

Mehr aus der Schlagwortnormdatei herausholen

Implementierung einer geographischen Facette in den Online-Katalogen der UB Heidelberg und der UB Mannheim

Heidrun Wiesenmüller, Leonhard Maylein und Magnus Pfeffer

Heutige Bibliothekskataloge werten nur einen Bruchteil der in den Schlagwortnormsätzen abgelegten Informationen aus, obwohl die Recherche dadurch signifikant verbessert werden könnte. So können die ISO-Ländercodes, die nicht nur bei Geographika, sondern z.B. auch bei Personen erfasst werden, als Basis für eine Einschränkung nach dem geographischen Raum dienen. Ein erster Ansatz zur Einbeziehung dieser Ländercodes in die Recherche stellt die Einführung einer geographischen Facette dar, wie sie in den auf Suchmaschinentechologie basierenden Katalogen der UB Heidelberg und der UB Mannheim realisiert wurde.

Making the most of the subject authority file – implementing a geographic facet in the catalogs of the Heidelberg and Mannheim University Libraries

In current catalogs, only a fraction of the information stored in subject authority records is actually made use of, although it could significantly enhance retrieval. The ISO country codes, which are recorded not only for geographic subject headings but also e.g. for persons, can be used for limiting a search geographically. As a first approach to integrating these country codes into the retrieval, a geographic facet can be introduced. This has been implemented in the catalogs of the Heidelberg and the Mannheim University Library, which are both based on search engine technology.

Einleitung

■ Datensätze der Schlagwortnormdatei (SWD) enthalten eine Fülle von Informationen, die bisher nur zu einem erschreckend geringen Teil für die Recherche in Bibliothekskatalogen nutzbar gemacht werden. Als Beispiel seien die folgenden beiden Datensätze – ein Sachschlagwort und ein Geographikum – betrachtet, die hier im PICA-Format präsentiert werden (leicht gekürzt¹):

Beispiel 1:

800 |s|Web log
808 |a|Networks Internet-Wörterbuch
810 30
816 006.752
830 |s|Netztagbuch
830 |s|Blog <Web-Seite>
830 |s|Weblog *Du. Informatik
845 |s|Web-Seite / Server / Logdatei
850 |s|Soziale Software
860 |s|Mikroblog

Beispiel 2:

800 |g|Schwäbische Alb

808 |a|B 1996, Geo-Du.
810 19.1b
811 XA-DE-BW;XA-DE-BY
816 T2--43473
830 |g|Schwabenalb
830 |g|Schwäbischer Jura
830 |g|Rauhe Alb
830 |g|Jura / Schwaben
850 |s|Mittelgebirge

Abgesehen von der Angabe der Quelle für die Ansetzungsform (Kat. 808), die i.d.R. nur für Erschließer von Interesse ist, könnten alle erfassten Kategorien für die Recherche von Nutzen sein. Standardmäßig werden jedoch nur die Ansetzungsform (Kat. 800) und die Synonyme (Kat. 830) ausgewertet. Die übrigen Thesaurusrelationen – der Oberbegriff (Kat. 850), der verwandte Begriff (Kat. 860) und ggf. der so genannte mehrgliedrige Oberbegriff² (Kat. 845)

2 Der mehrgliedrige Oberbegriff (auch „definitorische Verweisung“ bzw. „definitorischer Oberbegriff“) stellt eine Sonderform der hierarchischen Verweisung dar, die vor allem bei Individualbegriffen verwendet wird. Dabei wird eine Schlagwortfolge erfasst, „die der Beschlagwortung (...) ohne Verwendung des Individualnamens entsprechen würde“ (RSWK § 12,4). Beispielsweise erhalten alle Naturschutzgebiete in Baden-Württemberg den mehrgliedrigen Oberbegriff „g. Baden-Württemberg / Naturschutzgebiet“.

