Nelson Goodmans Symboltheorie gehört und möchte mehr darüber erfahren. Sie gibt eine entsprechende Suche bei Google ein und landet beim Wikipedia-Eintrag zu Nelson Goodman (siehe Abbildung 1).

Der Wikipedia-Eintrag liefert rudimentäre Informationen zu Goodman und seinem Werk sowie eine Liste seiner Schriften. Auch haben fleißige Wikipedianer bereits eine Verlinkung in den Katalog der Deutschen Nationalbibliothek (DNB) geschaffen, so dass mit einem Klick Literatur von und über Nelson Goodman aus dem Bestand der DNB angezeigt werden kann.[10]

Für einen ersten Eindruck hat die Kölnerin also bereits einige Informationen erhalten und sie weiß auch, welche Bücher sie sich näher anschauen könnte, wenn sie tiefer in das Thema einsteigen will. Allerdings gibt es für sie keine Gelegenheit direkt herauszufinden, welche dieser Bücher in Kölner Bibliothekskatalogen verzeichnet sind und von ihr ausgeliehen werden können. Kennt sie die Rechercheangebote wie KölnBib[11] oder einzelne Kataloge Kölner Bibliotheken nicht, so wird sie sich das Buch im Internet bestellen oder einfach auf eine weitergehende Lektüre verzichten.

Was aber könnte passieren, wenn die Daten aus deutschen Bibliothekskatalogen als Linked Open Data vorlägen? Wie die Verlinkung in den Katalog der DNB zeigt, arbeiten Wikipedianer seit einiger Zeit daran, auf Angebote von Bibliotheken zu verlinken. Mit Linked Open Bibliographic Data wäre es technisch nicht schwierig, Informationssuchenden im Zusammenhang eines Wikipedia-Artikels eine Liste von Bibliotheken aus ihrer Umgebung anzeigen zu lassen, die das jeweilige Buch im Bestand haben.[12] Damit könnten zusätzliche Informationen über Ort, Öffnungszeiten und Ausleihmöglichkeiten verbunden sein, so dass die Auswahl einer geeigneten Bibliothek leicht fällt. Was vorerst Zukunftsmusik ist, könnte in nicht allzu ferner Zeit Realität werden, sobald Bibliotheken und Verbünde in großem Maßstab ihre Katalogdaten in das Linked-Open-Data-Netz migrieren. Was das bedeutet und wie die Migration vonstattengehen könnte, wird im weiteren Verlauf des Textes erläutert.

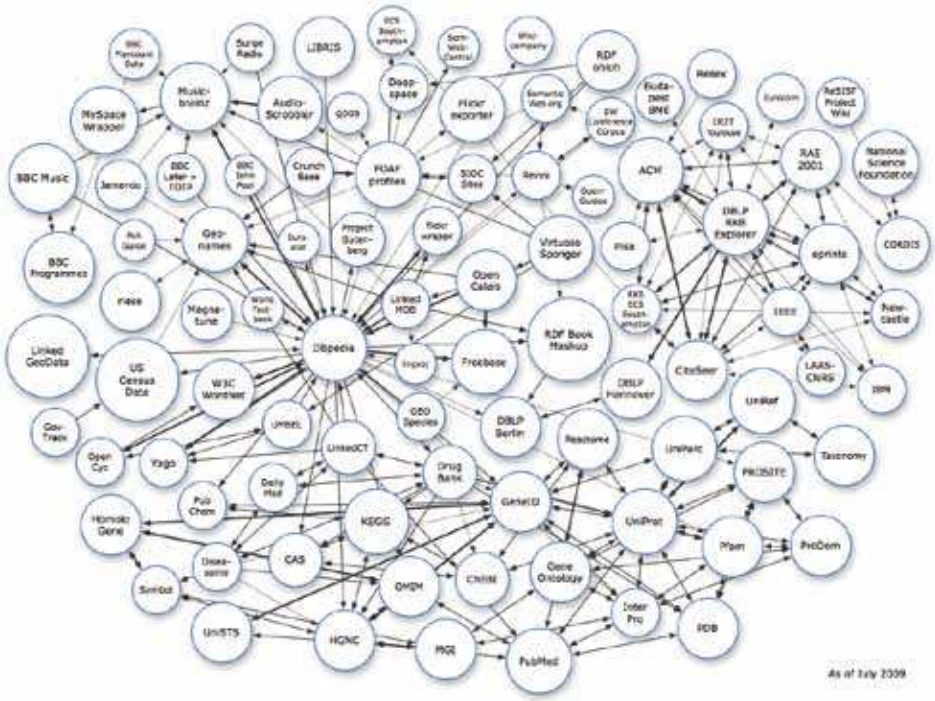


Abbildung 2: Die Linked-Open-Data-Cloud im Juli 2009 (Quelle: http://richard.cyganiak.de/2007/10/lod/)

4 Linked-Data-Prinzipien

Die von Tim Berners-Lee aufgestellten Linked-Data-Prinzipizen dienen als anerkannte Konventionen, an die sich halten sollte, wer Linked Data publizieren möchte:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs. so that they can discover more things.[13]

Es wurden bereits eine Menge Daten gemäß diesen Linked-Data-Prinzipien publiziert. Abbildung 2 zeigt die Linked-Data-Cloud vom Juli 2009, eine Darstellung der zu diesem Zeitpunkt gemäß Linked-Data-Prinzipien veröffentlichten Datensammlungen. Zu diesem Zeitpunkt war die klassische Bibliothekswelt noch nicht besonders gut vertreten, allein am oberen Rand (Mitte links) findet sich *Libris*[14], der schwedische Verbundkatalog. Mittlerweile dürfte die Linked-Data-Cloud um einiges angewachsen sein und auch immer mehr Daten aus Bibliotheken werden nach Linked-Data-Prinzipien veröffentlicht, verwiesen sei hier z.B. auf die Aktivitäten der Library of Congress (LoC)[15] oder der DNB[16].

Was sich hinter den Abkürzungen *URI*, *RDF* und *SPARQL* verbirgt, wird im nächsten Abschnitt näher erläutert werden.

9 Die meisten Bibliothekskataloge genügen bis heute nicht den gängigsten Web-Konventionen: Einträge sind nicht verlinkbar und die Inhalte können nicht in Suchmaschinen indexiert werden. Vgl. dazu Blogbeitrag und Kommentare: Hauschke, Christian (2010): Permalinks für Katalogisate, einsehbar unter http://infobib.de/blog/2009/10/27/permalinks-fur-katalogisate/

10 Diese Art der Verlinkung ist nur möglich, weil die Titel- und Normdaten der DNB permanente URLs – sogenannte Permalinks – haben, im Unterschied zu vielen anderen Katalogen. Vgl. hierzu Fußnote 9

11 http://www.koelnbib.de/

12 Die Wikipedia scheint mittlerweile auch darauf zu setzen, das in ihr gesammelte Wissen mit RDF anzureichern, vgl. Simonite, Tom (2010): Wikipedia to Add Meaning to Its Pages. Einsehbar unter http://www.technologyreview.com/web/25728/?a=f

13 http://www.w3.org/DesignIssues/LinkedData.html

14 http://libris.kb.se/ Die Daten des Libris liegen allerdings nicht als Komplettabzug zum Download bereit, weshalb hier nicht von Linked Open Data, sondern von Linked Data gesprochen werden kann.

15 Die LoC hat im Mai 2009 begonnen, ihre Normdaten in RDF zu veröffentlichen, vgl. http://id.loc.gov/authorities/

16 Auch die DNB hat die Normdaten als Linked Data veröffentlicht, vgl. die Dokumentation unter https://wiki.d-nb.de/display/LDS/Dokumentation+des+Linked+Data+Services+der+DNB. Allerdings sind die Daten nicht unter einer offenen Lizenz veröffentlicht.

5 Linked Data: Standards und Grundfunktion

Welche Technologie steckt hinter Linked Data? Welche sind die Prinzipien, auf deren Fundament das Linked-Data-Netz wächst? Bereits bestehende Web-Standards bilden auch für Linked Data die Basis: Das *Hypertext Transfer Protocol* (HTTP) als Datenaustauschprotokoll und HTTP-*URIs* (Uniform Resource Identifier)[17] zur Identifikation von Dingen[18] sind Eckpfeiler von Linked Data. Darüber hinaus kommen weitere Standards hinzu. Im Folgenden wird ein knapper Überblick über die grundlegenden Standards geliefert.

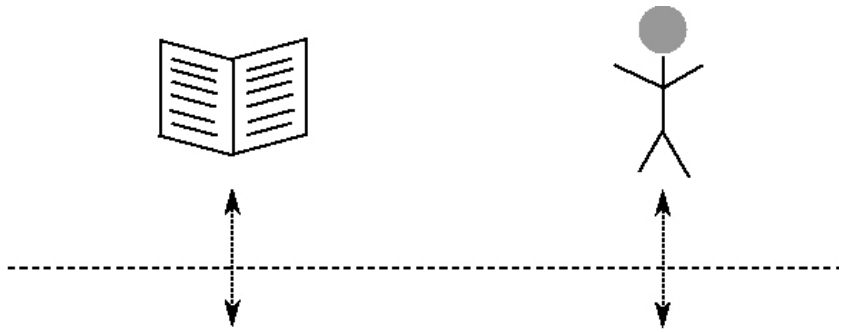


Abbildung 3: URIs als Namen von Dingen

URIs

Der Unterschied zum aktuellen Web besteht darin, dass HTTP-URIs nicht mehr allein zur Identifikation und Adressierung von abrufbaren Webressourcen (HTML-Seiten, Dateien), sondern auch zur Bezugnahme auf Nicht-Web-Ressourcen wie Personen, Bauwerke, Orte etc. benutzt werden. Abbildung 2 zeigt, wie URIs als Namen von Dingen wie einem Buch oder einer Person genutzt werden.[19]

Das besondere an URIs ist: Sie sind öffentliche Identifikatoren. Jeder kann sie verwenden. Das heißt, jeder Mensch kann überall RDF-Aussagen mit jedem URI veröffentlichen und dadurch das Wissen über das referenzierte Ding erweitern. URIs können auch als Dubletten verknüpft werden: Auch das ist eine Aussage.[20]

RDF-Tripel

Als wichtigste Ergänzung zu bestehenden Standards kommt das Resource Description Framework (RDF) hinzu, ein Modell zur Repräsentation einzelner Aussagen. Es ist so konzipiert, dass Maschinen optimal über RDF-Daten operieren können und ist somit gewissermaßen für Linked Data das, was HTML für das – in erster Linie menschenlesbare – World Wide Web ist. [21]

Hier ein Beispiel zu Veranschaulichung: Die Aussage „Platon ist der Autor des *Sophistes*." ist für jeden Menschen verständlich, der Deutsch spricht und ein gewisses Hintergrundwissen mitbringt, der etwa weiß, dass ‚Platon' auf einen antiken Philosophen bezugnimmt und ‚Sophistes' auf ein in Dialogform geschriebenes philosophisches Werk. Maschinen können mit diesem Satz nicht allzu viel anfangen. Dies ändert sich, wenn diese Aussage in Form einer RDF-Aussage vorliegt und mit anderen Aussagen verknüpft ist. RDF-Aussagen bestehen aus drei Teilen, aus diesem Grund spricht man auch von einem RDF-*Tripel*. Die drei Teile werden ‚Subjekt', ‚Prädikat' und ‚Objekt' genannt. Schauen wir uns einmal eine solche Aussage an:

<http://dbpedia.org/resource/Sophist_(dialogue)> <http://purl.org/dc/elements/1.1/creator> <http://dbpedia.org/resource/Plato>.[22]

Wie wir sehen, werden die drei Teile der RDF-Aussage (Subjekt, Prädikat und Objekt) durch HTTP-URIs repräsentiert. Warum ist das so? Was für Menschen nur schwer lesbar erscheint, erweist sich für eine verteilte Datenbank im Web als sehr vorteilhaft. Zum einen sind URIs – idealerweise – eindeutige Identifier, d.h. sie werden zur Referenz auf genau ein Ding benutzt, denn die Regel ist, einen URI nicht als Name für zwei verschiedene Dinge zu verwenden. Abbildung 3 illustriert, auf was die Bestandteile des Tripels bezugnehmen. Zum anderen ermöglichen HTTP-URIs es Maschinen, die Namen auf Basis des http-Protokolls aufzulösen, d.h. bei Eingabe in einen Browser Informationen über das referenzierte Ding zu liefern.[23]

Typisierte Verlinkungen

Wie wir in unserem Beispiel sehen, sind nicht nur die referenzierten Dinge durch einen URI identifiziert, sondern auch die Verbindung zwischen beiden: das Prädikat. Im konkreten Fall handelt es sich um den URI[24] für das Dublin-Core-Metadatenelement „creator". Hier wird ein grundlegender Unterschied im Vergleich zum World Wide Web deutlich. Zwar basiert auch das Web auf Links, allerdings sind diese nicht typisiert, d.h. in einem HTML-Dokument steht zwar, dass dieses Dokument auf ein anderes verlinkt, welcher Art diese Verlinkung ist, bleibt allerdings allenfalls für Menschen nachvollziehbar. Ob auf das Dokument zum Zweck eines Hinweises, einer Rezension etc. verlinkt wird, ist nicht explizit angeben. Im Linked-Data-Netz ist dies anders, hier ist *jede* Verlinkung typisiert, in unserem Fall als eine Werk-Autor-Beziehung.

Literale

Natürlich müssen die URIs in RDF mit natürlicher Sprache, Bildern usw. in Verbindung gesetzt werden, um letztlich uns Menschen Nutzen zu bringen. Deshalb ist es möglich an die Objektstelle eines Tripels Freitext (in RDF-Sprech ein ‚Literal') zu schreiben, während an Subjekt- und Prädikatstelle immer ein URI stehen muss. [25] Literale spielen zum Beispiel eine Rolle, wenn es darum geht, eine Aussage über den Titel eines Werkes zu machen. Ein Beispiel:

<http://dbpedia.org/resource/Sophist_(dialogue)> <http://purl.org/dc/elements/1.1/title> „Sophistes" .

RDF verfügt über zwei Möglichkeiten, um

17 URIs kennen wir alle. Der Link in der vorherigen Fußnote ist eine URI, aber auch andere Identifikatoren folgen dem URI-Schema, z.B. URNs oder auch Mailboxangaben der Form mailto:pohl@hbz-nrw.de . Zum URI-Schema und weiteren Beispielen vgl. Wikipedia, The Free Encyclopedia (2010): Artikel „URI scheme". Bearbeitungsstand: 9. Juli. URL: http://en.wikipedia.org/w/index.php?title=URI_scheme&oldid=372490112

18 ‚Ding' wird hier – sehr allgemein – im Sinne von ‚Entität' verwendet und umfasst auch abstrakte Konzepte.

19 In diesem und den folgenden Beispielen benutzen wir dbpedia-URIs aus dem Namensraum http://dbpedia.org. DBpedia ist ein Projekt der Freien Universität Berlin und der Universität Leipzig, das strukturierte Information aus der Wikipedia als Linked Data bereitstellt.

20 Wenn auch die Eineindeutigkeit von URIs relativ leicht herzustellen sein mag, so lässt sich Redundanz, Die durch die Bezeichnung eines Dings durch mehrere URIs entsteht, nicht immer verhindern. Dann wird es nötig, die URIs als gleichwertig zu bestimmen.

21 Darüber hinaus gibt es mit RDFa einen Standard, der HTML und RDF verschmilzt, indem er die Einbettung von RDF in HTML ermöglicht. Siehe hierzu etwa W3C (2008): RDFa Primer - Bridging the Human and Data Webs. Einsehbar unter http://www.w3.org/TR/xhtml-rdfa-primer/

22 Das Prädikat dc:creator und seine Nutzung können für manche verwirrend sein. Hier hilft, RDF wie ein Formular zu lesen: Das Subjekt ist das Ding, über das im Formular geredet wird. Dazu gibt es ein Element „creator", dessen Inhalt durch das Objekt angegeben wird.

23 Das Ding selbst kann freilich nicht geliefert werden, solange es sich nicht um eine Web-Ressource handelt. Eine Eingabe von http://dbpedia.org/resource/Plato in meinen Browser wird mir nicht Platon in mein Zimmer liefern, sondern leitet mich auf eine andere Seite (http://dbpedia.org/page/Plato) weiter, die Informationen über Platon enthält.

24 Die Verwendung von URIs für Prädikate macht es möglich, das Prädikat selbst im Web nachschlagen zu können, wenn z.B. Der Verwendungszweck nicht klar ist. Siehe hierzu auch den folgenden Abschnitt über Ontologien.

25 Die abstrakte RDF-Syntax, die bestimmt, was alles RDF genannt werden darf, ist hier zu finden: http://www.w3.org/TR/rdf-concepts/#section-Graph-syntax. Der Einfachheit halber lassen wir Blank Nodes in der vorliegenden Darstellung außen vor.

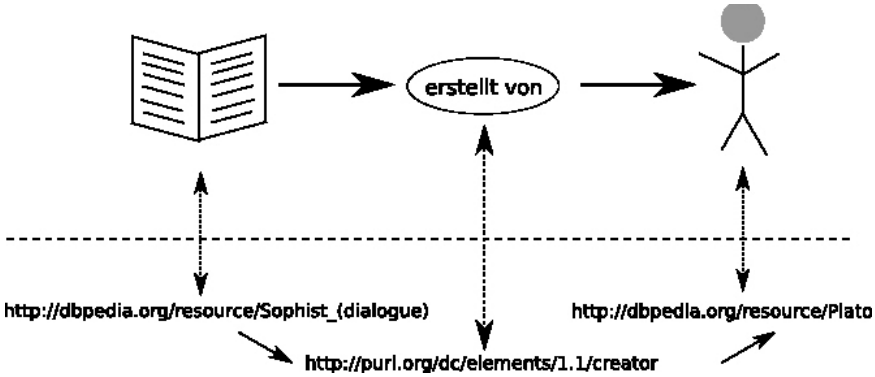


Abbildung 4: RDF-Tripel

solche Literale näher zu charakterisieren: die Angabe der Sprache, in der das Literal verfasst ist und die Angabe eines Datentyps wie etwa „Ganzzahl" oder „Datum". Es ist dabei je Literal nur eine Art der Markierung erlaubt, zum Beispiel:

<http://dbpedia.org/resource/Sophist_(dialogue)> <http://purl.org/dc/elements/1.1/title> „Sophistes"@de .
<http://dbpedia.org/resource/Sophist_(dialogue)> <http://purl.org/dc/elements/1.1/title> „Sophist"@en .
<http://dbpedia.org/resource/Sophist_(dialogue)> <http://purl.org/dc/elements/1.1/title> „Σοφιστής"@el .
<http://dbpedia.org/resource/Sophist_(dialogue)> <http://purl.org/dc/elements/1.1/created> „-360"^^xsd:gYear .

Die Markierung von Literalen mit Attributen ist die eine Form der in RDF vorgesehenen Möglichkeiten zur Typisierung. Um auch Aussagen über den Typ von Dingen treffen zu können, muss man sich auf das Gebiet der Ontologien wagen.

Ontologien

Wie bereits erläutert, handelt es sich bei RDF um ein abstraktes Modell, in dem Aussagen in Form von Tripeln gemacht werden können. Im Prinzip können dabei einem Subjekt über beliebige Prädikate beliebige Objekte zugeordnet werden. Um diese Beliebigkeit einzuschränken, kann anhand von Ontologie-Sprachen wie RDF-Schema (RDFS) und der Web Ontology Language (OWL) festgelegt werden, welche Klassen von Entitäten im Kontext der jeweiligen Ontologie als sinnvoll erachtet werden, welche Beziehungen – also Prädikate – zwischen Entitäten herrschen können und welche Schlüsse man aus diesen Beziehungen ableiten kann.

Das folgende Ontologie-Fragment definiert beispielhaft die Klassen *Person*, *Dokument* und *Buch*, wobei *Buch* als Unterklasse von *Dokument* modelliert wird. Darüber hinaus wird die Relation bzw. das Prädikat *autor* definiert[26]:

@prefix : <#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
:Person rdf:type rdfs:Class .
:Dokument rdf:type rdfs:Class .
:autor rdf:type rdf:Property .
:Buch rdfs:subClassOf :Dokument .
:autor rdfs:domain :Dokument .
:autor rdfs:range :Person .

In dieser Ontologie stecken folgende explizite Definitionen:

- Es gibt die Klassen *Person und Dokument*.
- Es gibt das Prädikat *autor*.

Darüber hinaus werden drei Regeln definiert:

- Wenn in einem Tripel das Prädikat *autor* verwendet wird, dann ist das Subjekt notwendigerweise ein *Dokument*.
- Wenn in einem Tripel das Prädikat *autor* verwendet wird, dann ist das Objekt notwendigerweise eine *Person*.
- Wenn etwas ein *Buch* ist, dann ist es immer auch ein *Dokument* (aus diesem Verhältnis zwischen Ober- und Unterklasse geht auch hervor, dass es sich bei *Buch* ebenfalls um eine Klasse handeln muss).

Anhand dieser Regeln wird zum einen deutlich, wie die Instanzen einer Klasse zueinander in Beziehung stehen können. Darüber hinaus ermöglicht es die Ontologie, implizite Aussagen inferieren zu können. So kann aus der Aussage

:Sophistes :autor :Plato

gefolgert werden, dass

- :Sophistes vom Typ :Buch ist,
- :Sophistes damit auch vom Typ :Dokument ist und dass
- :Plato vom Typ :Person ist.

Die Kernaufgaben einer Ontologie sind also die Definition von Klassen, möglichen Beziehungen und möglichen Schlüssen, die aus diesen Beziehungen hervorgehen. Ontologien sind damit ein sehr mächtiges Werkzeug für die Modellierung beliebiger Weltausschnitte.

Ontologien werden für bestimmte Zwecke geschaffen, idealerweise von Fachleuten des Gebietes, das die jeweilige Ontologie strukturieren soll. Es ist sinnvoll, wenn möglich auf bestehende Ontologien oder Teile derer zurückzugreifen, so dass nicht verschiedene Ontologien für ein und dieselbe Domäne entstehen. Im Umkehrschluss heißt dies für die Veröffentlichung von Ontologien, deren Nachnutzbarkeit sicherzustellen. Dies schließt eine ausführliche Dokumentation

26 Aufgrund der Abstraktheit der Syntax gibt es für RDF mehrere Notationssysteme. In diesem Beispiel wird zur Steigerung der Lesbarkeit die Turtle-Notation verwendet. Vgl. http://www.w3.org/2007/02/turtle/primer/http://www.w3.org/2007/02/turtle/primer/

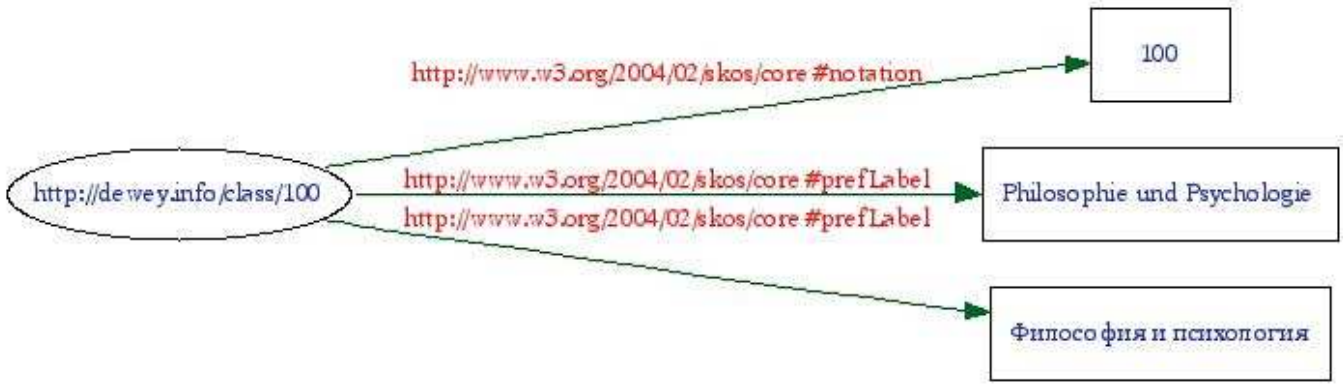


Abbildung 5: Darstellung einer DDC-Systemstelle in SKOS[31] *(siehe nächste Seite)*

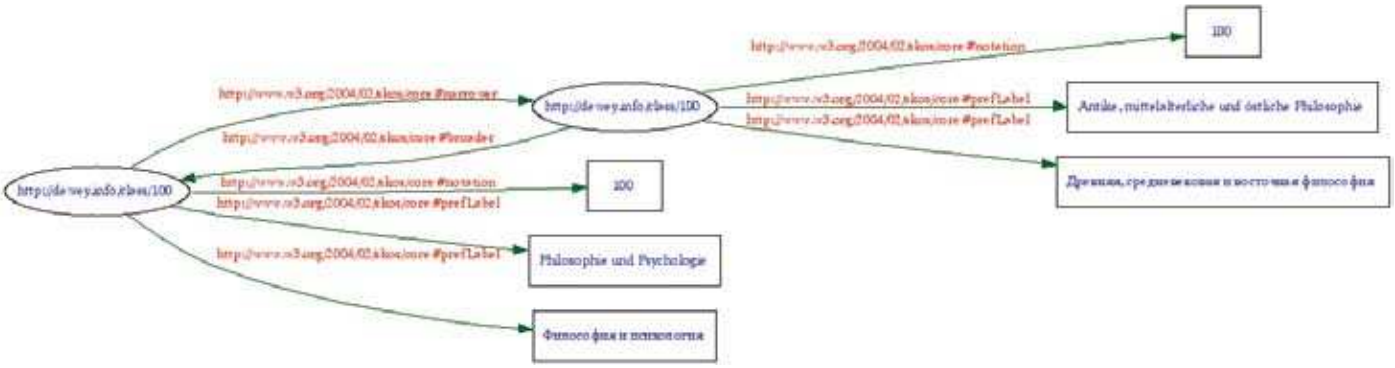


Abbildung 6: Darstellung der Verknüpfung von DDC-Systemstellen in SKOS

sowie die Lizenzierung der Ontologie unter einer offenen Lizenz mit ein.[27]

SKOS

SKOS steht für „Simple Knowledge Organization System"[28]. Es ist entwickelt worden, um kontrollierte Vokabulare – also Klassifikationen, Thesauri und Taxonomien – Linked-Data-konform zu kodieren. Dies ermöglicht es, solche Dokumentationssprachen (1) modular aufzubauen, um die Nachnutzbarkeit zu erhöhen, und somit (2) miteinander vernetzen und in Beziehung setzen zu können.