– bleiben in den meisten Katalogen ungenutzt. Als Ausnahmen sind u.a. die Kataloge der Deutschen Nationalbibliothek (DNB) und des Südwestdeutschen Bibliotheksverbands (SWB) zu nennen, in welchen diese Kategorien genauso indexiert sind wie Synonyme.³ Diese erweiterte Indexierung führt dazu, dass deutlich mehr Titel gefunden werden, die für den Suchenden von Interesse sein dürften: Recherchiert man etwa mit „Soziale Software“, so werden nicht nur Titel mit dem Schlagwort s. *Soziale Software* ausgegeben, sondern auch solche mit dessen Unterbegriffen (z.B. s. *Web log*, s. *Wiki*) und verwandten Begriffen (z.B. s. *World Wide Web 2.0*, s. *Social Tagging*). Die Methode bringt allerdings auch einen gewissen Ballast in der Treffermenge mit sich und sollte deshalb möglichst mit einem Ranking kombiniert werden.⁴

Die Kat. 816 wurde vor einigen Jahren im Zusammenhang mit dem Projekt CrissCross eingeführt, bei dem eine Konkordanz zwischen der SWD und der Dewey-Dezimalklassifikation (DDC) aufgebaut wird.⁵ In ihr werden die dem jeweiligen Schlagwort zugeordneten DDC-Notationen erfasst. Eine Nutzung dieser Informationen für die Recherche ist von der DNB seit langem beabsichtigt, bisher jedoch nicht umgesetzt worden. Im Heidelberger HEIDI-Katalog wird diese Information – neben anderen klassifikatorischen Angaben – für die neu eingerichtete Fachfacette ausgewertet. Darüber hinaus gibt es in den Datensätzen

3 Eine etwas andere Strategie wird an der UB Heidelberg verfolgt: Im HEIDI-Katalog werden nur die Oberbegriffe in die erweiterte Indexierung mit aufgenommen. Hintergrund ist die Überlegung, dass bei verwandten Begriffen auch Gegensatzpaare vorkommen.

4 Bei einer solchen weiten Indexierung wird etwa bei der Recherche nach „Mittelgebirge“ auch Literatur zu sämtlichen Mittelgebirgen wie z.B. der Schwäbischen Alb ausgeworfen. Das Ranking würde Treffer zu Unterbegriffen und verwandten Begriffen des eingegebenen Suchbegriffs erst weiter hinten in der Trefferliste anzeigen. Im Heidelberger HEIDI-Katalog wird dies künftig über einen niedrigeren Feld-Boost für Oberbegriffe umgesetzt.

5 Zum Projekt vgl. u.a. Jessica Hubrich: CrissCross: SWD-DDC-Mapping. In: VÖB-Mitteilungen 61 (2008) 3, S. 50-58.

1 Aus Gründen der Übersichtlichkeit wurden die folgenden Datenelemente weggelassen: Datensatzcodierung, Verwendungskennzeichen, SWD-Identnummer, Entitätencodierung für den Umstieg auf die Gemeinsame Normdatei (GND), Permutationen des mehrgliedrigen Oberbegriffs.

zwei Typen von systematisierenden Elementen: Zum einen die Notationen der SWD-Sachgruppen (Kat. 810), zum anderen den auf einem ISO-Standard basierenden Ländercode (Kat. 811). Die SWD-Sachgruppen (Kat. 810) untergliedern das Schlagwortvokabular in 36 Hauptgruppen, die in den meisten Fällen noch weiter unterteilt sind.⁶ Ein Katalog, in dem die Sachgruppen für die Benutzerrecherche genutzt werden, ist den Verfassern nicht bekannt.⁷ Bei den Ländercodes, die den Schwerpunkt des vorliegenden Beitrags bilden, ist es ganz ähnlich: Bis vor kurzem gab es nur einen einzigen Katalog – den Verbundkatalog des Kunstbibliotheken-Fachverbands Florenz-München-Rom⁸ – der sie auswertete. Dies ist umso erstaunlicher, da bereits im Jahr 1993 ein konkreter Anwendungsvorschlag publiziert wurde: Ursula Schulz wies damals im „Bibliotheksdienst“ auf die potenzielle Nützlichkeit der Ländercodes hin und forderte, diese Informationen im Katalog recherchierbar zu machen – freilich nicht als Codes, sondern in einer für Benutzer verständlichen verbalen Form.⁹

Im Frühjahr 2010 wurde das Thema an der Stuttgarter Hochschule der Medien (HdM) aufgegriffen und zunächst theoretisch in einer Bachelorarbeit untersucht.¹⁰ In einem zweiten Schritt wurde im Sommer 2010 eine Testinstallation beim Bibliotheksservice-Zentrums (BSZ) aufgesetzt, die es ermöglichte, die Überlegungen anhand realer Ländercode-Recherchen zu überprüfen: Die Erwartungen wurden dabei nicht

nur erfüllt, sondern sogar noch übertroffen! Im November 2010 wurden die Ergebnisse auf der Fortbildungsveranstaltung „Der Katalog der Zukunft (Teil 3): Die Benutzer in den Mittelpunkt!“¹¹ des VDB-Regionalverbands Südwest an der UB Stuttgart vorgestellt. In einer kleinen Arbeitsgruppe mit Vertretern der UB Heidelberg, der UB Mannheim, dem BSZ und der HdM wurde das Konzept Anfang 2011 weiter ausgearbeitet. Seither sind konkrete Umsetzungen im Heidelberger HEIDI-Katalog und im Mannheimer Primo-Katalog entstanden. Bevor diese Implementationen im Einzelnen vorgestellt werden, sei zunächst die Grundidee erläutert.

Ländercodes in der SWD

Die Ländercodes der SWD¹² beziehen sich in der Hauptsache auf die aktuelle politische Zuordnung¹³ und beinhalten die Kürzel für einzelne Staaten gemäß DIN EN ISO 3166-1 (z.B. DE für Deutschland, GB für Großbritannien, PL für Polen, EG für Ägypten, NG für Nigeria, CA für Kanada, RE für Reunion). Für die SWD wurde der ISO-Standard um zusätzliche Codes für Kontinente und Meere erweitert. Diese werden auch den Länderkürzeln vorangestellt, so dass sich u.a. die folgenden SWD-Ländercodes in einer zweistufigen hierarchischen Struktur ergeben:

XA	Europa
XA-DE	Deutschland
XA-GB	Großbritannien
XA-PL	Polen
XC	Afrika
XC-EG	Ägypten
XC-NG	Nigeria
XD	Amerika
XD-CA	Kanada
XL	Indischer Ozean
XL-RE	Reunion

Für Deutschland, Österreich und die Schweiz ist auch noch eine dritte Ebene mit den Verwaltungseinheiten (Bundesland bzw. Kanton) gemäß DIN ISO 3166-2 definiert¹⁴, z.B.:

XA-DE-BW	Baden-Württemberg
XA-DE-NW	Nordrhein-Westfalen
XA-AT-1	Burgenland
XA-CH-SG	Sankt Gallen <Kanton>

Die hierarchische Struktur ist für die Recherche ausgesprochen günstig, da man z.B. mit einer trunkierten Suche nach „XD?“ sowohl diejenigen Schlagwörter erhält, die den übergreifenden Ländercode für Amerika tragen, als auch diejenigen mit Codes für spezifische amerikanische Staaten. Entsprechend erhält man mit einer Recherche nach „XA-DE?“ Schlagwörter, die entweder den Code für Deutschland oder den für eines der deutschen Bundesländer enthalten.

Die Ländercodes werden bei allen Schlagwörtern erfasst, die einen geographischen Aspekt besitzen. Dies sind keineswegs nur Geographika, sondern auch viele andere Typen von Schlagwörtern – z.B. Ethnographika, Sprachen, Bauwerke, Körperschaften, Historische Ereignisse sowie Sachschlagwörter oder Werke mit einem eindeutigen geographischen Bezug. Der Ländercode für Baden-Württemberg (XA-DE-BW) beispielsweise wurde u.a. bei den folgenden Schlagwörtern erfasst: *g. Schwäbische Alb, g. Schluchsee, g. Vogelherdhöhle, g. Heidelberg, g. Rems-Murr-Kreis, g. Naturschutzgebiet Bühler Tal, g. Hohenzollern <Burg>, g. Ulm / Stadthaus, g. Stuttgart / Wilhelma, g. Badener, g. Schwäbisch, k. Württembergischer Ingenieurverein, k. Deutscher Orden / Kommende Kapfenburg, c. Karlsruhe / Badische Landesbibliothek, s. Badischer Aufstand, s. Ludwigsburger Porzellan, s. Reichenauer Malerschule, s. Reuchlinpreis der Stadt Pforzheim, t. Weingartner Liederhandschrift, t. Stuttgarter Nachrichten.*

Auch bei Personen werden Ländercodes vergeben, allerdings nur auf der zweiten Hierarchiestufe. Auch baden-württembergische Persönlichkeiten wie z.B. Carl Benz oder Theodor Heuss erhalten deshalb nur den Ländercode für Deutschland. Bei Bedarf können auch mehrere Ländercodes vergeben werden, i.d.R. jedoch maximal drei. Wären mehr als drei Ländercodes nötig, so wird der Ländercode für die nächsthöhere geographische Einheit vergeben: Beim Schlagwort *g. Alpen* ist deshalb nur der Ländercode für Europa (XA) erfasst – und nicht die Einzelcodes der acht Staaten, die Anteil an den Alpen haben.