SKOS ist demnach eine Ontologie für kontrollierte Vokabulare. Der Aufbau von SKOS ist denkbar einfach. Grundbausteine sind *Konzepte*, denen Notationen und Bezeichnungen[29] zugeordnet werden[30]:

Einzelne Konzepte können anhand der hierarchischen Relationen *narrower* und *broader* miteinander Verknüpft werden[32]:

Neben den Relationen, die innerhalb eines kontrollierten Vokabulars bestehen, können darüber hinaus Verbindungen zu anderen Vokabularen etabliert werden. So können Konzepte der Dewey Decimal Classification (DDC) beispielsweise wie folgt anhand des Prädikats *closeMatch* auf die Konzepte der Library of Congress Subject Headings (LCSH)[33] abgebildet werden:

<http://dewey.info/class/100> <http://www.w3.org/2004/02/skos/core#closeMatch> <http://id.loc.gov/authorities/sh99005065#concept> .

Weitere bereits in SKOS veröffentlichte Dokumentationssprachen sind der Standard Thesaurus Wirtschaft[34], der Thesaurus Sozialwissenschaften[35] und die Normdaten der Deutschen Nationalbibliothek[36].

Bibliographische Ressourcen können durch RDF-Tripel mit den in SKOS kodierten kontrollierten Vokabularen verknüpft werden:

<http://dbpedia.org/resource/Sophist_(dialogue)> <http://purl.org/dc/elements/1.1/subject> <http://dewey.info/class/180> .

Doch welchen Wert haben all diese Verknüpfungen? Sie ermöglichen es zum einen, über das Linked-Data-Netz zu navigieren und zu neuen relevanten Ressourcen aus anderen Kontexten zu gelangen. So nützlich und sinnvoll allein schon das Navigieren über Linked Data sein kann, so notwendig ist natürlich zum anderen auch die Möglichkeit der Recherche. Im nächsten Abschnitt wird der Standard skizziert, der Abfragen über RDF-Tripeln dient.

SPARQL

Ein großer Vorteil von Linked Data ist , dass komplexe Abfragen über Daten gemacht werden können, sobald eine hinreichend große und komplexe Datenbasis vorliegt. SPARQL ist die Abfragesprache für RDF und das Akronym wird aufgelöst in „SPARQL Protocol And RDF Query Language". Sie dient Entwicklern dazu, Anwendungen auf dieser Basis aufzubauen und ist nicht für den Einsatz durch Laien gedacht.

Eine einfache SPARQL-Anfrage an die DBpedia kann etwa so aussehen:

```
Select ?x
Where {
?x <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person> .
?x <http://dbpedia.org/ontology/era> <http://dbpedia.org/resource/Ancient_philosophy> .
?x <http://dbpedia.org/property/mainInterests> <http://dbpedia.org/resource/Epistemology> .
}
```

Die Antwort auf diese Abfrage liefert uns sämtliche in DBpedia verzeichneten Personen, die der antiken Philosophie zugeordnet werden und deren Hauptinteresse der Epistemologie galt.[37]

SPARQL-Abfragen können beliebig komplex sein, d.h. es lassen sich auch Anfragen der Art bilden: *Liefere mir alle weiblichen Autoren, die während des ersten Weltkriegs in Wien publiziert haben und vor 1880 geboren sind.*

Voraussetzung für die Durchführung komplexer Anfragen sind natürlich immer große Mengen sorgfältig erfasster, standardisierter und komplexer Daten. Der Aufwand zur Erstellung und Pflege von Daten wird einem durch Linked Data nicht abgenommen.

27 Beispielhaft sei hier die Dokumentation der DC-Terms genannt, wo allerdings (bisher) eine offene Lizenzierung versäumt wurde, vgl. http://dublincore.org/documents/dcmi-terms/

28 Es folgt an dieser Stelle nur eine sehr knappe Skizze. Für eine ausführliche Einführung siehe http://www.w3.org/TR/skos-primer/

29 Es sind grundsätzlich zwei Arten von Bezeichnungen vorgesehen: bevorzugte Bezeichnungen, von denen jedes Konzept für jede Sprache nur eine haben darf, und eine beliebige Menge von alternativen Bezeichnungen.

30 Dieses Beispiel stützt sich auf die Linked-Data-Variante der Dewey Decimal Classification, die die ersten drei Ebenen umfasst und welche unter http://dewey.info/http://dewey.info/ zu finden ist.

31 Die Abbildungen 4 und 5 exemplifizieren eine weitere Weise der Darstellung von RDF-Tripeln: die Darstellung als gerichteter Graph mit Knoten und Kanten. URI-Knoten werden als Oval dargestellt, Literale als Rechteck. Unter http://www.w3.org/RDF/Validator/ können aus RDF-Beschreibungen solche Grafiken generiert werden.

32 Darüber hinaus gibt es mit *related* auch eine nichthierarchische Relation, die Querverweise ermöglicht.

33 Wie bereits erwähnt, stellt die Library of Congress die LCSH unter http://id.loc.gov/authorities/ auch als Linked Data bereit.

34 http://zbw.eu/stw/, vgl. auch Borst, Timo / Neubert, Joachim (2009): Case Study: Publishing STW Thesaurus for Economics as Linked Open Data, http://www.w3.org/2001/sw/sweo/public/UseCases/ZBW/

35 http://www.gesis.org/dienstleistungen/tools-standards/thesaurus-sozialwissenschaften/, vgl. auch Benjamin Zapilko und York Sure (2009): Converting TheSoz to SKOS. Einsehbar unter http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2009/technicalreport_09_07.pdf

36 Vgl. Fußnote 16.

37 Die Antwort auf diese Anfrage lässt sich hier einsehen: http://dbpedia.org/snorql/?query=Select+%3Fx%0D%0AWhere+{+%0D%0A%3Fx+%3Chttp%3A%2F%2Fwww.w3.org%2F1999%2F02%2F22-rdf-syntax-ns%23type%3E+%3Chttp%3A%2F%2Fxmlns.com%2Ffoaf%2F0.1%2FPerson%3E+.%0D%0A%3Fx+%3Chttp%3A%2F%2Fdbpedia.org%2Fontology%2Fera%3E+%3Chttp%3A%2F%2Fdbpedia.org%2Fresource%2FAncient_philosophy%3E+.%0D%0A%3Fx+%3Chttp%3A%2F%2Fdbpedia.org%2Fproperty%2FmainInterests%3E+%3Chttp%3A%2F%2Fdbpedia.org%2Fresource%2FEpistemology%3E+.%0D%0A+}%0D%0A. Die geringe Trefferzahl spricht dafür, dass die Daten in der Wikipedia bisher nicht besonders sorgfältig erfasst bzw. nicht in RDF umgewandelt worden sind.

6 Der Migrationspfad ins Linked-Data-Netz

Wie gelangen die Daten aus Bibliotheks- und Verbundkatalogen nun in das Linked-Data-Netz? Was müssen wir dafür tun? Im Folgenden werden sechs grundlegende Schritte erläutert.

Daten analysieren

Wir müssen unsere Daten kennen und müssen wissen, welche Regelmäßigkeiten in den Katalogisaten zu finden sind. Die grundlegenden Fragen sind:

- Welche Dinge beschreiben unsere Daten?
- Wie geschieht dies?

Die erste Frage mag leicht erscheinen (Antwort: Personen und Medien), aber man muss sich nur einmal eine FRBR-Diskussion anschauen, um zu merken, dass es gar nicht so klar ist, welche Dinge in Katalogen eigentlich beschrieben werden (sollten).[38]
Die zweite Frage ist noch kniffliger. Zwar gibt es niedergeschriebene Katalogisierungsregeln und standardisierte Datenformate, allerdings kann man sich nicht darauf verlassen, dass bei der tatsächlichen Katalogisierung streng diesen Regeln gefolgt wurde bzw. die Regeln auf dieselbe Art und Weise interpretiert wurden. Es gibt in jeder Bibliothek und jedem Verbund individuelle Katalogisierungspraktiken.
Daten analysieren heißt also nicht nur Regelwerke analysieren, sondern auch statistische Daten über die tatsächliche Benutzung von MAB-/MARC-Feldern und deren Relationen untereinander zu erheben.

URIs prägen vs. URIS nachnutzen

Bevor RDF-Tripel aus Katalogdaten generiert werden können, müssen die beschriebenen Dinge erst einmal mit einem URI „getauft" sein. Es gibt zwei Wege, an URIs für katalogisierte Dinge zu kommen:

1. Ich präge in einem von mir kontrollierten Namensraum[39] für jedes zu beschreibende Ding neue URIs.
2. Ich ermittle, ob für das beschriebene Ding bereits eine URI existiert und benutze diese URI zur Beschreibung des Dings.

Für die erste Variante spricht die Einfachheit. Es ist momentan nämlich alles andere als trivial, existierende URIs zu finden. Allerdings spricht etwas dafür, diese Mühen, die die zweite Variante mit sich bringt, auf sich zu nehmen: es ergibt sich dadurch eine höhere Auffindbarkeit der Daten, da der Name des referenzierten Dinges bereits außerhalb der eigenen Daten bekannt ist. Darüber hinaus wächst durch die Nutzung bestehender URIs das Wissen über ein Ding sofort an, was die Abfragemöglichkeiten verbessert. Ein Mittelweg ist es, eigene URIs zu prägen und diese über das Prädikat owl:sameas nach und nach mit weiteren existierenden URIs für dasselbe Ding zu verknüpfen.[40]

Ontologien evaluieren, nachnutzen, kreieren

Wenn die Frage nach URIs für die beschriebenen Dinge geklärt ist, werden passende Ontologien oder Vokabulare benötigt. Fragen, die es hier zu beantworten gilt, sind: Wie werden die Dinge in unseren Katalogen klassifiziert, und welche Eigenschaften werden ihnen zugeschrieben? Gibt es bereits Ontologien, die die entsprechenden Klassen und Eigenschaften vorweisen? Dieser Schritt knüpft demnach nahtlos an den ersten Schritt der Datenanalyse an. Um die Interoperabilität der Daten zu maximieren, sollten bestehende Ontologien genau betrachtet werden.
Werden geeignete bereits existierende Klassen und Prädikate ausgemacht, so sollten diese auch verwendet werden. Ist dies nicht der Fall, so müssen eigene Ontologien aufgestellt werden. Häufig gibt es bestehende Ontologien, die aber nicht alle relevanten Klassen und Beziehungen aufweisen. In diesem Fall ist es günstig, die eigene Ontologie auf der bestehenden aufzubauen und neue Prädikate und Klassen zu ergänzen. Ein Beispiel für das Aufbauen auf relevanten Ontologien, die aber nicht spezifisch genug sind, ist die Bibliographic Ontology (Bibo)[41]: Sie baut auf den DCMI Metadata Terms[42] sowie der FOAF[43]- und der Event-Ontolgie[44] auf und ergänzt diese, wo es nötig ist, um eigene Klassen und Prädikate.
Für bibliographische Daten ist neben den bereits genannten Ontologien (Bibo und DCMI Metadata Terms) natürlich auch RDA (Resource Description and Access) interessant. Es gibt bereits erste Anfänge von RDA in RDF, wenn auch in einer schlecht konsumierbaren Version.[45]

Daten (ad-hoc) konvertieren

Steht fest, welche URIs und Ontologien verwendet werden sollen, kann mit der Konvertierung der Daten nach RDF begonnen werden. Diese Konvertierung ist in der Übergangsphase hin zum Linked-Data-Netz ein regelmäßig zu wiederholender Prozess, weil Anwendungen RDF und Ontologien noch nicht nativ unterstützen. Da die Daten in der jeweiligen Anwendung – etwa einem Katalogisierungssystem – gepflegt werden, müssen die regelmäßig vorgenommenen Änderungen aus dem verwendeten Format – MAB2, MARC21 etc. – in die RDF-Repräsentation übernommen werden.
Statt Daten in großen Paketen regelmäßig nach RDF zu konvertieren, können auch Schnittstellen geschaffen werden, die die Daten ad-hoc zu dem Zeitpunkt überführen, zu dem sie angefordert werden. Ein Beispiel für dieses Szenario ist der bereits angesprochene Linked-Data-Service der Deutschen Nationalbibliothek.[46]

Daten verlinken

Wie der Name „Linked Data" sagt, ist es essentiell, die eigenen Daten mit anderen zu vernetzen; nur so können die Vorteile von Linked Data voll ausgekostet werden. Das heißt, es sollten Literale wirklich nur dann in Tripeln vorkommen, wenn das Objekt nicht als Verweis auf ein anderes Ding aufgefasst werden kann – etwa bei dem Titel einer Publikation. Für viele Dinge existieren bereits geeignete Verknüpfungsmöglichkeiten; neben den bereits erwähnten SKOS-kodierten kontrollierten Vokabularen sei hier beispielhaft auf Geonames[47] für Ortsangaben verwiesen. Es ist beispielsweise wünschenswert, Geburts- und Verlagsorte nicht als Literale zu kodieren, sondern auf die URI des jeweiligen Ortes zu verlinken.
Nach der Veröffentlichung der Daten muss also daran gearbeitet werden, die Daten zunehmend mit anderen Daten zu verlinken. Es ist erstrebenswert, dass auf die eigenen URIs auch von außen verlinkt wird. Dies lässt sich nicht erzwingen. Eine offene Publikation von kompletten Datenabzügen kann sich hier aber positiv auswirken, da es die maschinelle Erzeugung von Links auf diese Daten stark vereinfacht.[48]

38 Ein Beispiel für eine solche Diskussion ist jene vom Mai/Juni 2010 auf der Mailingliste der Open Bibliographic Data Group, siehe http://lists.okfn.org/pipermail/open-bibliography/2010-May/000115.html.

39 ‚Namensraum' wird hier im Prinzip synonym mit ‚Domain' verwendet. Beispiele für Domains sind etwa wikipedia.org, hbz-nrw.de, w3.org.

40 In der Arbeitsgemeinschaft „Kooperative Verbundanwendungen" wird derzeit nach Mitteln und Wegen gesucht, die verhindern, dass in der deutschen Bibliothekswelt von verschiedenen Verbünden und Bibliotheken unzählige URIs für ein und dasselbe Ding geprägt werden. Zu diesem Zweck müsste so etwas wie ein Verzeichnis entstehen, in dem bestehende URIs mit Identifikatoren wie der ISBN, der OCLC-Nummer oder der Erstkatalogisierungs-ID (EKI) in Beziehung gesetzt werden, so dass auf diese Weise bereits geprägte URIs ermittelt werden können.

41 http://bibliontology.com/

42 http://dublincore.org/documents/dcmi-terms/. DCMI steht für ‚Dublin Core Metadata Initiative'.

43 http://www.foaf-project.org/. Das Akronym FOAF steht für ‚Friend of a Friend'.

44 http://motools.sf.net/event/event.html

45 Siehe RDA in der Metadata Registry der National Science Digital Library: http://metadataregistry.org/rdabrowse.htm. Vgl. auch Hillmann, Diane / Coyle, Karen / Phipps, John / Dunsire, Gordon: RDA Vocabularies: Process, outcome, Use. D-Lib Magazine 16, Nr. 1 (1, 2010). Einsehbar unter http://www.dlib.org/dlib/january10/hillmann/01hillmann.html.

46 Vgl. Fußnote 16.

47 http://www.geonames.org/

48 Vgl. zur offenen Lizenzierung von Daten den ersten

Im Web bereitstellen

Während ein Gesamtabzug der Daten für gewisse Zwecke sehr nützlich ist, ist dies nicht die einzige Form, in der die Daten publiziert werden sollten. So wie bei der Eingabe einer URL die entsprechende HTML-Seite von einem Webserver ausgeliefert wird, muss auch das Abrufen der RDF-Beschreibung eines Dings ermöglicht werden. In vielen Fällen, wenn es sich etwa um Personen oder Orte handelt, kann das Ding selbst nicht ausgeliefert werden.[49] Für diese Fälle muss der Webserver so konfiguriert werden, dass er auf ein beschreibendes Dokument weiterleitet, wenn ein solches Ding angefordert wird[50]. Schließlich bietet es sich auch an, die Daten über eine SPARQL-Schnittstelle zugänglich zu machen, um etwa Entwicklern eigene SPARQL-Abfragen zu ermöglichen.

7 Aktivitäten im hbz-Verbund

Es ist unschwer zu sehen, dass eine Migration bibliographischer Daten in das Linked-Data-Netz alles andere als trivial ist. In diesem Bereich muss noch eine Menge experimentiert werden, ehe diese Aufgabe befriedigend erledigt werden kann. Welche Linked-Data-Aktivitäten im hbz stattfinden soll nun abschließend knapp dargestellt werden.

Vernetzung und Kooperation

Da auf dem Weg zur erfolgreichen Implementierung von Linked Library Data noch eine Menge Erfahrungen gemacht werden müssen, ist es unerlässlich, sich mit anderen Akteuren über die auf diesem Gebiet gemachten Erfahrungen auszutauschen.

Seit etwa einem halben Jahr trifft sich im hbz – wie im ersten Teil dieser Reihe bereits erwähnt – eine Linked-Open-Bibliographic-Data-Arbeitsgruppe. Ursprünglich zur Vernetzung der Akteure im hbz und in der Region gedacht, ist diese Gruppe nun offen für alle, die an einer Überführung ihrer Daten in Linked Open Data interessiert sind.

Auch sucht das hbz regelmäßig den Erfahrungsaustausch in Arbeitstreffen und auf Tagungen, auf Mailinglisten oder durch die Ausrichtung der Konferenz *Semantic Web in Bibliotheken* (SWIB), die das hbz gemeinsam mit der Deutschen Zentralbibliothek für Wirtschaftswissenschaften (ZBW) im Jahr 2010 zum zweiten Mal veranstaltet.[51] Auf nationaler Ebene koordiniert das hbz seine Projekte zudem mit der DNB.

International engagiert sich das hbz vor allem in der Working Group on Open Bibliographic Data der Open Knowledge Foundation.[52] Seit Mai 2010 gibt es beim W3C eine Library Linked Data Incubator Group[53], die ebenfalls von Bedeutung ist. Bemerkenswert ist auch die kürzlich gestartete Resource Discovery Taskforce des britischen Joint Information Systems Committee (JISC), deren Vision wie folgt lautet: *UK researchers and students will have easy, flexible and ongoing access to content and services through a collaborative, aggregated and integrated resource discovery and delivery framework which is comprehensive, open and sustainable.*[54] Erste Kontakte zu dieser Task Force bestehen bereits.

Teil dieser Reihe: Pohl (2010).

49 Vgl. Fußnote 22.

50 Eine ausführliche Erläuterung dieser Problematik würde den Rahmen dieses Artikels sprengen, daher sei an dieser Stelle nur auf http://www.w3.org/TR/cooluris/#solutions verwiesen.

51 Siehe http://swib.org

52 Für mehr Informationen siehe die Wiki-Seite: http://wiki.okfn.org/wg/bibliography

53 Siehe http://www.w3.org/2005/Incubator/lld/

54 Vgl. das Visionspapier der Taskforce JISC (2010): One to Many; Many to One: The resource discovery taskforce vision. Einsehbar unter http://ie-repository.jisc.ac.uk/475/http://ie-repository.jisc.ac.uk/475/. Es ist bedauernswert, dass es in Deutschland nicht eine solche nationale impulsgebende Organisation wie die JISC gibt.

ONLINE BOOKS

Bereits vor einem Jahrzehnt war Wiley Vorreiter in Sachen Online Books. Heute zeichnet sich Wiley's Online Books-Sammlung durch eine unübertroffene Breite mit rund 10.000 Monografien aus, darunter zahlreiche wegweisende Buchreihen.

Finden Sie Spitzentitel berühmter Wissenschaftler, preisgekrönter Autoren und namhafter Forscher aus Biowissenschaft, Medizin und Naturwissenschaften, Gesellschafts- und Geisteswissenschaft in bekannten Imprints wie Wiley-Blackwell, Wiley-Liss, Wiley-VCH und Jossey-Bass.

Die Online Books stehen in der Wiley Online Library zur Verfügung. Als zukunftsweisende Plattform bietet Wiley's Online Library die nahtlose Integration von Must-Have-Inhalten in eine neue, flexible und einfach zu handhabende Forschungsumgebung.

Testen Sie die Möglichkeiten.

Besuchen Sie
wileyonlinelibrary.com/onlinebooks

WILEY ONLINE LIBRARY

Migration von Titeldaten nach Linked Data

Das wahrscheinlich aufwändigste Linked-Data-Projekt im hbz ist die Überführung von Titeldaten nach Linked Data. Da die Normdaten bereits von der DNB als Linked Data veröffentlicht werden, hat man sich im hbz direkt an die Aufgabe der Konvertierung von Titeldaten gemacht. Momentan wird ein MAB-RDF-Ansatz favorisiert.[55] Das heißt, die bestehenden MAB-Daten werden eins zu eins in RDF übersetzt: Jede einzelne Titelsatzzeile wird in ein RDF-Tripel umgewandelt, wobei die Prädikat-URIs aus der jeweiligen MAB-Feldnummer, dem Unterfeld und dem Indikator gebildet werden und in Objektstellung zunächst immer ein Literal – der jeweilige Katalogisierungseintrag – steht. Das Mapping der Prädikate auf bestehende oder neu zu erstellende Ontologien kann somit vollständig auf RDF-Ebene stattfinden. Darüber hinaus müssen URIs für die beschreibenden Objekte geprägt werden (vgl. Abschnitt 6) und Objekt-Literale in URIs überführt werden, ehe das Ganze Linked Data genannt werden kann.

Konvertierung von Dokumentationssprachen nach SKOS

Kleinere, weil leichter durchführbare Projekte sind Konvertierungen von Dokumentationssprachen nach SKOS. Dokumentationssprachen sind nicht so komplex wie Titeldaten und liegen bereits in einer klar strukturierten Form vor. Am hbz wurde bereits die *Regensburger Verbundklassifikation* (RVK) in SKOS umgewandelt. Das Ergebnis kann veröffentlicht werden, sobald einige offene rechtliche Fragen geklärt sind (die RVK ist nicht mit einer offenen Lizenz versehen).

Konzeption einer Open-Data-Plattform

Das hbz arbeitet an der Konzeption einer Open-Data-Plattform zur Migration von Informationen aus der und über die Bibliothekswelt ins Semantic Web. Derzeit finden Überlegungen statt, Daten über einzelne bibliothekarische Institutionen, über Kataloge und die durch sie beschriebenen Sammlungen in RDF zu erfassen, um damit die Basis für eine Migration der bibliographischen Daten ins Semantic Web zu schaffen. Eine detaillierte Darstellung dieser Überlegungen wird folgen.[56]

55 Dieser Ansatz ähnelt in Teilen dem in folgendem Text erläuterten: Styles, Rob / Ayers, Danny / Shapir, Nadeem (2008): Semantic MARC, MARC21 and the Semantic Web. Einsehbar unter http://events.linkeddata.org/ldow2008/papers/02-styles-ayers-semantic-marc.pdf

56 Der dritte Teil dieser Artikelreihe mit dem Titel „Eine Plattform für freie bibliographische Daten - Konzeption und Implementierung" wird im Tagungsband der DGI-Konferenz *Semantic Web & Linked Data - Elemente zukünftiger Informationsinfrastrukturen* erscheinen sowie – wie die anderen Teile auch – Open Access im Netz zur Verfügung stehen, u.a. anderem unter http://www.hbz-nrw.de/dokumentencenter/produkte/lod/.

8 Quellen

Für alle Webressourcen gilt: zuletzt eingesehen am 15.7.2010.

Berners-Lee, Tim / Hendler, James / Lassila, Ora (2001): The Semantic Web, Scientific American 284, no. 5: 34. Frei zugänglich unter http://www.scientificamerican.com/article.cfm?id=the-semantic-web.

Borst, Timo / Neubert, Joachim (2009): Case Study: Publishing STW Thesaurus for Economics as Linked Open Data, http://www.w3.org/2001/sw/sweo/public/UseCases/ZBW/.

Campbell, Lorna M. / MacNeill, Sheila (2010): The Semantic Web, Linked and Open Data. A Briefing Paper. Einsehbar unter http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf.

Hauschke, Christian (2010): Permalinks für Katalogisate. Blogbeitrag einsehbar unter http://infobib.de/blog/2009/10/27/permalinks-fur-katalogisate/.

Hillmann, Diane / Coyle, Karen / Phipps, John / Dunsire, Gordon: RDA Vocabularies: Process, outcome, Use. D-Lib Magazine 16, Nr. 1 (1, 2010). Einsehbar unter http://www.dlib.org/dlib/january10/hillmann/01hillmann.html.

Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (2010): Freigabe der Katalogdaten: Kölner Bibliotheken leisten Pionierarbeit. Einsehbar unter: http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe

JISC (2010): One to Many; Many to One: The resource discovery taskforce vision. Einsehbar unter http://ie-repository.jisc.ac.uk/475/

Pohl, Adrian (2010): Open Data im hbz-Verbund.: Was, Wie und Warum. Erscheint in ProLibris 3/2010. Preprint u.a. einsehbar unter http://www.hbz-nrw.de/dokumentencenter/produkte/lod/

Shotton, David (2009): Semantic publishing: the coming revolution in scientific journal publishing, Learned Publishing 22 (April 2009): 85-94. DOI: 10.1087/2009202.

Simonite, Tom (2010): Wikipedia to Add Meaning to Its Pages. Einsehbar unter http://www.technologyreview.com/web/25728/?a=f.

Styles, Rob / Ayers, Danny / Shapir, Nadeem (2008): Semantic MARC, MARC21 and the Semantic Web. Einsehbar unter http://events.linkeddata.org/ldow2008/papers/02-styles-ayers-semantic-marc.pdf

W3C (2008): RDFa Primer. Bridging the Human and Data Webs. Einsehbar unter http://www.w3.org/TR/xhtml-rdfa-primer/

Wikipedia, The Free Encyclopedia (2010): Artikel „URI scheme". Bearbeitungsstand: 9. Juli. URL: http://en.wikipedia.org/w/index.php?title=URI_scheme&oldid=372490112

Dieser Text ist unter folgende Creative-Commons-Lizenz veröffentlicht: Creative Commons Namensnennung 3.0 Deutschland.

AUTOREN

Adrian Pohl
ist im Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) im Bereich Open Data und Linked Data und in der Direktionsassistenz tätig.
Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz)
Jülicher Straße 6
50674 Köln
pohl@hbz-nrw.de

Felix Ostrowski
arbeitet im Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) in der Gruppe Publikationssysteme.
Jülicher Straße 6
50674 Köln
ostrowski@hbz-nrw.de

ibis e² –

Die Badische Landesbibliothek beschleunigt ihre Erwerbungsgeschäftsgänge mittels digitaler Verarbeitung des Neuerscheinungsdatendienstes der DNB

Bernt Walter und Mathias Landhäußer

Kosten zu senken, ohne den bestehenden Service einzuschränken, ist für wissenschaftliche Bibliotheken ein schwieriges Unterfangen. Die Einführung eines neuen, elektronischen Verfahrens für die Verarbeitung von Neuerscheinungen in der Badischen Landesbibliothek zeigt jedoch, dass die Analyse alter Geschäftsgänge zu großen Verbesserungen führen kann. Der neue Geschäftsgang basiert auf der elektronischen Ausgabe des Neuerscheinungsdienstes der DNB. Er verringert nicht nur die Bereitstellungszeiten für neue Medien und erleichtert die Bearbeitung von Pflichtexemplaren, sondern senkt darüber hinaus die laufenden Kosten für die Bibliothek.

ibis e² – How Digital Processing Speeds Up Acquisition

Cutting costs, while maintaining – or even improving – the quality of service, often seems unfeasible for research libraries. By introducing an electronic method for processing new publications, the state library of Baden achieved substantial improvements by overhauling long-standing business processes. The new method is based upon data digitally delivered by the German National Library. It has not only eased the handling of legal deposits, but has also reduced the running costs for the library. Furthermore, new publications can be delivered to the customers faster.

ibis e² – Grâce aux moyens digitaux, la BLB (Bibliothèque du Land de Bade) accélère l'acquisition.

Réduire le coût des bibliothèques scientifiques, sans restrictions des services appliqués, est une tentative audacieuse. L'introduction d'un nouveau procédé électronique pour l'adaptation des nouveautés à la BLB démontre bien que l'analyse des anciennes méthodes peut entraîner des améliorations considérables. La nouvelle méthode est basée sur l'édition électronique des nouveautés de la DNB (Bibliothèque Nationale d'Allemagne). Elle réduit non seulement le temps de mise en place pour les nouveaux médias tout en facilitant l'élaboration des exemplaires remis à titre de dépôt légal, mais en plus elle réduit les frais courants de la bibliothèque.

Die repräsentative Auswahl, das zeitnahe Beschaffen und die zügige Bereitstellung von aktueller gedruckter Literatur ist – auch im Internetzeitalter mit seinen vielfältigen elektronischen Angeboten – nach wie vor *die* zentrale Aufgabe für wissenschaftliche Bibliotheken. Angesichts der enormen Flut an wissenschaftlicher Literatur und unter dem Eindruck stagnierender, wenn nicht gar rückläufiger Beschaffungs- und Personaletats, stehen Bibliotheken heute mehr denn je vor einer großen Herausforderung: Sie müssen ihre begrenzten finanziellen und personellen Ressourcen möglichst wirtschaftlich und zielgruppenorientiert verwenden, ohne jedoch den Service für ihre Benutzer einzuschränken.

Auf den ersten Blick scheint dieser Spagat kaum zu bewältigen. Wie in anderen öffentlichen und privaten Bereichen, lassen sich auch in Bibliotheken nur wenige Kostenfaktoren spürbar verringern, ohne dass sich dies umgehend negativ auf den Service am Kunden, sprich Bibliotheksbenutzer, auswirkt. Um diesem Dilemma zu entgehen, müssen Alternativen gefunden, müssen neue Wege aufgezeigt und beschritten werden.

Im Folgenden soll am Beispiel der Einführung der digitalen Verarbeitung des Neuerscheinungsdienstes der Deutschen Nationalbibliothek (DNB) an der Badischen Landesbibliothek (BLB) aufgezeigt werden, dass es sich durchaus lohnen kann, im Bereich der Einarbeitung von Kauf- und Pflichtliteratur langjährig etablierte und ausdifferenzierte Geschäftsgänge zu überprüfen und zu reformieren.

1. Der alte Geschäftsgang

Wie in anderen großen wissenschaftlichen Bibliotheken wuchs auch in der BLB über viele Jahre ein filigraner Geschäftsprozess für die Erwerbung neuer Printmedien. Bei dieser Tätigkeit wurde zur Sichtung des aktuellen Buchangebotes, neben anderen Bestellunterlagen (Verlagsprospekte, Anschaffungsvorschläge, Rezensionen etc.), vor allem der gedruckte Neuerscheinungsdienst der DNB genutzt. Der Neuerscheinungsdienst wurde in Form von Karteikarten (sogenannte Titelkarten) bezogen. Der Geschäftsprozess wurde über die Jahre so stark verfeinert, dass er alle Bedürfnisse der beteiligten Personen erfüllte. Zugleich nahm seine Komplexität jedoch immer mehr zu, sodass bereits die Sortierung und Verteilung der Titelkarten an die beteiligten Personenkreise (Fachreferenten, Kauf- und Pflichtmitarbeiter) einen immer höheren Arbeitsaufwand erforderte und nur von wenigen Spezialisten überblickt und beherrscht wurde.

Bevor die eigentliche Bearbeitung der Titelkarten in der Abteilung Medienbearbeitung beginnen konnte, mussten diese Titelkarten zunächst ein komplexes Sortierschema durchlaufen. Die BLB ist neben ihren Aufgaben als wissenschaftliche Allgemeinbibliothek auch für das umfassende Sammeln von Pflichtexemplaren des badischen Landesteiles von Baden-Württemberg zuständig[1]. Daher mussten zunächst die vom baden-württembergischen Pflichtexemplargesetz erfassten Titelkarten der badischen Neuerscheinungen aussortiert und direkt an die Mitarbeiter des Pflichtteams der Medienbearbeitung weitergeleitet werden. In diesem Arbeitsschritt wurden neben den badischen bzw. württembergischen Erscheinungsorten noch weitere Kriterien berücksichtigt, wie beispielsweise eine automatische Lieferung von Verlagen (gegebenenfalls auch an die Schwesterbibliothek in Stuttgart, welche die gelieferten Medien dann unaufgefordert an die BLB weiterleitet).

Diese Sortierschritte mussten für die wöchentlichen Lieferungen auf mühsame Weise von Hand durchgeführt werden und banden somit in erheblichem Maße die wertvolle Arbeitszeit mehrerer hoch qualifi-

1 Sie teilt sich diese Aufgabe mit der Württembergischen Landesbibliothek in Stuttgart, welche den württembergischen Landesteil abdeckt

zierter Mitarbeiter. Der größte Teil der anfallenden Arbeit entfiel auf die Aussonderung der Pflichtliteratur-Titelkarten und die weitere Aufteilung der übrigen Karten nach verschiedenen formalen Kriterien (Monographien, Medienkombinationen, AV-Materialien usw.). Lediglich das umfassende Spezialwissen eines einzelnen, sehr erfahrenen Mitarbeiters ermöglichte es, die durchschnittlich 4.000 wöchentlich anfallenden Titelkarten einigermaßen zügig zu bearbeiten. Bei Krankheit oder urlaubsbedingter Abwesenheit dieses quasi unverzichtbaren Spezialisten kam es in der Folge zu unvermeidlichen Verzögerungen im weiteren Erwerbungsgeschäftsgang, welche sich insgesamt negativ auf die Bearbeitungs- und Bereitstellungszeiten der Neuanschaffungen auswirkten.

Nach der Trennung der Titelkarten der Bereiche Pflicht und Kauf wurden letztere an die zuständigen Fachreferenten weitergegeben. Sie hatten die fachliche Kaufentscheidung zu treffen. Zu bestellende Titel wurden dann von den Fachreferenten an die Erwerbungsleitung weitergegeben, welche wiederum die Kaufentscheidung bestätigte oder ablehnte. Erst nach der Genehmigung durch die Erwerbungsleitung wurden die Titelkarten an die Kaufteams der Abteilung Medienbearbeitung übergeben. Dort wurden die aufgedruckten Titelangaben zur Einleitung des Beschaffungsvorganges weiter verarbeitet.

Danach hatten die Titelkarten ihren Zweck erfüllt; jedoch wurden sie von so manch sparsamem Bibliothekar auf ihrer unbedruckten Rückseite als Notizzettel weiterverwendet.

Dass dieses Verfahren zur Erfassung und Sichtung der Neuerscheinungen im Printmedienbereich – obwohl bis dato alternativlos – selbstverständlich nicht optimal war, kann aus dem (vereinfacht) dargestellten Abriss des alten Geschäftsganges unschwer abgeleitet werden. Noch deutlicher werden diese Mängel, wenn man sich die anfallenden jährlichen Kosten vergegenwärtigt:

Die durchschnittliche Zahl der Titelkarten einer Wochenlieferung betrug ca. 4.000; somit galt es also jährlich rund 210.000 Titelkarten zu sichten und manuell zu sortieren. Die Kosten für das Jahresabonnement beliefen sich im Jahr 2009 auf ca. 14.000 €. Der personelle Aufwand für das bloße Sichten, Sortieren und Weiterleiten kann auf der Basis der Stundenentgelte der Vollzeitäquivalente aller beteiligten Mitarbeiter auf ca. 10.000 € geschätzt werden. Addiert man hierzu die verdeckten, kaum ermittelbaren Kosten für den recht umständlichen Weitertransport und des sonstigen Umgangs mit den vielen Titelkarten bei der Einarbeitung in bibliothekarische Fachanwendungen (Erwerbungssystem, elektronische Kataloge) hinzu, kann man durchaus von 30.000 bis 40.000 € jährlichen Gesamtkosten ausgehen.

2 Planung und Realisierung

Im Rahmen weitgehender organisatorischer Umstrukturierungen zur Straffung der internen Geschäftsgänge wurden in der BLB die Probleme beim Umgang mit dem konventionellen Neuerwerbungsdienst der DNB Mitte des Jahres 2009 erneut diskutiert. Nachdem die Direktion des Hauses die grundsätzliche Entscheidung zur Vereinfachung und teilweisen Automatisierung dieses Teils der Erwerbungsgeschäftsgänge getroffen hatte, wurde eine Arbeitsgruppe aus erfahrenen Mitarbeiterinnen und Mitarbeitern aller relevanten Bereiche ins Leben gerufen. Diese Arbeitsgruppe wurde beauftragt, eine Alternative für den bisherigen Erwerbungsgeschäftsgang zu erarbeiten. Zielvorgabe war, die vergleichsweise hohen Kosten für das Jahresabonnement einzusparen und die Personalkosten beim aufwändigen Sortieren der Pflicht- und Kauf-Titelnachweise deutlich zu reduzieren. Darüber hinaus sollten weitere Synergien bei der späteren Verarbeitung der Titelinformationen erschlossen werden. Entscheidende Voraussetzung hierfür war die Abkehr von der Nutzung der gedruckten Titelkartenausgabe des DNB-Neuerscheinungsdienstes.

Als Alternative zum Bezug der gedruckten Titelkarten der DNB bot sich die elektronische Ausgabe der Datenlieferung der Neuerscheinungen an. Sie deckt ebenfalls das gesamte Spektrum des konventionellen Neuerscheinungsdienstes ab, ist in vielen maschinenlesbaren Formaten verfügbar und kann die papiergebundene Lösung vollständig ersetzen. Voraussetzung hierfür ist jedoch ein IT-gestütztes Verfahren, welches einerseits die automatische Auswertung dieser Daten ermöglicht und andererseits die Ergebnisse dieser Auswertung sinnvoll unterstützend in den Ablauf der bibliothekarischen Geschäftsgänge einbringt.

Für die tägliche Arbeit mit dem Neuerscheinungsdienst stellt das Bibliotheksservice Zentrum (BSZ) für die dem Südwestverbund angeschlossenen Bibliotheken bereits seit längerem eine webbasierte Lösung bereit. Sie ermöglicht die Recherche in den Neuerscheinungen und den Export ausgewählter Titelsätze. Darüber hinaus besteht die Möglichkeit, den Neuerscheinungsdienst über das BSZ als Datenlieferung[2] entweder als wöchentliche Komplett- oder als Teillieferung im MAB-Format zu beziehen. Dies wird z.B. von den Fachreferenten in einigen Bibliotheken des Landes genutzt, welche sich vom BSZ gezielt mit den Neuerscheinungen ihrer Fachgebiete versorgen lassen. Allerdings müssen diese Teil-Datenlieferungen konvertiert und für eine Auswertung durch den Fachreferenten aufbereitet werden, da sie als komprimierte CSV-Dateianhänge direkt per Mail an die Fachreferenten ausgeliefert werden. Vielerorts behilft man sich hierzu mit dem Tabellenkalkulationsprogramm MS Excel, das CSV-Dateien einigermaßen lesbar darstellen kann. Die Weiterleitung von zur Kaufentscheidung herangezogenen Titelinformationen an die Erwerbungsabteilungen stellt hier ein besonderes Problem dar. Es wird vielerorts durch einfaches „Copy and Paste" in eine E-Mail an den Erwerber gelöst.

Schnell wurde klar, dass weder die (sehr günstige) vom BSZ angebotene Webbasierte Lösung, noch der E-Mail-Bezug ausgewählter Sachgruppen für die Fachreferenten die notwendige Granularität aufwies oder gar eine Möglichkeit zur Automatisierung des Auswahlprozesses bot – mithin also die wichtigsten Voraussetzungen fehlten, diese Dienstleistungen des BSZ ohne aufwändigste Modifikationen an die abteilungsinternen Geschäftsgänge einer Pflichtexemplarbibliothek anpassen zu können. Für das Vorhaben der BLB kam daher nur die Verwendung der originären elektronischen DNB-Datenlieferungen in Frage. Der Bezug der elektronischen Ausgabe steht grundsätzlich jeder Mitgliedsbibliothek des Südwestdeutschen Bibliotheksverbundes (SWB) – nach vorheriger Zustimmung durch die DNB und das BSZ – im Rahmen eines Linking-Bezugs offen.[3]

Basierend auf den Erkenntnissen der hausinternen Arbeitsgruppe wurde ein Konzept zur Umsetzung der Anforderungen und Wünsche entwickelt. Ziel der Planung war es zunächst, den papiergebundenen Geschäftsgang digital abzubilden und – wo möglich – zu vereinfachen. Die Sortierschritte sollten vollständig automatisch ablaufen, wodurch die weitere Bearbeitung der Pflichtliteratur vollständig unabhängig von manuellen Eingriffen werden sollte. Doch das zukünftige Programm sollte nicht nur die Titelkarten digital erfassen und aufbereiten, sondern den Anwendern auch möglichst komfortable Filterfunktionen (beispielsweise nach Lieferwoche oder DDC-Sachgebiet) bieten. So sollte erreicht werden, dass die bibliothekarische Arbeit auch mit dem neuen Programm wie

2 Siehe hierzu auch http://wiki.bsz-bw.de/doku.php?id=v-team:daten:emailabo.

3 Nähere Auskünfte hierzu erteilt Frau Cornelia Katz, Bibliotheksservice-Zentrum Baden-Württemberg, Cornelia.Katz@bsz-bw.de, Tel: 07531/88 3189, Fax: 07531/88 3703.

the world's becoming
swetswise

Alles was Sie für Beschaffung, Abruf und Management Ihrer Abonnements benötigen.

Service für Ihre gedruckten Zeitschriften

- Auslagerung und Vereinfachung operativer Aufgaben
- Regalfertige Lieferung Ihrer Zeitschriften
- Elektronische Lieferscheine und Rechnungen
- Alle Daten online verfügbar

Service für Ihre E-Journals und E-Books

- Katalog mit mehr als 250.000 Zeitschriften
- Katalog mit mehr als 300.000 E-Books
- Integrierte ERM-Lösung
- Back Office Service, der Sie von der Angebotserstellung über Lizenzverhandlungen und Erneuerungen bis zur Bereitstellung Ihrer elektronischen Inhalte unterstützt.

www.swetswise.com

gewohnt strukturiert werden kann.
Resultat dieser Bemühungen ist das Programm ibis e², welches im Auftrag der BLB von 3brains IT entwickelt wurde und im Folgenden detaillierter vorgestellt wird.

3 Die Datenverarbeitung im Hintergrund

Sowohl der Programmablauf als auch das Programm selbst ist zweistufig aufgebaut. Auf der unteren Programmstufe werden die automatischen Vorverarbeitungen durchgeführt; darauf aufbauend erfolgt die manuelle Bearbeitung (Sichtung und Titelauswahl durch die Mitarbeiter) analog zum alten Geschäftsgang. Die Verarbeitung ist dabei auf verschiedene Rechner verteilt: Ein zentraler Datenbankserver übernimmt die Datenanalyse und die Datenaufbereitung; die Arbeit der Mitarbeiter mit dem elektronischen Titelmaterial erfolgt mithilfe einer Arbeitsplatzversion von ibis e².
Die BLB bezieht den Neuerscheinungsdienst als Linking-Partner über das BSZ, welches neue Datenpakete auf einem passwortgeschützten Server zum Download bereitstellt. Die Prüfung auf neue Datenpakete erfolgt täglich programmgesteuert. Konnten neue Datenpakete vom BSZ abgeholt werden, wird automatisch die weitere Verarbeitung angestoßen.

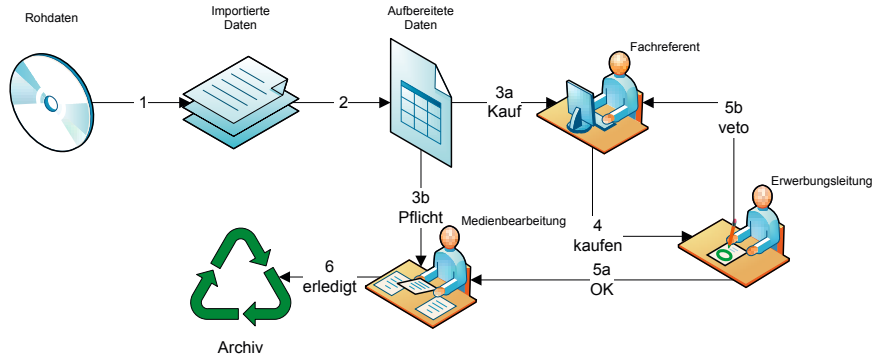


Abbildung 1: Die Verarbeitungskette von der Datenlieferung bis zur Archivierung.

Wie in Abbildung 1 zu sehen ist, werden zunächst die im MAB-Format gelieferten Daten aufbereitet und analysiert (1,2). Danach erfolgt die oben geschilderte Sortierung, welche programmseitig in wenigen Minuten erfolgen kann. Die gewählte Umsetzung ermöglicht eine flexible Anpassung der Sortierkriterien und erlaubt es, Zuständigkeiten und Sortierkriterien schnell anzupassen. Ebenso können unerwünschte Datensätze ausgeblendet und Zusatzinformationen (Titeldetails aus anderen Katalogen oder Lieferantenverzeichnissen) zu einzelnen Datensätzen ermittelt werden.
Nach diesem ersten Verarbeitungsschritt steht den Mitarbeitern der Medienbearbeitung bereits die gesamte Pflichtliteratur zur Verfügung (3b) – und das unabhängig von urlaubsbedingten oder krankheitsbedingten Verzögerungen. Gleichzeitig erzeugt ibis e² eine tabellarische Darstellung der Datenlieferung, um eine effiziente Überprüfung des Sortierergebnisses zu ermöglichen. Die Datensätze, die nicht als Pflichtliteratur eingestuft werden (sogenannte Kauftitel), werden an die zuständigen Fachreferenten weitergeleitet (3a), sodass auch in diesem Bereich das neue Datenmaterial umgehend bereitsteht. Die Schritte (4) bis (6) stellen den oben bereits geschilderten Geschäftsprozess der Fachreferenten, der Erwerbungsleitung und der Medienbearbeitung vereinfacht dar.

4 Neuerscheinungen im neuen Kleid – ibis e² im Überblick

Für die Arbeit der Fachreferenten war es zunächst nötig, die Datensätze so darzustellen, dass das große Liefervolumen in kurzer Zeit gesichtet werden kann. Insbesondere für den Bereich der Medienbearbeitung mussten jedoch die bibliografischen Informationen detailliert aufbereitet werden, um weitere Recherchen in anderen Systemen zu vermeiden und die Aufnahme von Bestellungen effizient zu gestalten. ibis e² deckt beide Perspektiven ab, indem es eine tabellarische Übersicht und eine Detailansicht der Einzeltitel bereitstellt. Sämtliche Geschäftsvorfälle können über beide Sichten abgewickelt werden, sodass die Benutzer ganz nach Belieben mit ihrer favorisierten Ansicht arbeiten können. Umfangreiche Tastaturkürzel ermöglichen darüber hinaus ein durchgängig zügiges Bearbeiten der Titelauswahl. Die Anpassung der Größe der Bildschirmschrift ist individuell einstellbar und unterstützt entscheidend die Ergonomie der Arbeitsplätze.

4.1 Die Lieferungen im Überblick

Abbildung 2 zeigt die tabellarische Übersicht aus der Perspektive eines Fachreferenten. Der Anwender entscheidet individuell, welche Spalten in welcher Reihenfolge dargestellt werden sollen. Zusammen mit den Sortier- und Filterfunktionen ergibt sich ein mächtiges Werkzeug zum Sichten der Datenlieferungen. Neben den bibliographischen Daten lassen sich auch programminterne Zusatzinformationen einblenden, wie beispielsweise der Name des bestellenden Fachreferenten. Über die Filterfunktionen in der Funktionsleiste lassen sich einzelne Wochenlieferungen und/oder Sachgebiete auswählen.
Die farbige Markierung der Datensätze kennzeichnet ihren Status. Neben dem Ausgrauen bereits erledigter Einträge versieht ibis e² Datensätze mit einer orangenen Veto-Markierung, wenn die Erwerbungsleitung eine Kaufentscheidung abgelehnt und den Datensatz an den bestellenden Fachreferent zurückgeleitet hat. Abweichend vom ursprünglichen Geschäftsgang ist die rote Hinterlegung; sie kennzeichnet Datensätze, die zugleich auf mehrere Sachgebiete aufgeteilt wurden. Vor der Einführung von ibis e² wurden Titelkarten aus überlappenden Sachgebieten von Fachreferent zu Fachreferent weitergegeben, bis alle möglichen Zuständigkeiten berücksichtigt wurden. Heute werden die betroffenen Datensätze zeitgleich allen (möglicherweise) Zuständigen zugeteilt, sodass keine Wartezeiten entstehen.

4.2 Eine Titelaufnahme im Detail

Abbildung 3 zeigt die Detailansicht eines Datensatzes. Unter der klassischen ISBD-Ansicht der Titelaufnahme befindet sich ein Kommentarbereich, in den alle bearbeitenden Personen einen Freitextkommentar einfügen können, um programmexterne Rückfragen oder Notizzettel überflüssig zu machen. Auf der rechten Seite können im Funktionsbereich bibliothekarische Zusatzvermerke hinterlegt werden, die Standardentscheidungen wie besondere Aufstellungsorte oder statistische Kennzeichnungen abdecken. Darüber hinaus können einzelne Datensätze an andere Fachreferenten oder Teams der Medienbearbeitung weitergereicht werden, beispielsweise bei einer thematischen Überschneidung der Sachgebiete oder bei einer fehlerhaften Eingruppierung eines Pflichtexemplars.
In der rechten oberen Hälfte des Bildschirms finden sich Verknüpfungen zu externen Zusatzinformationen, die direkt für den aktuell angezeigten Titel geöffnet werden. Anwender haben so die Möglichkeit, häufig genutzte Zusatzinformationen direkt abzurufen. Dies umfasst nicht nur das Öffnen von elektronischen Ressourcen oder Zusatzinformationen der DNB und der Verleger. Vielmehr können konkrete Recherchen in Web-OPACs durchgeführt werden,

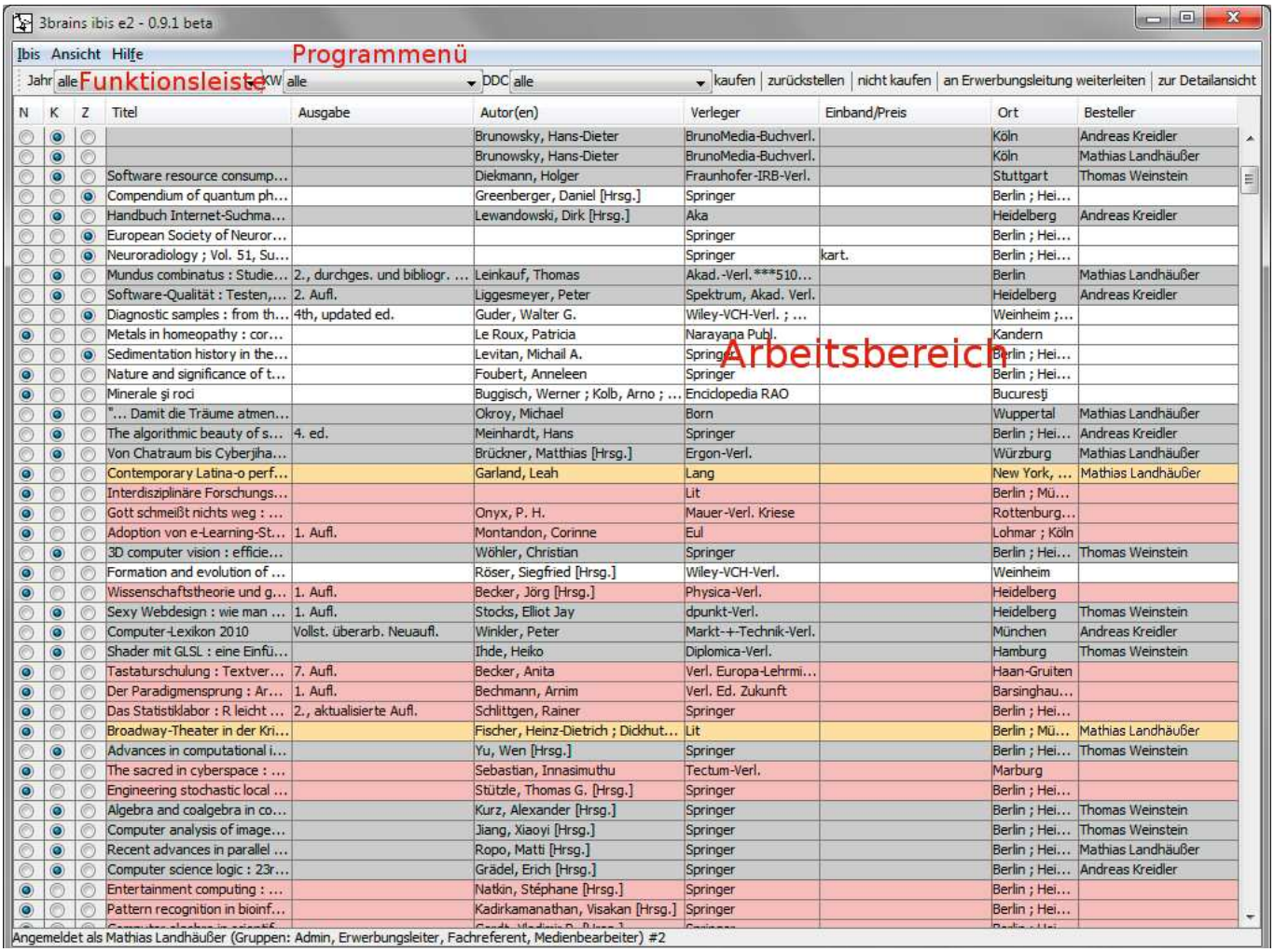


Abbildung 2: Listenansicht von ibis e².

um Kaufentscheidungen von Fachreferenten zu untermauern oder die Verfügbarkeit eines Titels in der Region zu prüfen. Selbstverständlich stehen diese Funktionen auch der Erwerbungsleitung und der Medienbearbeitung zur Verfügung.

5 Administrative Betreuung

Der Betreuungsaufwand für die IT-Abteilung wurde möglichst klein gehalten. Lediglich für das Registrieren eines Benutzerkontos wird ein Systemadministrator benötigt. Durch die Kopplung von ibis e² an einen Verzeichnisdienst (beispielsweise an ein MS Active Directory oder einen LDAP-Dienst) konnte jedoch auch hier der Aufwand minimiert werden: Der Systemadministrator muss lediglich ein (oder mehrere) Recht(e) über den Verzeichnisdienst vergeben. Der Abgleich mit ibis e² erfolgt dann automatisch. Nach der Freischaltung sind alle weiteren Anpassungen entweder Aufgabe der ibis e² Administratoren (beispielsweise Zuweisen von Sachgebieten zu Fachreferenten) oder liegen in den Händen des Anwenders selbst (wie beispielsweise die Konfiguration der Listendarstellung).

Zu den Aufgaben der ibis e² Administratoren zählen unter anderem das Pflegen der diversen Sortierkriterien sowie die Verwaltung der Sachgebiete und deren Zuordnung zu den einzelnen Fachreferenten. Alle derartigen Aufgaben können mithilfe einer einfachen Benutzeroberfläche direkt ausgeführt werden, ohne dass technisches Fachwissen benötigt wird. Diese Administrationsoberfläche ist Teil der Arbeitsplatzversion von ibis e², sodass keine zusätzlichen Installationen notwendig sind.

Die Authentifizierung an ibis e² erfolgt mit demselben Benutzernamen, den der Anwender auch in seiner gewohnten Windows-Arbeitsumgebung verwendet. Durch diese enge Integration kann weiterer Verwaltungsaufwand für die Pflege von Benutzerkonten vermieden werden. Darüber hinaus ermöglicht die Authentifizierung eine Personalisierung des Programms unabhängig vom Stamm-Arbeitsplatz eines Mitarbeiters. Ob ein Anwender mit ibis e² in seinem Büro arbeitet oder bei Spät- und Samstagsdiensten an den Auskunftstheken des Info-Zentrums oder des Lesesaals, spielt daher keine Rolle.

Abschließend ist die einfache Installation der Arbeitsplatzversion von ibis e² hervorzuheben. Die Auslieferung des Programmes erfolgt als MSI-Paket, weswegen die Installation auf den Arbeitsplatz-Rechnern der Mitarbeiter völlig problemlos über entsprechende Gruppenrichtlinien im lokalen Active Directory durchgeführt werden kann. Die Aktualisierung der Anwendung erfolgt web-basiert und zentral gesteuert, sodass die Systemadministratoren im Allgemeinen nicht eingreifen müssen. So wird der personelle administrative Aufwand minimiert und damit auch die bei der Anwendungsbetreuung anfallenden Kosten.

6 Ergebnisse der Pilotphase und Ausblick

Nach der Planungs- und Umsetzungsphase Mitte bis Ende des Jahrs 2009 wurde ibis e² zum Januar 2010 in der BLB eingeführt und konnte bereits nach einer sehr kurzen Eingewöhnungsphase in den Produktivbetrieb überführt werden. Der ursprüngliche Bezug der Titelkarten wurde ebenfalls zum Jahresbeginn 2010 beendet und durch das Abonnement der elektronischen Datenlieferung ersetzt. Einige wenige Medien aus vergangenen Lieferungen wurden nach wie vor mithilfe der vorhandenen Titelkarten bearbeitet – die neuen Bestellungen ab 2010 werden jedoch ausschließlich mit ibis e² bearbeitet.

Im ersten Halbjahr des Jahres 2010 wurden in 27 Lieferungen 110.752 Datensätze aufbereitet. Aus dieser Menge konnten 12.406

(ca. 11%) als Pflichtliteratur identifiziert und automatisch aussortiert werden. Ein Großteil dieser Titel wurde bereits von der Medienbearbeitung bearbeitet.

6.1 Kosteneinsparungen

Einen großen Vorteil der digitalen Verarbeitung stellen die geringen laufenden Kosten dar. Mit dem Ende des Bezugs der Titelkarten in Papierform entfällt ein großer Teil der Gesamtkosten. Die ansonsten notwendige IT-Infrastruktur war in der BLB bereits vorhanden und konnte ohne direkte Kosten mitbenutzt werden. So konnte nicht nur der Geschäftsgang beschleunigt, sondern das Einsparpotential für die Beschaffung, Sortierung und Verteilung der Datensätze voll ausgenutzt werden. Betrachtet man lediglich die direkten, monetären Einsparungen, so amortisieren sich die Entwicklungskosten nach ca. einem Jahr.

6.2 Beschleunigung der Sortierprozesse

Die offensichtliche Beschleunigung der bereichsspezifischen Sortierprozesse ist zweifellos das hervorstechendste Leistungsmerkmal von ibis e². Die Sortierung erfolgt heute umgehend nach Eingang neuer Lieferungen und die gesamte Vorverarbeitung erfolgt binnen weniger Minuten.
Durch die integrierte Verwaltungsoberfläche lassen sich die Grundparameter für alle Sortieroperationen feingranular für alle Mitarbeiterbereiche anpassen und die weiteren Geschäftsgänge somit effektiv steuern. Urlaubszeiten und Krankenstand beeinflussen und verzögern diese Arbeit nicht mehr, wovon der gesamte nachfolgende Geschäftsgang profitiert.

6.3 Zukünftige Weiterentwicklung

Bis zur aktuellen Programmversion bildet ibis e² hauptsächlich den alten, papiergebundenen Geschäftsgang ab. Dies erleichterte die Akzeptanz des Verfahrens in allen direkt angeschlossenen Bereichen sehr.
In enger Zusammenarbeit mit den betroffenen Abteilungen wird derzeit geprüft, wie sich die Arbeit mit ibis e² weiter optimieren lässt und die Bedienung verbessert werden kann. Ebenso konnten im Zusammenhang mit dem neuen, abgewandelten Geschäftsgang bereits Verbesserungsmöglichkeiten aufgezeigt werden. So wird beispielsweise erwogen, EAN-Barcodes auf den Medien auszuwerten, um die Eingangskontrolle der Pflicht- und Kauftitel zu vereinfachen.

6.4 Anbindung an Bibliothekssysteme

Eine Schnittstelle zu Bibliothekssystemen anderer Hersteller wurde bis dato noch nicht entwickelt. Geplant ist jedoch ein automatischer Abgleich zwischen ibis e² und dem Erwerbungsmodul von (integrierten) Bibliothekssystemen. Insbesondere im Bereich der Pflichtliteratur, jedoch auch bei Kauftiteln, könnte damit ein Teil der manuellen Arbeiten im Bereich der Medienbearbeitung erledigt werden. Auf die konkrete Umsetzung dieses Vorhabens in der BLB wurde bis jetzt bewusst verzichtet. Zum einen sollten zunächst erste Erfahrungen mit der automatischen Verarbeitung von Bestellunterlagen wie dem DNB-Datendienst gewonnen werden. Zum anderen war zu Beginn der ibis e² Entwicklung noch nicht klar, welches Bibliothekssystem in Zukunft in der BLB eingesetzt werden würde. Nachdem diese Entscheidung nun zu Gunsten von aDIS/BMS von |a|S|tec| gefallen ist, streben 3brains IT und |a|S|tec| an, die Funktionalitäten von ibis e² für aDIS/BMS zu erschließen.

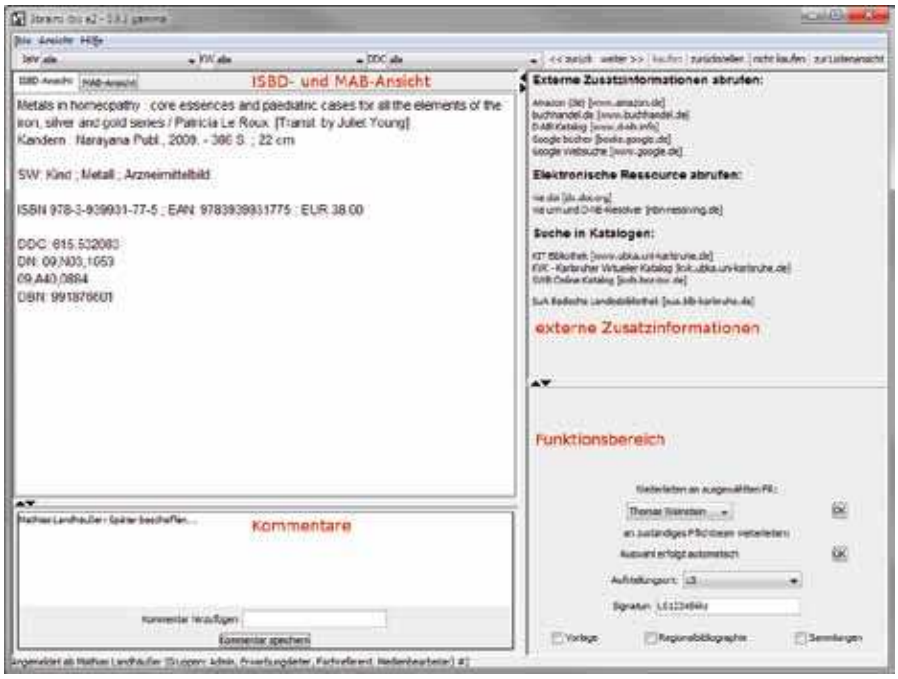


Abbildung 3: Die Detailansicht von ibis e².

Fazit

Das Projekt als solches kann auch in anderen Bibliotheken vergleichbarer Größe ein Anstoß zur Weiterentwicklung sein. Viele – über Jahre gewachsene und somit auch sehr ausdifferenzierte – Geschäftsprozesse können mithilfe von spezialisierten Softwarelösungen wie ibis e² vereinfacht werden. Insbesondere ist dann ein großes Verbesserungspotential zu erwarten, wenn diese Prozesse Medienbrüche aufweisen und somit viel manuelle Arbeit erfordern. Auch bereits digital ablaufende Prozesse bieten nach wie vor Verbesserungsmöglichkeiten, die erst durch eine genaue, objektive Analyse zu Tage treten.
Mithilfe von ibis e² konnte im ersten Halbjahr 2010 bereits ein erheblicher Teil der anfallenden Arbeit im Zusammenhang mit dem Neuerscheinungsdienst in der BLB automatisiert werden. Die Erfahrungen, die in der Pilotphase gewonnen werden konnten, fließen in die stete Weiterentwicklung des Programms ein. Neben Funktionserweiterungen wird eine Verbesserung der Ergonomie angestrebt und es wird geprüft, ob und wie der neue elektronische Geschäftsgang weiter verbessert werden kann.
Die tatsächlichen Einsparungen und Verbesserungen – neben den offensichtlichen monetären Größen – machen sich im Alltagsbetrieb bereits allenthalben bemerkbar. Bereits jetzt zeichnet sich ab, dass die beschleunigte Auswertung des DNB-Neuerscheinungsdienstes durch ibis e² dazu führt, dass Neuerscheinungen aus dem Kauf- und Pflichtbereich früher in den Bibliotheksbestand integriert werden und damit die Nutzer der Bibliothek direkt durch den Einsatz von ibis e² profitieren.

AUTOREN

Dipl.-Bibl. Bernt Walter
Sachgebietsleiter
Informationstechnologie
Badische Landesbibliothek
Karlsruhe
Postfach 1429
76003 Karlsruhe
walter@blb-karlsruhe.de
www.blb-karlsruhe.de

Dipl.-Inform.Wirt Mathias Landhäußer
3brains IT UG
(haftungsbeschränkt)
Rosenstraße 68
76287 Rheinstetten
landhaeusser@3brains.de
www.3brains.de

Sachliche Erschließung und die neue Suchmaschine des Österreichischen Bibliothekenverbundes

Markus Knitel, Josef Labner und Otto Oberhauser

Mit der Einführung der neuen Suchmaschine Primo im Österreichischen Bibliothekenverbund wurde eine Neubewertung und Neuausrichtung der Anzeige, Indexierung und der erstmals möglichen Facettierung sacherschließender Elemente notwendig. Nach einem kurzen Überblick über die im Verbund zum Einsatz kommende Software sowie über Umfang und Art der inhaltlichen Erschließung im zentralen Katalog wird detailliert auf die angestrebte Umsetzung derselben in der neuen Suchmaschinenumgebung eingegangen, wobei der verbalen Erschließung durch RSWK-Ketten als wichtigste Form der intellektuellen Erschließung erhöhte Aufmerksamkeit zukommt. In einem zweiten Teil wird die konkrete technische Umsetzung in Primo beleuchtet, um abschließend einen Ausblick auf zukünftige Entwicklungen sowie projektierte Verbesserungen zu geben.

Subject indexing and the new search engine of the Austrian Library Network

Recently, the Austrian Library Network implemented its new search engine, Primo. Hence, a reappraisal and realignment was needed for the display, indexing and – for the first time – faceted navigation of all data elements relevant for subject access. After a short overview of the software and the status of subject indexing and classification in the union catalogue of the consortium, the article deals in detail with the intended implementation in the new search engine. The main emphasis is on subject headings (according to the German subject cataloguing rules) as the most important means of intellectual indexing. In the second part, the actual implementation in Primo is outlined. Finally, future developments and improvements are discussed.

Sujet d'indexation et le nouveau moteur de recherche du réseau autrichien de bibliothèques

L'article traite la configuration du catalogage matières pour le nouveau moteur de recherche Primo du réseau autrichien de bibliothèques. Après un court tour d'horizon du réseau autrichien et du logiciel utilisé suit une analyse de l'utilisation des différentes formes du catalogage matières, c'est à dire des mots-sujet, des classifications et des résumés analytiques dans le catalogue central. Ensuite les auteurs présentent leurs idées pour l'emploi le mieux possible de ces éléments dans l'affichage de notices complètes et courtes, l'indexation et les facettes pour ensuite expliquer la réalisation téchnique dans Primo. Finalement l'article s'occupe de développements et de perfectionnements futures.

1 Einleitung

Der Österreichische Bibliothekenverbund ist der nationale Verbund universitärer, wissenschaftlicher und administrativer Bibliotheken des Landes. Über 75 Bibliotheken von ca. 60 Trägerinstitutionen nehmen aktiv durch Online-Katalogisierung an diesem Verbund teil, darunter die Österreichische Nationalbibliothek sowie nahezu alle bundesstaatlichen Universitätsbibliotheken. Von weiteren 310 Einrichtungen werden die Zeitschriftenbestände sowie Buchbestände aus einem früheren Gesamtkatalog nachgewiesen. Seit Anfang 1999 wird verbundweit das integrierte Bibliothekssystem *Aleph 500* des Herstellers Ex Libris eingesetzt. Unter diesem System werden sowohl der *Österreichische Verbundkatalog*,[1] der von der Verbundzentrale OBVSG[2] betriebene Gesamtkatalog, als auch die lokalen Online-Kataloge der einzelnen Verbundbibliotheken – einige davon auch mit dem kleineren System *Alephino* – frei zugänglich angeboten.

Schwerpunktmäßig umfasst der Verbundkatalog Literatur ab dem Erscheinungsjahr 1980. Von den meisten Verbundteilnehmern wurde und wird auch ältere Literatur durch Nachbearbeitung einzelner Titel oder im Rahmen größerer Retrokatalogisierungsprojekte erfasst. Zunehmend werden auch elektronische Dokumente nachgewiesen. Als Service zur Anreicherung des Verbundkataloges mit letzteren wurde von der OBVSG vor mehreren Jahren das System *eDOC* eingerichtet, das Objekttypen wie Inhaltsverzeichnisse, Abstracts, Rezensionen, Umschlagbilder und Klappentexte sowie Volltextdokumente unterstützt und mit dem Verbundkatalog automatisch verlinkt. Die Recherche[3] in eDOC erfolgt mittels der Suchmaschine *Swish-e*.

Der Gesamtkatalog des Österreichischen Bibliothekenverbundes weist mit Stand Ende Juni 2010 rund 7,7 Mio. Titel, 14,5 Mio. Exemplare und 0,7 Mio. Zeitschriften-Bestandsangaben nach. In eDOC sind zum selben Zeitpunkt rund elektronische 420.000 Dokumente enthalten.

Das jüngste Produkt in der Softwarepalette des Verbundes ist die neue Retrieval-Software *Primo,* ebenfalls aus dem Hause Ex Libris. Die Einführung eines neuen Retrievalinstrumentes war lange überfällig. Der herkömmliche Aleph-OPAC, der vom Hersteller schon seit geraumer Zeit nicht mehr weiterentwickelt wurde, kann benutzerseitigen Erwartungen und Gepflogenheiten, die aus dem täglichen Umgang mit bekannten Volltextsuchmaschinen resultieren, nicht adäquat begegnen. Aus Statistiken und Beobachtungen ist bekannt,

1 http://opac.obvsg.at/acc01 [alle im folgenden zitierten Web-Adressen überprüft am 06.07.2010].

2 Die Österreichische Bibliothekenverbund und Service GmbH (OBVSG) ist ein Unternehmen im Eigentum der Republik Österreich mit Sitz in Wien (http://www.obvsg.at).

3 http://media.obvsg.at/

```
-<record>
 +<control></control>
 -<display>
    <type>article</type>
    <title>Die Paradoxie des Entscheidens</title>
    <creator>Luhmann, Niklas</creator>
  -<ispartof>
      In: Paradoxien der Entscheidung / Friedrich Balke ...(Hg.). - Bielefeld, 2004. - S. 17-
    </ispartof>
    <language>ger</language>
    <source>ACC_aleph</source>
    <availlibrary>$$lACC_B_UBW$$Scheck_holdings$$LACC_B_UBW</availlibrary>
    <availlibrary>$$lACC_B_UBI$$Scheck_holdings$$LACC_B_UBI</availlibrary>
    <lds02>AC04113921</lds02>
    <lds30>AC05094114</lds30>
    <lds31>!!!-UBI-!!!</lds31>
    <availpnx>available</availpnx>
  </display>
 +<links></links>
 +<search></search>
 -<sort>
    <title>DIE PARADOXIE DES ENTSCHEIDENS</title>
    <creationdate>2004</creationdate>
  </sort>
 +<facets></facets>
 +<dedup></dedup>
 +<frbr></frbr>
 +<delivery></delivery>
 +<ranking></ranking>
 +<addata></addata>
</record>
```


Abbildung 1

dass die anspruchsvolleren Funktionalitäten (Index blättern, Expertenmodus) nur von einer kleinen Minderheit qualifizierter User genutzt werden, während die große Mehrheit die simple Stichwortrecherche im Suchschlitz der „Alle-Felder-Suche" bevorzugt. Dass dieses Suchverhalten nur allzu oft zu unbefriedigenden Ergebnissen führt – insbesondere da der OPAC keine sinnvolle Relevanzsortierung der Treffer erlaubt[4] –, ist ein manifester Nachteil, den es auszugleichen galt. Darüber hinaus war in den letzten Jahren im Bereich der lokalen Bibliotheken, aber auch des Gesamtkataloges (vgl. das oben erwähnte eDoc) eine zunehmende Zersplitterung des Informationsangebotes in unterschiedliche Sucheinstiege, Datenbanken, Teilkataloge oder Repositorien feststellbar. Die Kombination all dieser Angebote in einer einheitlichen, einfachen Oberfläche war folglich eine der Kernanforderungen an einen neuen Sucheinstieg. Neben dem Relevance Ranking war auch die Verfügbarkeit einer facettierten Navigation[5] ein wichtiges Kriterium im Hinblick auf einen „next-generation library catalog".[6] Die Entscheidung für Primo fiel nach Prüfung durch eine Arbeitsgruppe der Vollversammlung des Verbundes im Spätherbst 2008. Aus Gründen der Kostenökonomie sowie auf Grund des bereits sehr hohen Integrationsgrades des Österreichischen Bibliothekenverbundes entschied man sich damals für den Betrieb einer *einzigen zentralen Primo-Instanz* (angesiedelt bei der OBVSG), die durch mehrere Einrichtungen gemeinsam genutzt werden sollte. Trotz der abzusehenden Nachteile bzw. Einschränkungen bei manchen instanzweiten Konfigurationen wurde davon ausgegangen, dass die Synergieeffekte alle Nachteile überwiegen würden. Inzwischen sind die Kataloge des Verbundes sowie der Universitätsbibliotheken Wien und Innsbruck mit Primo produktiv;[7] weitere lokale Bibliotheken werden in Kürze folgen.[8]

2 Primo

Herzstück von Primo und Basis der Indexierung ist das *Primo Normalized XML* (PNX), ein programminternes, einfach lesbares XML-Format, das in Sektionen („sections") gegliedert ist. In dieses Format werden alle Quelldatensätze auf der Basis bestimmter Normalisierungsregeln transformiert. Dabei werden die MAB-Kategorien bestimmten Tags im PNX zugeordnet, welche ihrerseits wiederum für verschiedene Verarbeitungsschritte herangezogen werden: Während die *display section* anzeigerelevante Informationen enthält (z.B. Autor, Titel oder ISBN), befinden sich in der *search section* die zu indexierenden Anteile des Datensatzes und in der *link section* Links zu Volltexten oder auch zur OPAC-Anzeige.[9] Als Beispiel zeigt *Abbildung 1* einen PNX-Satz mit den aufgeklappten Sektionen „display" und „sort".

Die Quelldaten, in der Regel MAB-XML Datensätze,[10] können für jede Sektion unterschiedlich aufbereitet werden. Das heißt, dass z. B. die Darstellung, die Suchbarkeit und die Facettierung getrennt behandelt werden können. Dass diese Aussage eigentlich sogleich wieder einer Einschränkung bedürfte, wird weiter unten noch deutlich werden. Entgegen gelegentlich geäußerter Befürchtungen macht eine Suchmaschine wie Primo kontrollierte Metadaten keineswegs obsolet – vielmehr kommt die tiefgehende und mühevolle Erschließungsarbeit der letzten Jahre jetzt erst richtig zur Geltung. Auch die in der bisherigen Katalogisierungspraxis ein wenig unterschätzten „festen Felder" (z.B. MAB 050 = Datenträger, 051 = Veröffentlichungsspezifische Angaben zu begrenzten Werken, 052 = Veröffentlichungsspezifische Angaben zu fortlaufenden Sammelwerken, …) gewinnen in der neuen Umgebung an Bedeutung.

Die Suchoberfläche selbst sollte bei alledem so schlicht wie möglich gehalten werden. Letztlich obliegt es aber jeder an der zentralen Primo-Instanz teilnehmenden Institution, sich um ihren Internetauftritt zu bemühen. Die Gestaltungsmöglichkeiten sind vielfältig, wenn auch teilweise mit viel Aufwand verbunden. *Abbildung 2* zeigt als Beispiel die Kurzanzeige mit integrierter Suchmaske in der Version des Verbundes. Am frappierendsten und für den Benutzer wohl am offensichtlichsten ist hier einerseits die Sortierung der Treffer nach Relevanz und andererseits das Gruppieren der gefundenen Titel nach unterschiedlichen Facetten, die ein gezieltes Weiternavigieren und sukzessives Einschränken der Treffermenge in die Richtung des gewünschten Zieles ermöglichen. „Relevanz" ist das voreingestellte Sortierkriterium; alternativ dazu sind „Erscheinungsjahr" und „Bekanntheitsgrad" auswählbar. Mit letzterer Option wird erstmals das Benutzerverhalten in die Treffer-Sortierung einbezogen. Ein Titel kann auf drei Arten an Bekanntheit („popularity") gewinnen: durch Betrachten in der Vollanzeige, Hinzufügen zur Merkliste oder – im Falle der lokalen Bibliotheken – durch

4 Vgl.: Oberhauser, O. (2010). Relevance Ranking in den Online-Katalogen der „nächsten Generation". *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare.* 63 (1/2). 25–35.

5 Vgl. z.B.: Imhof, A. (2006). RSWK/SWD und Faceted Browsing: Neue Möglichkeiten einer inhaltlich-intuitiven Naviagation. *Bibliotheksdienst.* 40 (8/9). 1015–1025; Olson, T. A. (2007). Utility of a faceted catalog for scholarly research. Library hi tech. 25 (4). 550–561.

6 Vgl.: Breeding, M. (2007). Next-generation library catalogs. *Library technology reports.* 43 (4).

7 http://search.obvsg.at/OBV (Verbund); http://usearch.univie.ac.at/ (UB Wien); http://search.obvsg.at/UIB (UB Innsbruck).

8 Zu Primo im OBV vgl. folgende aktuelle Beiträge: Knitel, M. (2010). Architektur und Parametrisierung von Primo im Österreichischen Bibliothekenverbund: Ein Überblick. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare.* 63 (1/2). 56–68; Babitchev, V.; Brandauer, J. (2010). Die konsortiale Primo-Implementierung im OBV. Teil 1: Aktueller Stand, Herausforderungen, Lösungen. Vortrag, *DACHELA* 2010, Wien, 15.–16. Juni 2010. Online: http://www.obvsg.at/fileadmin/files/obvsg/publ/primo-dachela-2010_fin.pdf; Krabo, U. (2010). Die konsortiale Primo-Implementierung im OBV. Teil 2: Integration zusätzlicher Dienste. Vortrag, *DACHELA* 2010, Wien, 15.–16. Juni 2010. Online: http://www.obvsg.at/fileadmin/files/obvsg/publ/primo-dachela-2010_teil2_v1.pdf

9 Primo ist derzeit mittels einer einfachen Verlinkung mit dem jeweils korrespondierenden Aleph-OPAC verbunden. Die im Bereich der lokalen Bibliotheken wichtigen Funktionen wie Bestellung, Einsicht des Benutzerkontos oder Verlängerung ausgeliehener Bücher können bislang nur im herkömmlichen OPAC abgewickelt werden. Dies wird sich mit der neuen Version 3 von Primo ändern.

10 http://www.d-nb.de/standardisierung/formate/mabxml.htm

DABIS.eu
Gesellschaft für Datenbank-InformationsSysteme mbH

Ihr Partner für Archiv-, Bibliotheks- und DokumentationsSysteme

BIS-C 2000

Archiv- und Bibliotheks-InformationsSystem

DABIS.com · alle Aufgaben · ein Team

Synergien: Qualität und Kompetenz
Software: Innovation und Optimierung
Web · SSL · Warenkorb und Benutzeraccount
Lokalsystem zu Aleph-Verbünden

Software · State of the art · Open Source

Leistung	Sicherheit
Standards	Offenheit
Stabilität	Verläßlichkeit
Generierung	Adaptierung
Service	Erfahrenheit
Outsourcing	Support
Dienstleistungen	Zufriedenheit

GUI · Web · Wap · XML · Z 39.50

Archiv		Bibliothek
singleUser	System	multiUser
Lokalsystem		Verbund
multiDatenbank		multiServer
multiProcessing		multiThreading
skalierbar		stufenlos
Unicode		multiLingual
Normdaten		redundanzfrei
multiMedia		Integration

Portale

http://www.vthk.eu
http://www.landesbibliotheken.eu
http://www.bmlf.at
http://www.volksliedwerk.org
http://www.dabis.org:18093

DABIS.com
Heiligenstädter Straße 213
1190 · Wien, Austria
Tel.: +43-1-318 9 777-10
Fax: +43-1-318 9 777-15
eMail: office@dabis.com
http://www.dabis.com

DABIS.de
Herrgasse 24
79294 · Sölden/Freiburg, Germany
Tel.: +49-761-40983-21
Fax: +49-761-40983-29
eMail: office@dabis.de
http://www.dabis.de

Zweigstellen: 61350 · Bad Homburg vdH, Germany / 1147 · Budapest, Hungary / 39042 · Brixen, Italy

Abbildung 2

Anklicken des sogenannten „getIt"-Links, der zur Anzeige der Bestandsinformation in den lokalen OPACs führt. In der Abbildung sind zwei thematische Facetten dargestellt – „Thema" (RSWK-Schlagwörter) und „RVK". Welche Facetten bei der Trefferliste aufscheinen (und in welcher Reihenfolge), kann jede teilnehmende Bibliothek selbst bestimmen. In der Version des Verbundes werden neben den beiden genannten auch die Facetten „Medium", „Form", „Erscheinungsjahr", „Sprache", „Bibliothek", „Autor/in" und „Region" angeboten („Medium" kann grob als Facette für den Materialtyp wie Buch, CD, Artikel etc. betrachtet werden, während die Facette „Form" eher nach Kriterien wie Aufsatzsammlung, Kunstführer, Ratgeber etc. gliedert).

Einem bewährten Prinzip folgend, stehen dem Suchenden grundsätzlich eine einfache und eine erweiterte Suche zur Verfügung. Die *einfache Suche* führt eine Recherche im Index „Alle Felder" durch, während die *erweiterte Suche* eine Beschränkung der Recherche auf bestimmte Indizes, Jahre oder Materialarten erlaubt.

Die beiden eben genannten Hauptfunktionen von Primo und das völlig neue Datenmodell machten es notwendig, den Einsatz und die Präsentation von sacherschließenden Elementen bei der Recherche neu zu überdenken. Was möchte man sichtbar und suchbar, nur sichtbar, oder gar nur suchbar jedoch nicht sichtbar machen? Welche thematischen Facetten sollen angeboten werden? Lohnt sich der Einsatz von Klassifikationen (wenn ja, welcher)? Wie gewichtet man Sacherschließungselemente in der Berechnung der Relevanz? Basis und Ausgangspunkt der Beantwortung dieser Fragen war die sorgfältige Analyse des Status Quo des Einsatzes von sacherschließenden Elementen. Wir beziehen uns im Folgenden immer nur auf die Ebene des Gesamtverbundes.

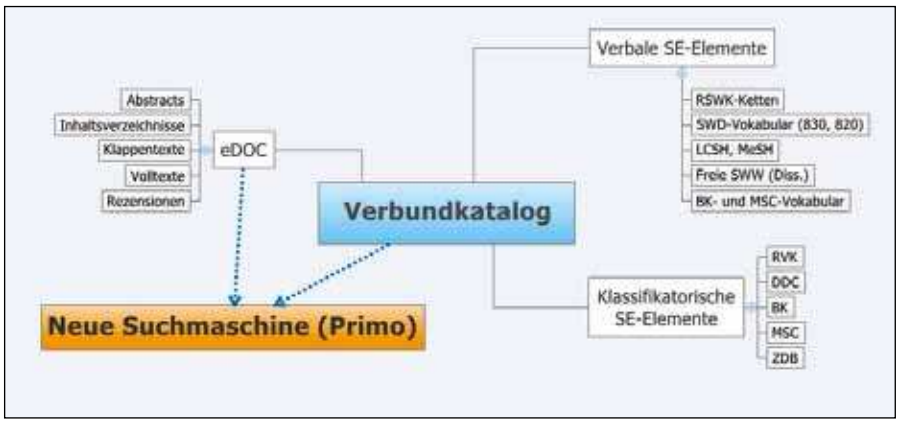


Abbildung 3

3 Sachliche Erschließung im Österreichischen Verbundkatalog

Das gegenwärtig bestehende Erscheinungsbild der sachlichen Erschließung im Österreichischen Bibliothekenverbund ist historisch gewachsen und gleicht – wie in anderen Verbünden auch – ein wenig einem Patchwork-Muster. Es gibt nämlich eine ganze Reihe unterschiedlicher Sacherschließungselemente innerhalb und außerhalb des bisherigen Verbundkatalogs, wie in *Abbildung 3* veranschaulicht wird.

Die **verbale** Sacherschließung, an der sich der überwiegende Teil der Verbundbibliotheken beteiligt, erfolgt seit 1989 – kurz nach der Einführung des ersten automationsunterstützten Verbundsystems – kooperativ nach RSWK und SWD. Heute weisen 38,3% der Titeldatensätze mindestens *eine* Schlagwortkette nach RSWK bzw. SWD auf. Dieser Anteil betrug 2005 noch 44%,[11] ist aber seitdem aufgrund mehrerer großer Retrokonversionsprojekte auf den derzeitigen Stand abgesunken. *Nur eine* Kette weisen 5,3% auf, mindestens 2 Ketten 33%, 3 oder mehr Ketten 19,5%, 4 oder mehr Ketten 8,7%. Durch die Normdatenverknüpfung wird das Vokabular der Schlagwortketten im Aleph-Katalog mit äquivalenten Bezeichnungen (Synonymen) und Alternativformen aus der SWD weiter angereichert. Die sachliche Erschließung nach RSWK / SWD spielt unter allen Sacherschließungselementen bei weitem die größte Rolle.

Die übrigen *verbalen* Sacherschließungselemente, die in den Verbunddaten enthalten sind, sind in quantitativer Hinsicht weniger bedeutend. Der Anteil der im Zuge der Fremddatennutzung (v.a. BNB) übernommenen Library of Congress Subject Headings (LCSH) beträgt nur 1,8%; jener der auf demselben Weg übernommenen Medical Subject Headings (MeSH) liegt noch weit darunter. Beide Datenelemente gelangen jedoch im Aleph-Katalog nicht nur in den Index „Alle-Felder", sondern auch in einen kombinierten Schlagwort-Titelstichwort-Index, in den auch die im Rahmen der Erfassung von österreichischen Dissertationen entstehenden Autoren-Keywords (deutsch und englisch, jeweils deutlich unter 1%) mit eingehen. Schließlich sind noch die Klassenbenennungen aus Basisklassifikation (BK) und Mathematics Subject Classification (MSC) zu erwähnen, mit denen die Titeldaten im Zuge der Normdatenverknüpfung – BK und MSC werden als

11 Vgl.: Oberhauser, O. (2005). Klassifikatorische Erschließung und Recherche im Österreichischen Verbundkatalog. B.I.T. online. 8(2). 123–132.

Normdatei geführt[12] – angereichert werden. Im Bereich der **klassifikatorischen** Sacherschließung gibt es keine verbindliche Verbundklassifikation, sondern fünf Systeme, die in den Verbunddatensätzen zugelassen sind. Von diesen spielt die Regensburger Verbundklassifikation (RVK), die auch in Österreich immer mehr aktive Anwender findet, mit ca. 7,5% (575.000 Titel) die bedeutendste Rolle. Notationen der Dewey Decimal Classification (DDC), die mehrheitlich durch die Nutzung der DNB-Fremddaten in den Verbund gelangen (aktive Anwenderbibliotheken der DDC sind nur die Oberösterreichische Landesbibliothek und Teile der UB Wien), finden sich bei 3,7% der Datensätze. Die BK, die über ein beträchtliches Potential als Grundlage einer sachlichen Facette in einem neuen OPAC verfügen würde,[13] wird erst seit kurzem von zwei großen Bibliotheken aktiv eingesetzt (Österreichische Nationalbibliothek; UB Wien). Derzeit tragen nur ca. 168.000 Verbundtitel (2,2%) BK-Notationen; eine Datenübernahme (aus dem GBV), mit der es möglich wäre, einen weit höheren Anteil von Titeln mit BK-Notationen anzureichern, ist bisher leider nicht zustande gekommen. Zu erwähnen ist schließlich auch, dass nahezu alle Zeitschriftentitel des Katalogs (ca. 271.000) mit Notationen der ZDB-Systematik versehen sind. Die nur in Fachbereichsbibliotheken eingesetzte MSC ist im Verbundkatalog bei rund 26.000 Titeln vertreten.

In dem als Ergänzung des Verbundkatalogs geführten System **eDOC** werden gegenwärtig rund 420.000 elektronische Objekte mit dem Verbundkatalog verlinkt bzw. indexiert (soweit textuelle Elemente vorhanden sind). Etwa 3,6% der Verbundtitel (ca. 273.000 Datensätze) weisen solche Links auf. Bei 73% der elektronischen Dokumente handelt es sich um Inhaltsverzeichnisse und bei 14,5% um Abstracts. Der Rest verteilt sich auf Volltexte, Klappentexte, Umschlagbilder, Rezensionen und Titelseiten.

4 Konzeptionelle Überlegungen

Im folgenden werden die Überlegungen zur Rolle der einzelnen sacherschließenden Datenelemente in der neuen Suchmaschine dargelegt, d.h. die funktionellen Soll-Vorstellungen, wie sie *vor* der tatsächlichen Implementierung in Primo entstanden sind (nicht immer ist es allerdings dabei möglich, letztere außer Acht zu lassen). Angesichts der quantitativen Gegebenheiten geht es hier hauptsächlich um die RSWK-Schlagwörter; danach wird auch kurz auf andere verbale Sacherschließungselemente sowie die Verbundklassifikationen eingegangen.

4.1 RSWK

Überlegungen zur sachlichen Facette

Wie oben erwähnt, sind im Verbundkatalog die Schlagwortketten nach RSWK das am weitesten verbreitete Mittel der Sacherschließung, auch wenn eine Quote von knapp 40% bei weitem nicht flächendeckend ist. Trotz der von ihm thematisierten Unzulänglichkeiten von RSWK bzw. SWD spricht sich Imhof[14] *für* die Verwendung der RSWK-Ketten – insbesondere des jeweiligen Hauptschlagwortes – im Rahmen einer facettierten Navigation aus, wobei er für die differenzierte Behandlung der unterschiedlichen Schlagworttypen in verschiedenen Facetten plädiert. Das Problem der auch in Deutschland alles andere als flächendeckenden Erschließung nach RSWK/SWD wird im Rahmen seiner Argumentation freilich außer Acht gelassen, wenngleich Imhof darauf verweist, dass eine „gute" Facette idealerweise auf ein Kriterium zurückgreifen können sollte, welches bei 100% der Datensätze anzutreffen ist. Nicht eingegangen wird auch auf die Gefahr, die u.E. bei der Verwendung einer sehr großen Zahl von (spezifischen) Schlagwörtern im Rahmen einer sachlichen Facette besteht: Würde dies nicht dazu führen, dass die Ausprägungen dieser Facette das jeweilige Suchergebnis – anstatt es in handliche Cluster zu untergliedern – in eine Vielzahl von lediglich ganz schwach besetzten Grüppchen aufsplittern? In Katalogen mit facettierter Navigation kann man bereits jetzt Ähnliches bei der Facette „Autor" bzw. „Körperschaft" beobachten. Dennoch erschien vorerst die Verwendung der Schlagwörter für die sachliche Facette die einzige sinnvolle Alternative zu sein, da derzeit viel zu wenige Verbunddatensätze mit klassifikatorischen Sacherschließungselementen, die sich hier vermutlich besser eignen würden, ausgestattet sind.

SW 1.SW-Kette	902		g	Wien
			a	Stephansdom
			9	4079282-1
SW 1.SW-Kette	902		s	Architektur
			9	4002851-3
SW 1.SW-Kette	902		f	Führer
Erl. 1.SW-Kette	904	a	a	UBWFRA
SW 2.SW-Kette	907		g	Wien
			a	Stephansdom
			9	4079282-1
SW 2.SW-Kette	907		z	Geschichte
Erl. 2.SW-Kette	909	a	a	UBWFRA
SW 3.SW-Kette	912		g	Wien
			a	Stephansdom
			9	4079282-1
SW 3.SW-Kette	912		z	Geschichte 1277-1466
SW 3.SW-Kette	912		f	Quelle
Erl. 3.SW-Kette	914	a	a	UBWFRA

Abbildung 4

Extraktion von RSWK-Datenelementen

Welche Elemente der RSWK-Ketten sollten nun für eine sachliche Facette extrahiert werden, welche sind für die künftige Sachrecherche und welche für das Display (Vollanzeige bzw. Weiternavigieren) relevant? (Als Sachrecherche bezeichnen wir hier die im Primo-Suchmodus *erweiterte Suche* mögliche Recherche im Feld „Schlagwort").

- Für die sachliche Facette wären zweifellos Einzel-Schlagwörter und nicht komplette Ketten zu bevorzugen, da letztere zu einer noch stärkeren Aufsplitterung von Ausprägungen mit geringer Fallzahl führen würde. Gemeint ist allerdings jeweils der Inhalt einer Kategorie 9xx (902, 907, …) und nicht etwa Stichwörter daraus. Beispiele wären etwa: Hundefutter; Wien / Steinhofkirche; Physikotheologischer Gottesbeweis.
- Die vollständigen RSWK-Ketten werden hingegen für das Display sehr wohl benötigt.
- Form- und Zeitschlagwörter sollen in der Facette unterdrückt werden, für Suche und Vollanzeige hingegen zur Verfügung stehen. Ursprünglich war daran gedacht worden, sie in der Vollanzeige zwar darzustellen, aber auf ihrer Basis kein Weiternavigieren zuzulassen.
- Geographika ohne Unterfeld $9, d.h. ohne Normdatenverlinkung, wären in der Facette ebenfalls zu unterdrücken (dies betrifft in erster Linie Kongress-/Ausstellungsorte), sollen aber selbstverständlich such- und anzeigbar sein. Auch hier galt ursprünglich die oben erwähnte Restriktion hinsichtlich des Weiternavigierens.
- Die ID-Nummer des jeweiligen Normdatensatzes (Unterfeld $9) wird zwar nicht für die sachliche Facette gebraucht, wäre aber, schon im Hinblick auf die Suchmöglichkeiten, ebenfalls mitzuführen.
- Die Formschlagwörter könnten eventuell mit bereits bestehenden formalen Kriterien bezüglich Dokumenttyp bzw. Erscheinungsform kombiniert werden.

Beispiele hiezu zeigt *Abbildung 4*. Aus der ersten Kette (902) wäre zu extrahieren:

12 Vgl.: Oberhauser, O.; Labner, J. (2006). Entwurf und Implementierung einer neuen Multi-Klassifikations-Normdatei für den Österreichischen Verbundkatalog. B.I.T. online. 9 (4). 293–302.

13 Vgl.: Oberhauser, O.; (2008). Sachliche Erschließung im Österreichischen Verbundkatalog: Status und Perspektiven. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare.* 61 (3). 59–77.

14 Imhof, op. cit.

- für die sachliche Facette:
  1. Wien / Stephansdom
  2. Architektur
     nicht aber: Führer (da Formschlagwort [Unterfeld $f])
- für die Suche:
  1. Wien / Stephansdom / Architektur / Führer [Kette]
  2. Wien / Stephansdom [Phrase]
  2. Wien
  2. Stephansdom
  2. 4079282-1
  3. Architektur
  3. 4002851-3
  4. Führer
- für das Display:
  1. Wien / Stephansdom; Architektur; Führer [Kette mit Unterscheidung zwischen einfachen und komplexen Schlagwörtern][15]
- für das Weiternavigieren aus der Vollanzeige:
  1. Wien / Stephansdom
  2. Architektur
  3. Führer [ursprünglich nicht intendiert]

Aus der zweiten Kette (902) wäre zu extrahieren:

- für die sachliche Facette:
  1. Wien / Stephansdom
     nicht aber: Geschichte (da Zeitschlagwort [Unterfeld $z])
- für die Suche:
  1. Wien / Stephansdom / Geschichte [Kette]
  2. Wien / Stephansdom [Phrase]
  2. Wien
  2. Stephansdom
  2. 4079282-1
  3. Geschichte
- für das Display:

Wien / Stephansdom; Geschichte [Kette mit Unterscheidung zwischen einfachen und komplexen Schlagwörtern]

- für das Weiternavigieren aus der Vollanzeige:
  1. Wien / Stephansdom
  2. Geschichte [ursprünglich nicht intendiert]

Aus der dritten Kette (902) wäre zu extrahieren:

- für die sachliche Facette:
  1. Wien / Stephansdom
     nicht aber: Geschichte 1277-1466 (da Zeitschlagwort [Unterfeld $z])
     nicht aber: Quelle (da Formschlagwort [Unterfeld $f])
- für die Suche:
  1. Wien / Stephansdom / Geschichte 1277-1466 / Quelle [Kette]
  2. Wien / Stephansdom [Phrase]
  2. Wien
  2. Stephansdom
  2. 4079282-1
  3. Geschichte 1277-1466
  4. Quelle
- für das Display:
  1. Wien / Stephansdom; Geschichte 1277-1466; Quelle [Kette mit Unterscheidung zwischen einfachen und komplexen Schlagwörtern]
- für das Weiternavigieren aus der Vollanzeige:
  1. Wien / Stephansdom
  2. Geschichte 1277-1466 [ursprünglich nicht intendiert]
  3. Quelle [ursprünglich nicht intendiert]

Ein zweites Beispiel zeigt Abbildung 5. Hier soll extrahiert werden:

- für die sachliche Facette:
  1. Hund
  2. Tierernährung
     nicht aber: Kongress (da Formschlagwort [Unterfeld $f])
     nicht aber: Cambridge <1985> (da Geographikum ohne Unterfeld $9)
- für die Suche:
  1. Hund / Tierernährung / Kongress / Cambridge <1985> [Kette]
  2. Hund
  2. 4026181-5
  3. Tierernährung
  3. 4060095-6
  4. Kongress
  5. Cambridge <1985>
- für das Display:
  1. Hund; Tierernährung; Kongress; Cambridge <1985>
- für das Weiternavigieren aus der Vollanzeige:
  1. Hund
  2. Tierernährung
  3. Kongress [ursprünglich nicht intendiert]
  4. Cambridge <1985> [ursprünglich nicht intendiert]

SW 1.SW-Kette	902		s	Hund
			9	4026181-5
SW 1.SW-Kette	902		s	Tierernährung
			9	4060095-6
SW 1.SW-Kette	902		f	Kongress
SW 1.SW-Kette	902		g	Cambridge <1985>

Abbildung 5 und Abbildung 6

Hauptschlagwort	800		s	Landschaftsgebundene
Quelle	808	a	a	SWL
s.Verw./Bemerk	830		s	Regionale Bauform
s.Verw./Bemerk	830		s	Landschaftsgerechtes
ÜbergeordSW	850		s	Architektur
ÜbergeordSW	850		s	Bauform
VerwandtesSW	860		s	Bauernhaus
VerwandtesSW	860		s	Volksarchitektur

Extraktion von Normdatenelementen

Wie bereits erwähnt, verfügte bereits der bisherige Verbundkatalog über die Möglichkeit, die katalogisierten Ansetzungsformen mit definierten Datenelementen aus den verwendeten Normdateien (PND, GKD, SWD) anzureichern. Dabei handelte es sich primär um eine Funktionalität zur Erweiterung der Indizes (Wortindex, Phrasenindex) – daher beschränkte sich diese Anreicherung auf die Siehe-Verweise. Im Falle der SWD kamen dafür also die Kategorien 830 (Äquivalente Bezeichnung = Synonym) und 820 (Alternativform zum Hauptschlagwort) in Frage.

Künftig können auch die übrigen Relationierungen – Mehrgliedriger Oberbegriff (845), Übergeordnetes Schlagwort (850) und Verwandtes Schlagwort (860) – miteinbezogen werden. Dies kann etwa am Beispiel des Schlagworts „Landschaftsgebundenes Bauen" veranschaulicht werden (*Abbildung 6*).

Für die Vollanzeige bzw. das Weiternavigieren aus dem Display spielt diese Anreicherung keine Rolle, sehr wohl dagegen ist sie für die Suche sowie auch die sachliche Facette von Relevanz. Dafür würde im obigen Beispiel die Extraktion folgende Elemente liefern:

1. Regionale Bauform
2. Landschaftsgerechtes Bauen
3. Architektur
4. Bauform
5. Bauernhaus
6. Volksarchitektur

Nicht vergessen werden soll dabei, dass bestimmte Unterfelder wie z.B. Unterfeld $v (Bemerkungen), deren Inhalte nicht zur Vokabularerweiterung beitragen, nicht extrahiert werden dürfen.

Problemkreis Homonymenzusatz

Ein Aspekt, der sowohl die Datenelemente aus den RSWK-Ketten als auch jene aus den Normdaten tangiert, ist der Problemkreis „Homonymenzusatz" (HZ), der im bisherigen Verbundkatalog nicht zufriedenstellend gelöst werden konnte. Ein künftiges – vielleicht nicht in jedem Detail realistisches –Szenario könnte wie folgt aussehen:

- in der sachlichen Facette soll der HZ angezeigt wer-

15 Die Verwendung des Semikolons als Trennzeichen zwischen den Schlagwörtern einer Kette ist in Primo aus programmtechnischen Gründen notwendig (Ansetzungsketten behalten dagegen den Schrägstrich).

den, da damit die Unterscheidbarkeit gewährleistet wird
- ebenso soll der HZ in der Vollanzeige sichtbar sein
  daraus folgt, dass der HZ auch beim Weiternavigieren mitgeführt wird
- für die einfache Suche soll dies ebenso gelten
- bei der **erweiterten Suche im Schlagwortindex** wäre folgendes Systemverhalten wünschenswert:
  – Stichwörter aus dem HZ sollten weder für sich allein noch in Kombination mit beliebigen anderen Suchbegriffen zu Treffern führen
  – hingegen sollten sie in Kombination mit dem Schlagwort, aus dessen HZ sie ursprünglich stammen, sehr wohl zu Treffern führen
  – nicht betroffen davon sind die gleichlautende Suchbegriffe, wenn sie nicht aus dem HZ stammen (also z.B. als Sachschlagwort verwendet werden)

Ein Beispiel wäre etwa das Schlagwort „Holbein, Hans <Künstler, 1497-1543>„: Eine Suche nach „Künstler" sollte Treffer mit diesem Schlagwort nicht erbringen, eine Suche nach „Holbein Künstler" hingegen schon.
Auch die nachfolgenden, willkürlich ausgewählten Beispiele veranschaulichen diese Problematik:

800 Kirchberg <Regen>
830 Kirchberg <Wald, Bayern>
800 Bank Austria <Schweiz>
800 Oldenburg <Oldenburg, Bezirk>
830 Oldenburg <Land>
830 Oldenburg <Staat>
800 Holbein, Hans <Künstler, 1497-1543>
830 Holbein, Hans <der Jüngere>
800 Ada <Holland, Gräfin>
800 Hunde <Familie>
800 Ölbaum <Motiv>

Kommunikations- und Parametrisierungsgrundlage

Viele der bisherigen Überlegungen sind in der tabellarischen Übersicht in *Abbildung 7*, die auch als Kommunikations- und Parametrisierungsgrundlage für das Primo-Setup gedient hat, zusammengefasst dargestellt. Die Spalten 2–7 beziehen sich auf die unterschiedlichen Sektionen im Primo-Internformat (PNX) bzw. die damit verbundenen Funktionen (Spalte 2 = Vollanzeige; Sp. 3 = Weiternavigieren aus der Vollanzeige; Sp. 4 = feldübergreifende Suche / „Alle-Felder-Index"; Sp. 5 = erweiterte Suche / Sachrecherche; Sp. 6 = Facette „Thema" (topic); Sp. 7 = Kandidat für weitere Facetten wie Medium oder Form).

1	2	3	4	5	6	7
Extraktion aus	**Display** display	**Display** navi	**Index** search	**Index** search/subject	**Facette** Topic	**Facette** weitere
Titeldaten						
902ff. $spckgta9	Y	Y	Y	Y HZ=N	Y	N
902ff. $f	Y	N	Y	N	N	Y
902ff. $z	Y	N	Y	N	N	Y
902ff. $g (ohne $9)	Y	N	Y	N	N	N
Normdaten						
820ff. $spckgta -$v	N	–	Y	Y HZ=N	N	N
830 $spckgt -$v	N	–	Y	Y HZ=N	N	N
845 $spckgt -$v	N	–	Y	Y HZ=N	N	N
850 $spckgt -$v	N	–	Y	Y HZ=N	N	N
860 $spckgt -$v	N	–	Y	Y HZ=N	N	N

Abbildung 7

In *Abbildung 7* symbolisiert die Farbe Grün die spätere wunschgemäße Implementierung der betreffenden Funktionalität, während die Farbe Rot ausdrückt, dass der ursprüngliche Wunsch (vorerst) nicht realisiert werden konnte. Das bedeutet z.B., dass in Sp. 3 der ursprüngliche Wunsch „N" bei 902ff. $f (Form-SW *nicht* für das Weiternavigieren anbieten) *nicht* umgesetzt wurde.
Ausgangspunkt sind die Kategorien 902ff.[16] der Titeldaten (Spalte 1 Bereich „Titeldaten"). Wenn bei den Unterfeldern $spckgta (Sach-SW, Personen-SW, Körperschafts-SW, Geo-SW, Werktitel-SW, Unterschlagwort) über die SWD-Identnummer (IDNR) in $9 ein Normdatensatz gefunden wird, so werden zusätzlich bestimmte Normdaten-Kategorien (Spalte 1 Bereich „Normdaten") mit eingebunden. Im anderen Fall (d.h. wenn kein Unterfeld $9 vorhanden ist bzw. der Inhalt von $9 zu keinem Normdatensatz führt), so werden nur die Datenelemente aus den Titelkategorien herangezogen.

16 d.h. 902, 907, 912, ..., 947.

Die Inhalte der Unterfelder $fz (Form-SW und Zeit-SW) und $g ohne $9 (Geo-SW ohne IDNR, z.B. Kongressorte) sollten zumindest für den Schlagwortindex und die Facette „Thema" unberücksichtigt bleiben. Dies konnte bei der Implementierung nur für die Facette (grünes N in Sp. 6), nicht aber für den Schlagwortindex realisiert werden (rotes N in Sp. 5). Da die genannten Inhalte aber jedenfalls im „Alle-Felder-Index" vorkommen sollen, konnte die ursprünglich gewünschte Deaktivierung für das Weiternavigieren aus der Vollanzeige nicht erreicht werden (daher rotes N in Sp. 3). Das oben angeführte wünschenswerte Szenario im Zusammenhang mit dem Homonymenzusatz konnte erwartungsgemäß ebenfalls nicht umgesetzt werden (rotes HZ=N in Sp. 5).
Die Möglichkeit der Einbeziehung der synonymen Begriffe bzw. alternativen Ansetzungsformen in der Facette „Thema" wurde zwar erwogen, schließlich aber doch verworfen (vgl. grünes N in Spalte 6), schon um die Facette nicht unnötig aufzublähen. In Kauf genommen werden mußte dabei allerdings, dass in Österreich bevorzugte Synonyme (z.B. „Traktor" anstelle von „Schlepper") oder auch ins Deutsche übersetzte Werktitel „außen vor" bleiben müssen. Ob künftig eventuell auch Oberbegriffe (845, 850) in die Facette „Thema" einfließen sollten, wird erst in Zukunft zu diskutieren sein.

4.2 Andere verbale Sacherschließungselemente

Auch die Einbeziehung der weiteren oben dargestellten verbalen Sacherschließungselemente in das künftige sachliche Such-Interface wurde erwogen. Schließlich kamen wir zu dem Schluss, dass die Recherche nach diesen Datenelementen über den Index „Alle Felder" realisiert werden soll, nicht aber über den in der erweiterten Suche angebotenen Schlagwortindex (keine Vermischung mit RSWK). Ebenso sollen etwa LCSH *nicht* in der Facette „Thema" aufscheinen. Sehr wohl hingegen ist es sinnvoll, diese Sacherschließungselemen-

te in der Vollanzeige darzustellen und auch mit ihnen ein Weiternavigieren zu ermöglichen. Ob die verbalen Elemente aus der BK (Klassenbenennungen, Synonyme) überhaupt für die Suche herangezogen werden sollen oder sogar *nicht* in den Index „Alle Felder" mit eingehen sollen, wurde ebenfalls diskutiert. Da aber ein Prinzip des Verbundkatalogs darin besteht, immer alle verbalen Datenelemente in diesen Index aufzunehmen, wurde entschieden, den Status Quo auch in der neuen Umgebung beizubehalten. Auch die verbalen Elemente aus der MSC sollen in diesen Index aufgenommen werden.

Primo hat, ganz dem Anforderungsprofil entsprechend, eine weitere, sehr wichtige Verbesserung mit sich gebracht, nämlich die Möglichkeit der völligen Integration der zu den Katalogisaten angelegten Abstracts, Inhaltsverzeichnisse und Klappentexte. Alle drei genannten Dokumenttypen sollen im Volltext (d.h. für den Index „Alle Felder") indexiert werden, wie das auch in eDoc der Fall war. Man kann diese Stichwörter als unkontrolliertes Vokabular mit potentiell mehr oder weniger ausgeprägtem inhaltserschließenden Charakter betrachten.

4.3 Klassifikatorische Sacherschließungselemente

Die im bisherigen Verbundkatalog angebotene Funktion „Index blättern" steht in Primo nicht mehr zur Verfügung. Dies betrifft auch die fünf oben angeführten Verbundklassifikationen, für die ein Sucheinstieg über den Index sehr nützlich war. Bei der auf dem Stichwortprinzip basierenden Suche sollen einerseits die Notationen bzw. ihre Bestandteile (z.B. im Fall der RVK: PB 1601 bzw. PB und 1601) als Suchbegriffe eingegeben werden können, andererseits – bei den beiden als Normdatei verfügbaren Systemen (BK, MSC) – auch die Stichwörter aus den verbalen Elementen. Neben der Möglichkeit, diese Datenelemente im Index „Alle Felder" (einfache Suche) zu recherchieren, soll die die „erweiterte Suche" einen feldspezifischen Einstieg pro Klassifikationssystem anbieten.

Für die Darstellung in der Vollanzeige und das Weiternavigieren aus dieser wären die vollständigen Notationen (inklusive der Klassenbenennungen bei BK und MSC) vorzusehen. Mit der im bisherigen OPAC gebotenen Möglichkeit der „Interpretation" der Notationen der RVK durch die Verlinkung mit der Regensburger Datenbank[17] bzw. der DDC und der ZDB-Systematik mittels hinterlegter Listen kann (vorerst) wohl nicht gerechnet werden.

Die Frage, inwieweit eine der Verbundklassifikationen für eine Facette (oder auch mehrere für mehrere Facetten) genutzt werden sollte, ist aufgrund des derzeitigen Anteils der mit Notationen ausgestatteten Datensätze (wie oben erwähnt, hält diesbezüglich die RVK mit 7,5% die Spitze) eher negativ zu beantworten. Sollte es dagegen gelingen, durch verbundübergreifende Anreicherungsprojekte eine „flächendeckendere" Ausstattung mit Notationen zu erreichen,[18] wären insbesondere die BK, aber auch die RVK oder die DDC für eine Facette geeignet. Wie aus anderen Primo-Applikationen bekannt ist,[19] bietet Primo für die Library of Congress Classification (LCC) sogar eine hierarchische Facetten-Option an, bei der zunächst nur die oberste Hierarchiestufe angezeigt wird. Bei Auswahl einer Hauptklasse wird die Facette neu berechnet und die nächste darunter liegende Hierarchiestufe zur Auswahl angezeigt. So kann ein Ergebnisset immer enger eingeschränkt werden, ohne sofort auf die kleinste Hierarchiestufe wechseln zu müssen. Letztere wäre aufgrund meist geringer Fallzahlen – ähnlich wie im Fall ganzer RSWK-Ketten – ohnedies für eine Facette weniger geeignet.

5 Implementierung

Wie bereits in Abschnitt 2 kurz angerissen, wird jeder Datensatz in Primo als XML-Datei abgespeichert. Die einzelnen Sektionen dieser XML-Datei werden für jeweils unterschiedliche Funktionen von Primo ausgewertet. Über die Normalisierung ist jede MAB-Kategorie, jeder Indikator, jedes Unterfeld getrennt ansprechbar und kann einer oder mehreren PNX-Abschnitten zugeordnet werden. Einzelne MAB-Felder und deren Unterfelder können zusammengeführt oder zerschlagen werden, zahlreiche String-basierte Funktionen können das Katalogisat anreichern oder beschneiden. Die Trennung von Erfassungs- und Präsentationssystem (Aleph und Primo), die übersichtliche Aufteilung der XML-Abschnitte und die einfache Handhabung der Transformation ermöglichen somit eine gezielte Manipulation der Quelldateien, einen maßgeschneiderten Zuschnitt für die spätere Verwendung und in der Folge eine nachträgliche Bereinigung der Datenbank. Es ist evident, dass ein sauberes Quellsystem erste Wahl bleibt – Primo sollte als Datenempfänger, d.h. abhängiges System, nicht zur Kaschierung fehlerhafter Quellen missbraucht werden. Dennoch: Eine Bereinigung von Altlasten oder fehlerhaften Katalogisaten scheitert oft an Ressourcenknappheit; für solche Fälle ergibt sich mit Hilfe der Normalisierung die Möglichkeit, mehr oder weniger umfassende Katalogbereinigungen durchzuführen.

Noch vor der Normalisierung muss ein Weg gefunden werden, die Daten in einem für Primo lesbaren Format aufzubereiten. In Aleph steht für diese Aufgabe ein eigenes Tool, der *Aleph Publishing Mechanism* (APM), zur Verfügung. Dieser erlaubt es, an bestimmten „Stellschrauben zu drehen" und damit den Aufwand auf der Seite von Primo bereits um einiges zu reduzieren. Die bestehende bibliographische Datenbank wird quasi im XML-Format dupliziert. Jeder einzelne Datensatz kann zusätzlich um Angaben aus übergeordneten Werken, Bestandsangaben für Exemplare und Zeitschriften sowie Normdaten-Verlinkungen angereichert werden, was gleichsam zu einer Aufhebung der relationalen Struktur und zu einer Vervielfachung der katalogisierten Felder führt. Ein abhängiger Band z.B. ist dann nicht mehr über einen Link mit der Gesamttitelaufnahme verbunden, sondern enthält alle Angaben des übergeordneten Werkes. Über den APM wurde auch die oben angesprochene Anreicherung der Sätze mit über- und gleichgeordneten Schlagwörtern realisiert.

Im Rahmen dieses Beitrags sind insbesondere drei PNX-Sektionen von Interesse: *display, search* und *facet*. Entsprechend der Tabelle in *Abbildung 7* wurden die in den Feldern 902ff. abgelegten RSWK-Ketten für die Facette „Thema" zerschlagen und jene Felder, bei denen es sich um Geographika oder Formschlagwörter handelt, entfernt. Dem besonderen Charakter einer RSWK-Kette, der syntaktischen Indexierung, wird durch die Zerschlagung übrigens einmal mehr *nicht* Rechnung getragen. Von der angesprochenen Möglichkeit der Katalogbereinigung wird auch in der thematischen Facette Gebrauch gemacht; so werden z.B. die unterschiedlichen Schreibweisen von Formschlagwörtern („Kongress" und „Kongreß") zusammengeführt.

Was die Titelvollanzeige betrifft, so wurde für diese von einer Zerschlagung der Ketten abgesehen. Deren Anzeige erfolgt wie gewohnt: eine Kette pro Zeile. In diesem Zusammenhang sei auf eine Standardfunktion von Primo hingewiesen, die aus bestimmten Feldern der Vollanzeige automatisch einen Hyperlink generiert, der dann für eine weiterführende Suche angeklickt werden kann (an jenen Stellen, an denen bereits von

17 Vgl.: Oberhauser, O. (2009). Das RVK-Interface der FU Berlin – modifiziert, zentral gehostet und lokal im Online-Katalog der WU Wien genutzt. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare.* 62(3). 43–56.

18 In diesem Fall wäre auch der Bedarf für ein hierarchisches Such-Interface à la *MelvilleSearch* für die DDC (http://melvil.d-nb.de/) oder *RVK-Online* (http://rvk.uni-regensburg.de/) gegeben.

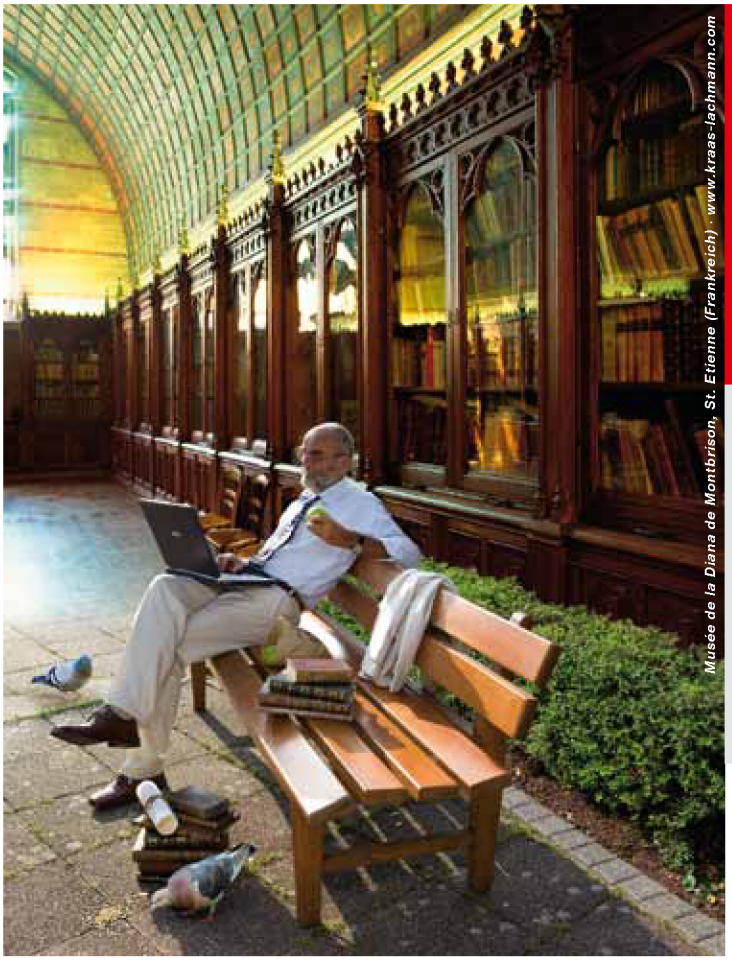
19 So z.B. in den Katalogen der Universitäten Vanderbilt (http://discoverlibrary.vanderbilt.edu/), Minnesota (http://www.lib.umn.edu/catalogs) oder Guelph (http://www.lib.uoguelph.ca/).

„Weiternavigieren" die Rede war, war genau diese Funktion gemeint). Dies gilt zunächst für das Feld „Autor/ in", aber auch für die einzelnen Elemente einer Schlagwortkette: Jedes Element einer Kette kann für eine weiterführende Suche ausgewählt werden. Wäre die gesamte Kette ein einziger Link und würde ein Klick darauf eine Suche nach Werken mit exakt dieser RSWK-Kette durchführen, dann wäre wegen der eher seltenen mehrfachen Verwendung ein und derselben Kette der Sucherfolg oft ziemlich bescheiden (nämlich ein einziger Treffer bzw. nur sehr wenige Treffer). Mit der Zersplitterung der Ketten in ihre Bestandteile für die Bildung der Ausprägungen in den Facetten bzw. die weiterführende Suche in der Vollanzeige geht einher, dass auch die Indexeinträge nicht aus ganzen Ketten, sondern aus einzelnen Schlagwörtern bestehen; andernfalls würde die weiterführende Suche ins Leere gehen. Unsere Parametrisierung sieht vor, dass im Anschluss an eine *einfache Suche* das Weiternavigieren mit Schlagwörtern im Index „Alle Felder" erfolgt, im Anschluss an eine *erweiterte Suche* hingegen im Schlagwortindex. Wenn im Index jedoch z.B. der *angezeigte* Homonymenzusatz nicht vorhanden wäre, würde eine weiterführende Suche ins Leere gehen, da etwa der Suchbegriff „Holbein, Hans <Künstler, 1497-1543>„ den Indexeintrag „Hohlbein, Hans" nicht finden würde. Somit ist der ursprünglich gewünschte Ausschluss der Homonymenzusätze aus dem Schlagwortindex nicht realisierbar. Ein entsprechender Wunsch bezüglich einer Einflussnahme auf die Linkbildung wurde bei der Herstellerfirma bereits deponiert.

Bei einer Suchmaschine, die in der Lage ist, eine Gewichtung der gefundenen Treffer vorzunehmen, stellt sich unweigerlich die Frage, wie thematische Aspekte zu gewichten sind. In diesem Zusammenhang tritt der Vorteil kontrollierter Metadaten gegenüber Volltextdokumenten deutlich zu Tage: Da der Suchmaschine eindeutig mitgeteilt werden kann, in welchem Feld welche Information zu finden ist, ist es ohne weiteres möglich, bestimmte Felder bei der Berechnung der Relevanz höher zu gewichten als andere.[20] Natürlich findet eine unterschiedliche Gewichtung verschiedener Felder auch bei Internet-Suchmaschinen statt. Was diese aber teilweise erst mühsam ermitteln müssen, liegt bei Katalogdaten unzweifelhaft vor. Auch handelt es sich um einwandfreie und nicht – wie bei manchen Web-Dokumenten – mit manipulativen Absichten (um ein besseres Ranking zu erzielen) vergebene Indexbegriffe. Im Gegensatz zu Web-Suchmaschinen dürfte die thematische Suche, die Suche, die ein Informationsbedürfnis zu einem bestimmten Wissensgebiet befriedigen soll, in Bibliothekskatalogen neben der Suche von Werken von bestimmten Autoren bzw. nach mehr oder weniger bekannten Werktiteln die häufigste Form der Suche darstellen. Ein kurzer Blick auf die bis dato in Primo eingegebenen Suchstrings bestätigt diese Vermutung. Benutzer suchen zumeist nach Autoren („Simone de Beauvoir"), Werken („Erich Fromm Kunst des Liebens") oder, am häufigsten, thematisch („Essstörung Kinder"). Demgegenüber zeigt die Auswertung von Suchen in Internet-Suchmaschinen ein viel breiter gefächertes Spektrum an Intentionen, die der Suche zu Grunde liegen.[21] Folgerichtig erhalten

20 Vgl. z.B.: Dellit, A.; Boston, T. (2007). Relevance ranking of results from MARC-based catalogues: From guidelines to implementation exploiting structured metadata. Paper presented at *Information online* 2007, Sydney, NSW; 30th January – 1st February 2007.
Online: http://conferences.alia.org.au/online2007/Presentations/30Jan.B1.relevance.ranking.pdf;
Langenstein, A.; Maylein, L. (2009). Relevanz-Ranking im OPAC der Universitätsbibliothek Heidelberg. *B.I.T. online.* 12 (4). 408–413; Lewandowski, D. (2009). Ranking library materials. *Library hi tech.* 27 (4). 584–593.

21 Rose, D. E.; Levinson, D. (2004): Understanding user goals in web search. *Proceedings of the 13th International Conference on World Wide Web, WWW ,04.* New York: ACM. 13-19. DOI: http://doi.acm.org/10.1145/988672.988675



Als wär man da.

Ihre Nutzer wollen mal eben im Park auf die Inhalte Ihrer wertvollen Originalausgaben zugreifen? Kein Problem!
Wir beherrschen mit unseren Digital- und Analogsystemen alle Prozesse der Dokumenten-Erfassung, -Archivierung, -Verarbeitung und -Bereitstellung. Seit mehr als 40 Jahren.

Zeutschel, die Zukunft der Vergangenheit.

ZEUTSCHEL

Zeutschel GmbH · Heerweg 2 · 72070 Tübingen · Tel.: +49 7071 9706-0
Fax: +49 7071 9706-44 · info@zeutschel.de · www.zeutschel.de

alle RSKW-Schlagwörter einen Multiplikator, der sie vor gleichlautenden Begriffen in anderen Abschnitten des Datensatzes auszeichnet. Dies ist allerdings eine Voreinstellung des Herstellers, in die bisher von unserer Seite nicht eingegriffen wurde.

Was die weiteren verbalen Erschließungselemente betrifft, so wurde während der Phase der technischen Implementierung ganz besonderes Augenmerk auf die Indexierung von Abstracts, Inhaltsverzeichnissen sowie Klappentexten gelegt. Wie sich herausstellte, waren Planung und Umsetzung schwieriger und erforderten mehr Zeitaufwand als vorher angenommen. Dennoch konnten schließlich die textuellen Anteile aller 300.000 betroffenen, im System eDoc abgelegten Objekte extrahiert und in Primo geladen werden. Dies wurde über ein sogenanntes Plug-in – Primo bietet die Möglichkeit, die fertigen PNX-Sätze noch weiter mit eigenen Java Programmen zu bearbeiten – realisiert. Anfängliche Befürchtungen, die Volltextindexate könnten eine „Verwässerung" der Metadaten mit sich bringen, haben sich bislang nicht bestätigt. Sollte jedoch Bedarf bestehen, so kann die Gewichtung von Volltextindexaten beeinflusst werden. Letztlich ist deren Menge jedoch noch zu gering, um einen ernsthaften Einfluss auf Suchergebnisse auszuüben.[22] Darüber hinaus steht es jeder teilnehmenden Einrichtung frei, ob in ihrer Sicht eine Volltextindexierung überhaupt zum Einsatz kommen soll oder nicht.

Die Normalisierung der Klassifikationskategorien erwies sich im Grunde als kaum problembehaftet, da der Kategorieninhalt – d.h. die Notation – meist vollständig kopiert wird. Leider stehen, wie oben erwähnt, Klassenbenennungen und andere verbale Elemente nur für die BK und die MSC (für die eine Normdatei vorgehalten wird), zur Verfügung. So wird etwa die in Aleph erfasste BK-Notation über den APM übernommen, aus der Normdatei mit Klassenbenennung und Synonymen ergänzt, wodurch Notation und verbale Bezeichnung angezeigt und facettiert werden können. Dass die BK wegen ihrer zu geringen Verbreitung derzeit noch nicht als Facette angezeigt werden kann, wurde bereits weiter oben vermerkt. Dagegen werden im Fall der RVK in Aleph nur die Notationen katalogisiert, was für eine Anzeige bzw. Facette wenig benutzerfreundlich ist. Ex Libris bietet im Auslieferungsumfang von Primo für dieses Problem ein Anreicherungsverfahren an, welches eine bestimmte RVK-Notation in eine verbale Bezeichnung übersetzt, die als die jeweilige Ausprägung der RVK-Facette angezeigt wird. Von diesem Verfahren wird aber bei uns kein Gebrauch gemacht, da diese Übersetzung auf dem Hinzufügen der Klassenbenennungen aus der gesamten Hierarchie beruht und so zu unübersichtlichen und viel zu langen Ausprägungsbezeichnungen führt (so etwa für die Notation ZB 86161: „Land- und Forstwirtschaft. Gartenbau. Fischereiwirtschaft. Hauswirtschaft – Wirtschafts- und Sozialwissenschaften – Agrarmarktpolitik – Interventionsinstrumente – Asien, Orient, Östliche Welt – Zentral und Ostasien, Hinterer Orient – China (einschl. Tibet, Mandschurei, Taiwan, Hongkong, Macao)"). Ideal wäre dagegen eine Form der hierarchischen Facettierung wie bereits oben für die LCC beschrieben. Als Kompromiss zwischen der allerengsten, über das Anreicherungsverfahren ermittelten Klassenbenennung und gar keiner Benennung findet daher die Facettenbildung derzeit auf zweiter Hierarchieebene statt. ZB 86161 fällt somit in die Ausprägung „ZB: Agrarwissenschaftliche Ausbildung, Agrarwissenschaftliche Forschung, Agrarverwaltung und -recht sowie Wirtschafts- und Sozialwissenschaften". Die Umsetzung Notation à Benennung findet ausschließlich in der Facette statt – eine verbale Suche nach Klassenbezeichnungen ist weder in der einfachen, noch in der erweiterten Suche möglich. Inwieweit hier noch eine Verbesserung (d.h. Suchbarkeit der verbalen Elemente) erzielt werden kann, müsste erst eruiert werden. Aus bereits erläuterten Gründen muss der Index wiederum der Anzeige entsprechen – andernfalls würde ein Weiternavigieren keine Resultate erbringen.

Was die Gewichtung der Notationsfelder betrifft, besteht auch hier die Möglichkeit, steuernd einzugreifen. Bisher hat sich dafür aber keine Notwendigkeit ergeben. Die ausschließliche Suche mit Notationen (d.h. alphanumerischen Codes) ist im Benutzerverhalten so gut wie überhaupt nicht registrierbar, was auch das Hervorheben oder Herabstufen der entsprechenden Felder unnötig erscheinen lässt. Sollte (dereinst) die erwähnte Suchbarkeit der Klassenbenennungen der RVK erreicht werden, könnte eine Neubewertung dieses Aspekts erforderlich sein.

Die grundsätzliche Problematik der Facettierung bei Beständen, in denen nicht jeder Datensatz jeder Facette zugeordnet werden kann, wurde bereits gestreift (vgl. Abschnitt 4.1, erster Absatz). Sie liegt darin, dass durch Auswahl einer Facetten-Ausprägung nicht nur die Datensätze ausgeschlossen werden, die *einer anderen* Ausprägung zugeordnet sind, sondern auch jene, die *gar keine* Ausprägung besitzen. Aus diesem Grund wird bei der Bildung der Facetten für jeden Satz zumindest ein Dummy-Wert vergeben. So enthält jede Facette auch eine Ausprägung „Ohne Angabe" oder „Nicht klassifiziert", welches meist jenes ist, das die höchste Trefferzahl aufweist. Ob dadurch tatsächlich das Bewusstsein dafür gestärkt werden kann, dass die Auswahl einer Facette immer mit dem Verlust einer großen Anzahl dort nicht vertretener Datensätze, die gleichwohl relevant gewesen sein könnten, mit sich bringt, sei allerdings dahingestellt.

6 Ausblick

Mit der Einführung von Primo hat eine neue Ära der wissenschaftlichen Recherche im Österreichischen Bibliothekenverbund begonnen. Suchmaschinentechnologie (Ergebnissets gewichtet nach Relevanz, facettierte Navigation) sowie die Vereinigung von Metadaten und Volltexten sind die Hauptcharakteristika der Innovation. Die daraus folgende Umstrukturierung und Neuordnung der sacherschließenden Elemente im Index und in der Suchoberfläche war nicht unbeträchtlich und erlaubt eine noch gezieltere Recherche nach inhaltlichen Aspekten.

Die Vorzüge der neuen Lösung können aber eigentlich erst voll zum Tragen kommen, wenn das vorhandene Datenmaterial bestimmten Qualitäts- und Quantitätsansprüchen genügt. War etwa die lediglich partielle Ausstattung der Datensätze mit RSWK-Ketten auch schon bisher ein Problem – die Schlagwortrecherche reduzierte, meist unbemerkt, die Suchbasis um mehr als die Hälfte –, so gilt dies umso mehr für die Facette „Thema" (bei der sich regelmäßig die Ausprägung „Ohne Angabe" als erste positioniert), zumal „gute" Facetten – wie oben bereits zitiert – im Idealfall auf Kriterien basieren sollten, die bei allen Datensätzen vorkommen. Das bedeutet, dass zukünftige Aktivitäten in vermehrtem Masse darauf abzielen müssen, die Verbunddatensätze durch Kataloganreicherungen (in verbundübergreifenden Kooperationsprojekten) mit RSWK-Ketten, aber auch mit Elementen der klassifikatorischen Sacherschließung möglichst flächendeckend auszustatten. Notationen der BK, RVK und DDC wären hier besonders hervorzuheben, aber auch verbale Elemente (Klassenbenennungen, Verweisungsformen usw.), die bei RVK und DDC zur Zeit noch nicht über Normdateien oder andere Quellen in zufriedenstellender Weise zur Verfügung stehen. Zu den Sacherschließungselementen zählen im weiteren Sinn auch die Stichwörter aus den elektronischen Dokumenten (Abstracts, Klappentexte, Rezensionen, Inhaltsverzeichnisse etc.) Die Forderung nach einer verstärkten Kataloganreicherung gilt daher auch für diese Objekte.

22 Das könnte sich eventuell ändern, wenn die Österreichische Nationalbibliothek den Primo-Betrieb aufnimmt. Vor kurzem wurde bekannt, dass Google einen guten Teil der wertvollen Bestände der Nationalbibliothek einscannen wird (http://www.onb.ac.at/bibliothek/austrianbooksonline.htm).

Aus dem oben Gesagten geht hervor, dass bisher nicht alle Intentionen hinsichtlich der inhaltlichen Recherche realisiert werden konnten. Als Defizit kann bspw. vermerkt werden, dass aus bereits genannten Gründen der Schlagwortindex derzeit auch die Form-, Zeit- und Geo-Schlagwörter (ohne IDNR) enthält, sodass eine Recherche nach einem davon betroffenen Begriff, der als Sach-SW gesucht wird, nahezu aussichtslos ist. Eine denkbare Lösungsvariante kann kurz wie folgt skizziert werden: Der bisherige Schlagwortindex, den auch die *display section* benötigt, um die vollständigen RSWK-Ketten in der Vollanzeige darzustellen, würde unverändert bleiben. Daneben könnte aber ein zweiter Sachindex definiert werden, der die unerwünschten Form-, Zeit- und Geo-Schlagwörter *nicht* enthält. In der „erweiterten Suche" würde dann dieser neue Index anstelle des vollständigen Schlagwortindex angeboten werden. Die oben ebenfalls diskutierten Probleme hinsichtlich der Homonymenzusätze und der nicht unterdrückbaren Weiternavigation mit Form-, Zeit und Geo-Schlagwörtern lassen sich durch eine solche Lösung allerdings nicht bereinigen.

Kurz erwähnt wurde bereits, dass für die klassifikationsbasierten Facetten (RVK, künftig ev. auch BK und DDC) ein hierarchischer Navigationsmechanismus nach dem Muster der LCC (in nordamerikanischen Primo-Implementierungen) wünschenswert wäre. In diesem Zusammenhang werden vermutlich neue Herausforderungen entstehen, da sich RVK und BK aufgrund ihrer nicht bzw. nur teilweise die Hierarchien abbildenden Notationen nicht ohne zusätzlichen Aufwand in ein hierarchisch arbeitendes Interface integrieren lassen.

Primo – ein relativ junges Produkt – wird vom Hersteller laufend weiterentwickelt. Zurzeit ist die neue Version 3 bei der OBVSG in einer Testphase; sie bringt als Neuerungen schwerpunktmäßig lokale OPAC-Funktionalitäten (Ausleihen, Vormerken etc.) Wir sind optimistisch, was auch die Weiterentwicklung auf dem Retrievalsektor betrifft und erwarten Verbesserungen hinsichtlich der inhaltsbezogenen Recherche in naher Zukunft.

Danksagung

Die Autoren danken Johann Brandauer und Martin Hekele (beide OBVSG) für die Durchsicht des Manuskripts.

AUTOREN

Mag. Markus Knitel

Systembibliothekar in der OBVSG, betreut die zentrale Primo-Instanz des Österreichischen Bibliothekenverbundes.
markus.knitel@obvsg.at

Ing. Josef Labner

Systemanalytiker im Österreichischen Bibliothekenverbund, in den Bereichen Normdateien, Datenformate und Systemparametrisierung.
josef.labner@obvsg.at

Dr. Otto Oberhauser MPhil MSc MLIS

Im Österreichischen Bibliothekenverbund vor allem in den Bereichen Online-Kataloge (zentrale und lokale Systeme) und Systemparametrisierung tätig.
otto.oberhauser@obvsg.at

Alle:
Die Österreichische Bibliothekenverbund und Service GmbH (OBVSG), Brünnlbadgasse 17/2a, A 1090 Wien, Österreich
http://www.obvsg.at

easy Check
library technologies

STAPELWEISE
MEHR
KOMFORT
MIT EASYCHECK!

RFID- und EM-Technik
für Bibliotheken in jeder Art und Größe

. Konvertierung
. Thekenverbuchung
. Selbstverbuchung
. Mediensicherung
. Rückgabe und Sortierung
. Softwarelösungen

Zahlreiche Eigenentwicklungen und hochwertige Einzelkomponenten starker Partner stehen für zeitgemäße und zukunftssichere Lösungen. Wir unterstützen Arbeitsabläufe in Bibliotheken – einfach, sicher, kundenfreundlich.

EasyCheck GmbH & Co. KG
Steinbeisstraße 12
73037 Göppingen
DEUTSCHLAND
Fon +49 (0)7161 808600-0
Fax +49 (0)7161 808600-22
mail@easycheck.org

www.easycheck.org

Die Grenzen der Modernität – das neue Regelwerk „Resource Description and Access“

Heidrun Wiesenmüller

Der Beitrag stellt den AACR2-Nachfolger „Resource Description and Access“ (RDA) und das Online-Werkzeug „RDA Toolkit“ vor. Er erläutert wichtige Prinzipien des neuen Regelwerks wie z.B. die Orientierung an FRBR und die Definition von Kernelementen. Es wird gezeigt, dass RDA aufgrund allzu großer Rücksichtnahme auf die bisherige amerikanische Katalogisierungspraxis seine Ziele nur zum Teil erreichen kann.

The limits of modernity – the new cataloguing code „Resource Description and Access“ (RDA)

„Resource Description and Access“ (RDA), which is supposed to replace AACR2, and the online tool „RDA Toolkit“ are presented. Important principles of the new cataloguing code, such as the alignment with FRBR and the definition of core elements, are discussed. It is shown that RDA does not achieve all of its objectives as it makes too much allowance for current American cataloguing practices.

Les limites de la modernité – le nouveau livre de normes « Ressource Description and Access »

On présente ici « Ressource Description and Access », qui doit remplacer AACR2, et l'outil en ligne « RDA Toolkit ». On explique d'importants principes du nouveau livre de normes, tels que l'alignement sur FRBR et la définition d'éléments centraux. On montre que RDA n'atteint pas tous ses objectifs parce qu'il prend trop en considération les pratiques américaines connues de catalogage.

■ Nach über sechsjähriger Entwicklungsarbeit erschien im Juni 2010 die erste Ausgabe des Nachfolgers der Anglo-American Cataloguing Rules (AACR2): „Resource Description and Access“ (RDA). Anfangs hatte man noch von „AACR3“ gesprochen. Der 2005 gewählte neue Titel (auf Deutsch: Beschreibung von Ressourcen und Zugang zu ihnen) signalisiert, dass mit RDA etwas wirklich Neues intendiert war, das weder auf den angloamerikanischen Raum noch auf den Bibliotheksbereich beschränkt bleiben sollte. Die beiden Aspekte „description“ (bibliographische Beschreibung) und „access“ (Zugang, d.h. Eintragungen und Ansetzungen) entsprechen freilich einem durchaus traditionellen Bild von Erschließung.

Das RDA Toolkit

Ursprünglich sollte es RDA ausschließlich in elektronischer Form geben; angesichts heftiger Proteste wurde mittlerweile aber auch eine Loseblattausgabe angekündigt. Die Online-Ausgabe, das so genannte „RDA Toolkit“[1] (Abb. 1), stand während einer Schnupperphase kostenlos zur Verfügung; regulär fallen Lizenzkosten an: Bei Zahlung des Grundpreises von 285 Euro pro Jahr können beliebig viele Nutzer das Toolkit verwenden, jedoch nicht zur selben Zeit. Sollen beispielsweise fünf Nutzer gleichzeitig zugreifen können, so liegen die jährlichen Gesamtkosten schon bei 485 Euro. Und was bekommt man für sein Geld? Zum einen eine wenig spektakuläre Aufbereitung des Regelwerkstexts mit einer etwas schwerfälligen Navigation, zum anderen weitere Ressourcen wie z.B. den kompletten Text von AACR2 sowie – wie die Bezeichnung „tool kit“ (Werkzeugkasten) bereits andeutet – verschiedene Werkzeuge. Das vielleicht interessanteste davon ermöglicht es, eigene Schritt-für-Schritt-Anweisungen (workflows) anzulegen und diese z.B. mit Regeln und Glossareinträgen aus RDA oder externen Quellen zu verlinken (Abb. 2). Mittelfristig wird das Toolkit auch Übersetzungen von RDA in andere Sprachen enthalten.

In den USA läuft seit 1. Juli 2010 ein groß angelegter Test unter Beteiligung von 26 Institutionen. Die Ergebnisse werden die Basis für die gemeinsame Entscheidung der drei US-amerikanischen Nationalbibliotheken (Library of Congress, National Library of Medicine, National Agricultural Library) sein, „whether or not to imple-

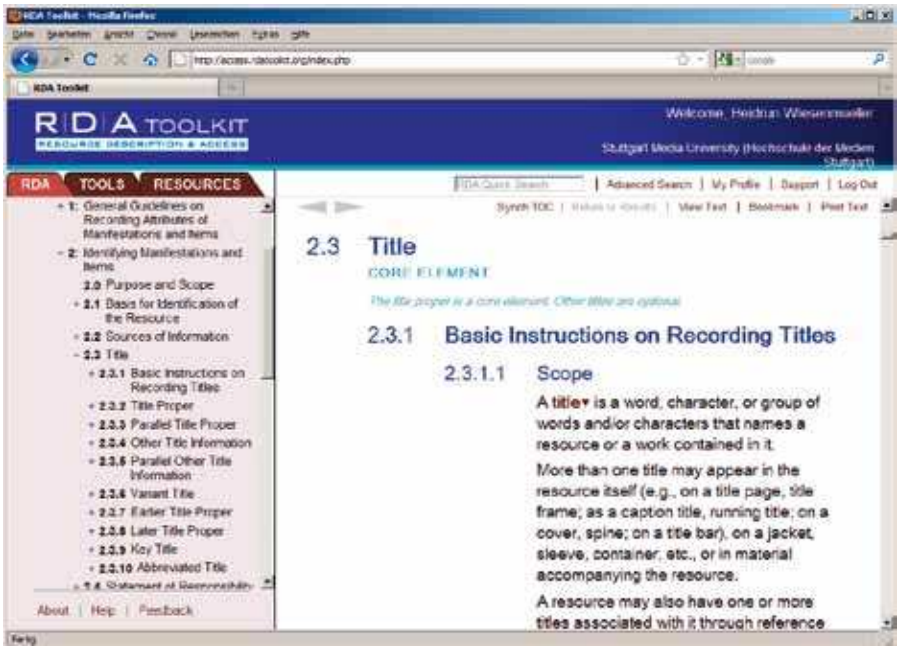


Abbildung 1: Darstellung des Regelwerkstexts im „RDA Toolkit“

1 http://www.rdatoolkit.org/ [09.08.2010].

ment RDA"[2]. Der Test, bei dem sowohl das Regelwerk selbst als auch das Toolkit in Hinblick auf Betriebsabläufe, technische Fragen und Wirtschaftlichkeit evaluiert werden sollen, teilt sich in eine dreimonatige Übungs- und eine ebenfalls dreimonatige Praxisphase. Er ist eine Reaktion auf die scharfe Kritik an RDA, die Anfang 2009 von einer hochkarätig besetzten Expertengruppe formuliert worden war.[3] Die endgültige Entscheidung der US-Nationalbibliotheken zur Einführung von RDA wird im zweiten Quartal 2011 erwartet. Die Testergebnisse aus den USA gelten – neben dem Vorliegen einer deutschen Übersetzung – auch als Voraussetzung für die Entscheidung in Deutschland.

Einstieg mit Hindernissen

Der Einstieg in das neue Regelwerk wird Neulingen nicht leicht fallen, denn RDA baut in mehrfacher Hinsicht Hürden auf. Die erste ist bereits der Stil: Anstatt auf die versprochenen Formulierungen „in clear, concise, simple terms"[4] trifft man vielfach auf merkwürdig „mäandernde" Satzkonstrukte voller Umständlichkeiten und Redundanzen.[5] Ob der Text in deutscher Übersetzung lesbarer sein wird, darf bezweifelt werden.

Die zweite Hürde ist die völlig neuartige Gliederung: Die ersten vier Abschnitte (nach der mit „0" gekennzeichneten Einleitung) stehen unter der Überschrift „Merkmale" (attributes), die restlichen sechs unter der Überschrift „Beziehungen" (relationships). Der Hintergrund für diese 2007 eingeführte Struktur war der Wunsch, RDA stärker mit dem theoretischen Modell „Functional Requirements for Bibliographic Records"[6] (FRBR) in Einklang zu bringen. Ohne solide Kenntnisse dieses Modells ist deshalb der Aufbau von RDA nicht verständlich.

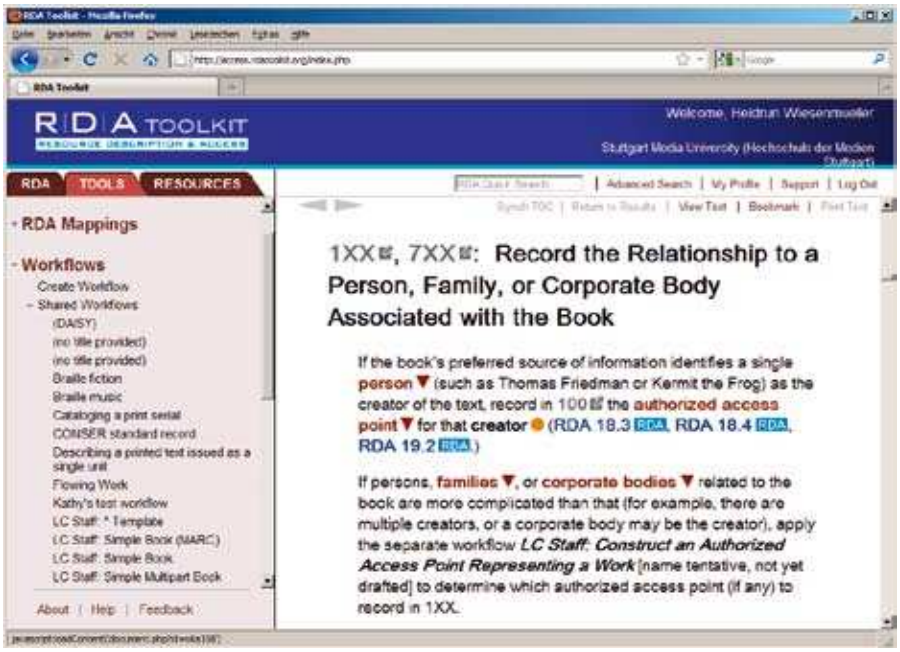


Abbildung 2: Ausschnitt aus einem von der Library of Congress angelegten Workflow

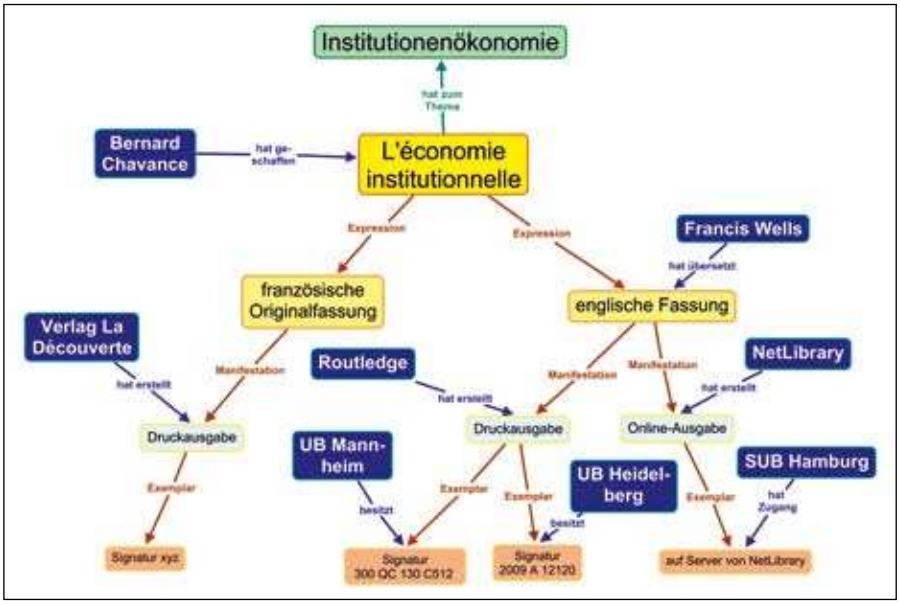


Abbildung 3: Darstellung der FRBR-Entitäten am Beispiel des Werks „L"économie institutionelle"

FRBR definiert drei Gruppen von Entitäten (Objekten) mit jeweils bestimmten Merkmalen sowie die zwischen diesen Entitäten bestehenden Beziehungen. Die Entitäten der Gruppe 1 sind Werk (work), Expression, (expression), Manifestation (manifestation) und Exemplar (item). Das in Abb. 3 dargestellte Werk „L"économie institutionelle" beispielsweise besitzt zwei Expressionen – die französische Originalfassung und die englische Übersetzung. Jede Expression wird in einer oder mehreren Manifestationen verkörpert: Die englische Fassung gibt es z.B. in einer Druckausgabe und als E-Book. Jede Manifestation schließlich liegt in mindestens einem konkreten Exemplar vor. In der Abbildung blau dargestellt sind die Entitäten der Gruppe 2: Diese repräsentieren Personen und Körperschaften, die für die Entitäten der Gruppe 1 verantwortlich sind, z.B. den geistigen Schöpfer (creator) des Werkes, den Übersetzer der englischsprachigen Expression, die Verlage als Produzenten der Manifestationen und die Bibliotheken als Besitzer der Exemplare. Die Gruppe 3 schließlich (hier grün dargestellt) umfasst alle Entitäten, die Thema eines Werkes sein können. Hier bewegt sich FRBR – und in seiner Folge auch RDA – in den Bereich der Sacherschließung hinein.

Stellt man RDA und FRBR gegenüber, so erschließt sich die innere Logik: In den ersten vier Abschnitten werden die unterschiedlichen FRBR-Entitäten behandelt, z.B. die Gruppe 2 im Abschnitt 3 – dort sind also u.a. die Ansetzungsregeln für Personen und Körperschaften zu suchen. Ab Abschnitt 5 folgen die verschiedenen Arten von Bezie-

2 http://www.loc.gov/bibliographic-future/rda/ [09.08.2010]; hier auch detaillierte Informationen zum Test.

3 On the record : report of the Library of Congress Working Group on the Future of Bibliographic Control, http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf [09.08.2010], S. 29: „Suspend further new developmental work on RDA until a) the use and business cases for moving to RDA have been satisfactorily articulated, b) the presumed benefits of RDA have been convincingly demonstrated, and c) more, large-scale, comprehensive testing of FRBR (...) has been carried out (...)."

4 RDA prospectus, Stand: 01.07.2009: http://www.rda-jsc.org/docs/5rda-prospectusrev7.pdf [09.08.2010], S. 2.

5 Vgl. die Kritik des Committee on Cataloging der American Libraries Association vom 20.01.2009, http://www.libraries.psu.edu/tas/jca/ccda/docs/chair46.pdf [09.08.2010], S. 7 und S. 9.

6 Deutsche Übersetzung der FRBR unter http://www.d-nb.de/standardisierung/pdf/frbr_deutsch_09.pdf [09.08.2010]; vgl. einführend: Heidrun Wiesenmüller: Zehn Jahre „Functional Requirements for Bibliographic Records" (FRBR) : Vision, Theorie und praktische Anwendung. In: Bibliothek : Forschung und Praxis 32 (2008), H. 3, S. 168-179.

hungen, die zwischen FRBR-Entitäten bestehen können: Im Abschnitt 6 geht es beispielsweise um die Beziehung zwischen der Entitätengruppe 2 und der Entitätengruppe 1 – hier findet man also die Regeln für Eintragungen unter Personen und Körperschaften. Die für die Sacherschließung relevanten Abschnitte wurden für die erste Ausgabe noch nicht ausgearbeitet – vermutlich, weil man die Ergebnisse einer Studie abwarten wollte, welche das FRBR-Modell speziell in diesem Bereich ergänzen soll.[7] Künftig werden in den entsprechenden Abschnitten vermutlich nur recht abstrakte Regeln zu finden sein, so dass RDA in Kombination mit unterschiedlichen Sacherschließungsregeln angewendet werden kann.

Speziell für deutsche Katalogisierer erhebt sich noch eine dritte Hürde: die Kompatibilität zum Vorgänger AACR2, der die Entwickler – aus nachvollziehbaren Gründen – eine sehr hohe Prioriät eingeräumt haben: „The need to integrate data produced using RDA into existing databases developed using AACR and related standards has been recognized as a key factor in the design of RDA" (0.2). Viele Regelungen versteht man wesentlich leichter, wenn man ihr Pendant in AACR2 kennt. Deshalb ist eine eingehende Beschäftigung mit AACR2 zum jetzigen Zeitpunkt die vielleicht beste Vorbereitung auf RDA.

Die Vielfalt von RDA

Vergleicht man den Informationsgehalt von Titelaufnahmen nach RAK-WB und AACR2, so stellt man fest, dass in der angloamerikanischen Welt ein merklich höherer Aufwand bei der Katalogisierung getrieben wird als bei uns: Die bibliographische Beschreibung ist nach AACR2 ausführlicher als nach RAK-WB, typischerweise werden mehr Eintragungen gemacht (z.B. unter bis zu drei Herausgebern) und Körperschaften in stärkerem Maße berücksichtigt. Auch die Entscheidungsprozesse sind oft komplexer, da AACR2 vielfach inhaltliche Kriterien anlegt, wo RAK-WB ausschließlich formal argumentiert. Nicht zuletzt gehört Sacherschließung, die i.d.R. durch dieselbe Person erfolgt wie die Formalerschließung, als Standard zu jeder Titelaufnahme dazu.

Wie hoch ist nun der Katalogisierungsaufwand bei RDA? Eine einfache Antwort darauf gibt es nicht, da das neue Regelwerk eine enorme Bandbreite erlaubt: Hochkomplexe, weit über die derzeitige AACR2-Praxis hinausgehende Katalogisate sind ebenso gut möglich wie die Verkürzung auf einige wenige Informationen. Ein Großteil der insgesamt fast 500 „RDA-Elemente" (d.h. Merkmale und Beziehungen) nämlich ist fakultativ – obligatorisch sind nur die so genannten Kernelemente (core elements), die für jeden Bereich von RDA definiert sind. Diese Kernelemente müssen stets erfasst werden, andere Elemente hingegen nur dann, wenn es ansonsten zu Verwechslungen mit einer anderen, ähnlichen Entität kommen könnte (0.6.1).

Bei den Merkmalen einer Manifestation (0.6.2) gehören beispielsweise der Sachtitel und die erste Verfasserangabe zu den Kernelementen, nicht aber Zusätze zum Sachtitel (obwohl diese nicht selten aufschlussreicher sind als der Sachtitel), Paralleltitel und weitere Verfasserangaben. Der (erste) Erscheinungsort wurde nachträglich noch zu einem Kernelement „befördert" – zu heftig war die Kritik der Community. Bei den Beziehungen zwischen einer Ressource und den mit ihr verbundenen Personen und Körperschaften (0.6.6), d.h. den Eintragungen, gilt nur der erste bzw. hauptverantwortliche geistige Schöpfer als Kernelement – also der erste Verfasser bzw. erste Urheber. Weitere Verfasser und Urheber sowie alle sonstigen beteiligten Personen (z.B. Herausgeber oder Übersetzer) und Körperschaften sind hingegen generell fakultativ, ebenso wie z.B. Eintragungen unter Schriftenreihen (0.6.8).

Ein erklärtes Ziel von RDA – die Orientierung an den Bedürfnissen der Benutzer (0.4.2.1) – wird mit diesem minimalistischen Ansatz in Frage gestellt. U.a. sollen mit RDA erstellte Daten es den Nutzern ermöglichen, „to find all resources associated with a particular person, family, or corporate body" (0.4.2.1). Aber wie soll das gehen, wenn etwa die fragliche Person als zweiter Verfasser oder Herausgeber gar keine Eintragung erhalten hat? Obligatorisch ist hingegen die Erfassung von mindestens einer Thema-Beziehung (0.6.7), d.h. gemäß RDA müssen sämtliche Werke auch sachlich erschlossen werden.

Freilich gibt es keinen Zwang, sich auf die Kernelemente zu beschränken: Verbotsregeln, wie man sie aus RAK-WB kennt (beispielsweise die Beschränkung auf maximal *eine* sonstige beteiligte Person), gibt es nicht. Aus Sicht von RDA spricht nichts dagegen, fünf Herausgeber oder 20 Verfasser zu berücksichtigen. Ja, sogar unter Verlagen können Eintragungen gemacht werden, wenn man dies für sinnvoll hält (21.3.1.3)! Ob und inwieweit solche fakultativen Elemente tatsächlich erfasst werden, ist Sache der jeweiligen Katalogisierungsstelle. Diese kann entweder Richtlinien dafür erlassen oder die Entscheidung in das Ermessen der einzelnen Katalogisierer stellen (0.6.1). Zur Vielfalt tragen außerdem zahlreiche alternative oder optionale Regelungen bei. Auch hier können die Katalogisierungsstellen entweder Anwendungsrichtlinien festlegen oder die Entscheidung dem Einzelnen überlassen (0.8).

Unter diesen Umständen wird RDA kaum zu einer Vereinheitlichung der Katalogisierung beitragen, wie man es sich eigentlich erhofft hatte. Vielmehr könnten die Titelaufnahmen noch heterogener werden als bisher – mit entsprechenden Nachteilen für den Datenaustausch. Bei einem Umstieg auf RDA wäre es deshalb wichtig, sich zumindest innerhalb der deutschsprachigen Welt auf einen gemeinsamen Mindeststandard zu einigen.

Näher an der Vorlage

Bei der bibliographischen Beschreibung bleibt RDA tendenziell näher an der Vorlage als die älteren Regelwerke. Dies entspricht einem Grundprinzip des 2009 erschienenen „Statement of International Cataloguing Principles" (ICP) – sozusagen des Nachfolgers der „Paris Principles" von 1961: „Beschreibungen und normierte Namensformen sollen sich danach richten, wie die Entität sich selbst beschreibt".[8]

Die Interpunktion wird im Normalfall nach Vorlage übernommen; aus Gründen der Klarheit können jedoch weiterhin Interpunktionszeichen ergänzt werden (1.7.3). Bei der Groß- und Kleinschreibung orientiert sich RDA – wie RAK-WB und AACR2 – an den Regeln der jeweiligen Sprache (1.7.2 und Anhang A). Eine Alternativregelung ermöglicht es allerdings, für Interpunktion, Groß- und Kleinschreibung u.ä. entweder eigene Hausregeln zu verwenden oder gar – sofern elektronische Daten maschinell übernommen werden – alles exakt nach Vorlage zu erfassen (1.7.1). Steht auf einem Titelblatt also beispielsweise „LEXIKON BUCH DRUCK PAPIER", so könnte dies in genau derselben Form in einer RDA-Titelaufnahme erscheinen – sicher keine benutzerfreundliche, wohl aber eine pragmatische Lösung für Materialarten, die nur noch maschinell erschlossen werden können.

Abkürzungen verschwinden weitgehend aus der bibliographischen Beschreibung; selbst „pages" (bzw. „Seiten") schreibt man aus. Auch sonst darf man sich auf mehr Schreibarbeit einstellen, etwa bei Verfasserangaben: Diese sind genau so zu übernehmen, wie sie in der Vorlage erscheinen (2.4.1.4) – also auch inkl. Personalangaben (akademische Titel, Institution etc.). Laut Standardregel sind außerdem sämtli-

7 Functional Requirements for Subject Authority Data (FRSAD), endgültige Fassung vom Juni 2010: http://www.ifla.org/files/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf [09.08.2010].

8 Zitiert nach der dt. Übersetzung: http://www.ifla.org/files/cataloguing/icp/icp_2009-de.pdf [09.08.2010], S. 2; vgl. RDA 0.4.3.4.

che Namen zu erfassen, auch wenn mehr als drei Personen in gleicher Funktion genannt sind. Unter Umständen muss man also mehrere Dutzend Verfasser abtippen (2.4.1.5). Doch halt – RDA lässt hier, wie an vielen anderen Stellen auch, ein „Schlupfloch", das das Festhalten an der bisherigen Praxis legalisiert: Über entsprechende Optionsregeln darf man auch weiterhin Personalangaben weglassen und bei mehr als drei Namen auf den ersten kürzen.[9] Auch Verlage sind künftig genau nach Vorlage zu erfassen (2.8.1.4), d.h. es gibt keine Weglassung von juristischen Wendungen oder Kürzung auf den Familiennamen mehr. Ob man den Benutzern damit wirklich immer einen Gefallen tut, sei dahingestellt. Die bisherige Praxis hatte jedenfalls den Vorteil, die oft unübersichtlichen Angaben in der Vorlage auf das Wesentliche zu reduzieren – etwa von „Peter Lang, Internationaler Verlag der Wissenschaften" auf ein schlichtes „Lang".

Das ICP fordert auch, dass bibliographische Beschreibungen „auf einem international akzeptierten Standard basieren"[10] sollen (für die Bibliothekswelt: ISBD). RDA hingegen steht auf dem – durchaus einleuchtenden – Standpunkt, dass zwischen der Erfassung der Daten und ihrer Darstellung eine klare Trennlinie zu ziehen ist (0.1). Definiert werden Datenelemente und ihre Inhalte; wie diese den Benutzern zu präsentieren sind, wird hingegen nicht vorgeschrieben. Eine ISBD-Darstellung ist weiterhin möglich (Details dazu finden sich im Anhang D), jedoch nicht zwingend. Dies ist z.T. auf Kritik gestoßen: „ISBD is much too important a part of cataloging to be relegated to an appendix in RDA."[11] Wirklich problematisch erscheinen jedoch nur die Bereiche, in denen RDA nicht mehr zur ISBD kompatibel ist: Beispielsweise weicht die Regelung zur Verwendung eckiger Klammern ab – diese (oder eine andere Markierung) werden gemäß RDA nur noch dann verwendet, wenn eine Information von außerhalb der Ressource stammt (2.2.4).

RDA – ein modernes Regelwerk?

Bei näherer Beschäftigung zeigt sich, dass das neue Regelwerk seinen eigenen Ansprüchen häufig nicht gerecht wird. RDA steckt in einem Dilemma: Auf der einen Seite steht das Streben nach Modernität – auf der anderen Seite eine übertriebene Rücksichtnahme auf die bisherige Praxis, die dies nicht selten konterkariert. Dies sei an einigen Beispielen demonstriert.

245 00 |a Playing the past : |b three plays by Australian women / |c edited by Kerry Kilner & Sue Tweg ; introduced by Kerry Kilner.
260 __ |a Sydney : |b Currency Press in association with the National Centre for Australian Studies and the Australian Drama Project, Monash University, |c 1995.
300 __ |a xx, 54 p. ; |c 22 cm.
440 _0 |a Current theatre series
440 _0 |a Australian drama project series ; |v no. 1
505 0_ |a Delphiniums (1942) / Catherine Shepherd -- The lighthouse keeper's wife (1922) / Mary E. Wilkinson -- Feet of clay (1928) / Betty Roland.
650 _0 |a Australian drama |x Women authors.
650 _0 |a Australian drama |y 20th century.
650 _0 |a Women |z Australia |x Drama.
651 _0 |a Australia |x Drama.
700 1_ |a Kilner, Kerry.
700 1_ |a Tweg, Sue.
700 12 |a Shepherd, Catherine, |d 1902-1976. |t Delphiniums.
700 12 |a Wilkinson, Mary E. |q (Mary Elizabeth). |t Lighthouse keeper's wife.
700 12 |a Roland, Betty. |t Feet of clay.
710 2_ |a Monash University. |b National Centre for Australian Studies.
710 2_ |a Monash University. |b Australian Drama Project.

Abbildung 4: Nebeneintragungen unter enthaltenen Werken in einer AACR2-Titelaufnahme der Library of Congress (Format MARC 21, Ausschnitt)

In der bibliographischen Beschreibung folgt RDA dem „modernen" Prinzip, eine Entität so wiederzugeben, wie sie sich selbst darstellt. Deshalb bleiben juristische Wendungen in der Verlagsangabe erhalten. Eine entsprechende Regelung hätte man nun auch bei der Ansetzung von Körperschaften erwartet. Dort jedoch findet man die AACR2-Regeln in praktisch unveränderter Form – inkl. der Anweisung, juristische Zusätze wegzulassen (11.2.2.10)!

Der Grund für diese merkwürdige Inkonsistenz war die Vorgabe, bei Ansetzungsformen möglichst keine Änderungen gegenüber AACR2 vorzunehmen – denn diese würden in den Bibliothekssystemen Probleme machen. Für deutsche Bibliothekare ist das schwer nachvollziehbar: Da die Titeldaten bei uns üblicherweise über eine Identnummer mit den Normsätzen verknüpft sind, genügt es, die Änderung im jeweiligen Normsatz durchzuführen – sofort sind auch alle zugehörigen Titeldaten wieder „richtig". Automatische Update-Verfahren sorgen dafür, dass solche Änderungen auch in den Lokalsystemen nachvollzogen werden. In der angloamerikanischen Welt steht hingegen in den Titeldaten nur die Ansetzungsform als Textstring. Ändert sich diese, so müssen auch alle betroffenen Titeldaten angepasst werden. Nicht wenige amerikanische Bibliotheken führen solche Änderungen überhaupt nicht oder nur in großen Abständen nach.[12] Diese Praxis wird – trotz ihrer offensichtlichen Untauglichkeit – von RDA sanktioniert: Die Beziehung einer Person oder Körperschaft zu einer Ressource kann entweder durch einen Identifikator (identifier) für die Person oder Körperschaft (also eine Nummer) oder durch einen „normierten Zugangspunkt" (authorized access point), d.h. die Ansetzungsform als Textstring, abgebildet werden (18.4.1).

Auch für Haupteintragungen galt „Bestandsschutz", so dass die Regelungen zumeist genau denen der AACR2 entsprechen – damit bleiben auch alle Unterschiede zu RAK-WB erhalten. Eine interessante Ausnahme findet sich bei Werken von mehr als drei Verfassern: Diese sind nun keine Sachtitelwerke mehr, sondern erhalten die Haupteintragung unter dem ersten Verfasser (6.27.1.3).

Aber warum, so fragt man sich, löst RDA sich nicht überhaupt vom antiquierten Konzept von Haupt- und Nebeneintragungen? Denn nur die Bezeichnung „main entry" ist verschwunden, nicht aber die Sache selbst. Deutlich wird dies bei den Regeln zur Bestimmung des normierten Zugangspunktes für ein Werk (6.27.1) – ganz klar spiegelt sich hier die alte Dreiteilung in Verfasserwerk, Urheberwerk und Sachtitelwerk wider: Für den Zugangspunkt wird der Vorzugstitel des Werkes entweder mit der Ansetzungsform einer Person kombiniert (z.B. „Swift, Jonathan, 1667-1745. Tale of a tub") oder mit der Ansetzungsform einer Körperschaft (z.B. „Hamline University. Biennial catalogue of Hamline University") oder er besteht – wenn weder eine Person noch eine Körperschaft als „creator" verantwortlich ist – nur aus dem Vorzugstitel (z.B. „Anthologie de la poésie baroque française").

Wiederum scheint die amerikanische Kata-

9 Allerdings heißt es nun nicht mehr z.B. „with contributions by Peter Williams … [et al.]", sondern „with contributions by Peter Williams [and twelve others]".

10 Statement of International Cataloguing Principles (wie Fußnote 8), S. 5.

11 Carlo Bianchini und Mauro Guerrini: From bibliographic models to cataloging rules: remarks on FRBR, ICP, ISBD, and RDA and the relationships between them. In: Cataloging & classification quarterly 47 (2009), H. 2, S. 105-124, hier S. 115.

12 Typischerweise werden dafür externe Dienstleister eingesetzt, die die lokalen Daten mit den aktuellen Normdateien abgleichen und die Titeldaten entsprechend aktualisieren.

```
0500 Aau
1100 2009
1130 druck
1500 eng$cfre
2000 978-0-415-44911-3
2240 BSZ: 326979549
3000 !166114731!Chavance, Bernard
3010 !181759748!Wells, Francis[Übers.]
3211 L'économie institutionelle <engl.>
4000 Institutional economics$hBernard Chavance. Transl. by Francis Wells
4020 1. publ.
4030 London [u.a.]$nRoutledge
4060 XIII, 94 S.
5550 !20896987X!|s|Institutionenökonomie
```

Abbildung 5: Vermischung der FRBR-Ebenen Werk, Expression und Manifestation (am Beispiel einer RAK-Titelaufnahme im PICA-Format des Südwestverbunds)

logisierungspraxis der Grund für dieses gerade zwanghafte Festhalten an Überkommenen zu sein: Auch Beziehungen zu Werken – beispielsweise Eintragungen unter Schriftenreihen oder enthaltenen Werken – werden nämlich mit Hilfe von Textstrings dargestellt. In der abgebildeten AACR2-Titelaufnahme (Abb. 4) sieht man solche Eintragungen in der Kategorie 700 (für die drei enthaltenen Werke, nach den Nebeneintragungen für die beiden Herausgeber). Die Textstring-Methode bleibt auch nach RDA erlaubt (24.4.2). Der dafür zu verwendende normierte Zugangspunkt ist (wie oben beschrieben) abhängig von der Haupteintragung, auf die man deshalb nicht verzichten kann. RDA konserviert damit ein aus der Zeit der Zettelkataloge stammendes Konzept, welches deutsche Katalogexperten bereits vor einem Jahrzehnt aufgeben wollten.[13]

„RDA is designed for use in an international context" (0.11.1) – doch die versprochene Internationalisierung bleibt an der Oberfläche: Sind etwa englische Formulierungen vorgeschrieben (z.B. „publisher not identified"), so können diese von Katalogisierungsstellen außerhalb des englischen Sprachraums in die jeweilige Arbeitssprache übersetzt werden (0.11.2). Nicht gerüttelt wurde hingegen beispielsweise an den AACR2-Regeln für Werke aus den Bereichen Recht, Religion und amtliche Mitteilungen, die viel zu sehr auf die angloamerikanische Welt zugeschnitten sind. Als Reaktion auf entsprechende Kritik in einer deutschen Stellungnahme von 2008 wies das Lenkungsgremium JSC (Joint Steering Committee for Development of RDA) – nicht wirklich überraschend – auf das Problem der Kompatibilität mit den vorhandenen Daten hin; eine stärkere Internationalisierung könne vielleicht in späteren Ausgaben von RDA erfolgen.[14]

FRBRisierung à la RDA

Auch bei der Umsetzung von FRBR fehlt es derzeit noch an Konsequenz. In herkömmlichen Katalogisaten nach RAK-WB oder AACR2 sind die Ebenen Werk, Expression und Manifestation vermischt. Dies zeigt etwa die Titelaufnahme für die Druckausgabe der englischen Übersetzung von „L"économie institutionelle" (Abb. 5): Die meisten Kategorien, z.B. Erscheinungsjahr, ISBN, Sachtitel- und Verfasserangabe, Ausgabebezeichnung, Erscheinungsvermerk und Umfangsangabe, gehören auf die Ebene der Manifestation. Auf der Ebene der Expression befinden sich Sprachangabe und Übersetzer; auf der Ebene des Werkes schließlich Verfasser, Einheitssachtitel und Sacherschließung.

In einem FRBRisierten Katalog sollen Benutzer zwischen den Ebenen navigieren und sich beispielsweise alle zu einem Werk gehörigen Expressionen oder Manifestationen anzeigen lassen können. Auf den ersten Blick scheint RDA dies zu ermöglichen: Die so genannten Primärbeziehungen – also die Beziehungen zwischen einem Werk, seinen Expressionen, Manifestationen und Exemplaren – sollen in RDA-Katalogisaten abgebildet werden (17.2), und das manifestierte Werk ist ein Kernelement (17.3). Auf den zweiten Blick sieht die Sache freilich anders aus: Zum einen darf die Ebene der Expression übersprungen werden (17.4.1), zum anderen findet sich wieder einmal ein „Schlupfloch": Die Beziehung zum Werk kann nämlich nicht nur über die üblichen Konventionen – Identifier oder normierter Zugangspunkt – ausgedrückt werden, sondern auch durch eine zusammengesetzte Beschreibung (composite description): „Prepare a composite description that combines one or more elements identifying the work and/or expression with the description of the manifestation" (17.4.2.3). Dies freilich beschreibt nichts anderes als den Status quo in unseren Katalogen. Wer möchte, kann also offenbar so weitermachen wie bisher.

Hoffen auf die Zukunft

RDA macht gegenüber AACR2 fraglos viele gute Schritte nach vorne – und bleibt doch hinter den Erwartungen zurück. Die erste Ausgabe spiegelt vermutlich das wider, was der angloamerikanischen Bibliothekswelt zum jetzigen Zeitpunkt zugemutet werden konnte. Für die Zukunft wünscht man sich in manchen Bereichen mutigere Entscheidungen. Bei den Verantwortlichen scheint – diesen Eindruck hat die Verfasserin in persönlichen Gesprächen gewonnen – die Bereitschaft dafür durchaus vorhanden zu sein. Auf einer Tagung in Kopenhagen im August dieses Jahres[15] betonten mehrere Vortragende, dass RDA ein „work in progress" sei. Man darf deshalb gespannt darauf sein, wie sich das neue Regelwerk weiter entwickeln wird. Gerade der Input aus europäischen Ländern, der bisher leider nur recht spärlich berücksichtigt wurde, könnte sich dabei als wertvoll erweisen.

AUTORIN

Prof. Heidrun Wiesenmüller M.A

Hochschule der Medien
Wolframstraße 32
70191 Stuttgart
wiesenmueller@hdm-stuttgart.de

13 In der – leider unfertig abgebrochenen – Überarbeitung der RAK-WB unter dem Arbeitstitel RAK2 war geplant, die Unterscheidung zwischen Haupt- und Nebeneintragung zugunsten von gleichwertigen Sucheinstiegen aufzugeben.

14 Vgl. die dt. Stellungnahme, http://www.d-nb.de/standardisierung/pdf/comments_rda_full_draft.pdf [09.08.2010], S. 24, und die Antwort des JSC, http://www.d-nb.de/standardisierung/pdf/jsc_response_0811.pdf [09.08.2010], S. 3.

15 RDA in Europe – making it happen, ausgerichtet von der European RDA Interest Group, http://www.slainte.org.uk/eurig/meetings.htm [09.08.2010].