Ergebnisse der Test-Recherchen

Für die Testinstallation änderte das BSZ die Indexierung des SWB-OPAC in der Weise ab, dass bei der Recherche nach einem Ländercode nicht mehr die Schlagwörter mit dem betreffenden Ländercode ausgegeben wurden, sondern stattdessen die Titel, die

6 Vgl. Deutsche Nationalbibliothek: SWD-Sachgruppen, URL: <http://www.d-nb.de/standardisierung/normdateien/swd_not_hilfe.htm>. Alle URLs wurden zuletzt am 29.07.2011 geprüft.

7 Nicht ausgeschlossen wäre eine Nutzung im Rahmen einer Fachfacette, vgl. dazu Julian Frick: Konzeption einer fachlichen Facette für einen Bibliothekskatalog am Beispiel der Universitätsbibliothek Mannheim. Bachelorarbeit an der Hochschule der Medien, Stuttgart 2011 (Betreuer: Heidrun Wiesenmüller, HdM Stuttgart; Magnus Pfeffer, UB Mannheim). Die Arbeit wird über den OPUS-Server der HdM verfügbar gemacht, URL: <<http://opus.bsz-bw.de/hdms/>>.

8 Dieser Katalog berücksichtigt in der Suchoption „Land oder Region“ auch die Ländercodes, allerdings nur bei geographischen Schlagwörtern. URL: <http://www.kubikat.org/mrbh-cgi/kubikat_de.pl>.

9 Vgl. Ursula Schulz: Einige Forderungen an die Qualität von Normdateien aus der Sicht der inhaltlichen Erschließung für Online-Kataloge. Teil 2: Geographischer Aspekt. In: Bibliotheksdienst 27 (1993) 3, S. 315-328, hier S. 323: „Die lokale und/oder Verbundsoftware muß (...) für folgende Möglichkeit der Datenaufbereitung und relationalen Datenhaltung sorgen: Wird ein Geographikum, Personen-SW oder Sach-SW, das in der SWD (...) einen Ländercode enthält, (...) vergeben, wird auch der Ländercode (...) in der Titeldatenbank recherchierbar gemacht. Eine interne Konkordanz zwischen Code und verbaler Benennung muß allerdings dafür sorgen, daß dem Benutzer kein Hantieren mit Codes zugemutet wird!“

10 Anja Siegel: Geografische Sacherschließungsdaten in Online-Bibliothekskatalogen : Analyse und Nutzungsmöglichkeiten. Unveröffentlichte Bachelorarbeit an der Hochschule der Medien, Stuttgart 2010 (Betreuer: Heidrun Wiesenmüller, HdM Stuttgart; Christof Mainberger, BSZ Konstanz).

11 Heidrun Wiesenmüller: Erschließungsdaten besser nutzen : geographische Recherche mit SWD-Ländercodes, URL: <http://www.vdb-online.org/veranstaltungen/522/wiesenmueller_laendercodes.pdf>.

12 Für eine ausführliche Übersicht über den Ländercode vgl.: Der Ländercode (LC) für SWD und PND : Leitfaden zu seiner Vergabe, URL: <http://www.d-nb.de/standardisierung/pdf/h_lc_leitfaden.pdf>. Eine Liste der Codes ist unter URL: <http://www.d-nb.de/standardisierung/normdateien/lc_hilfe.htm> abrufbar.

13 Einige wenige anwenderspezifische Codes orientieren sich an anderen Kriterien, z.B. „XW“ für Palästinenser in Geschichte und Gegenwart, „XT“ für das Römische Reich oder „XZ“ für fiktive Geographika.

14 Außerdem werden noch Südtirol (XA-IT-32) und Tibet (XB-CN-54) auf der dritten Ebene codiert.

mit diesen Schlagwörtern verknüpft sind. Dies ermöglichte es, Anfragen vom Typ „ein bestimmtes Thema in einem bestimmten geographischen Raum“ auf zweierlei Weise durchzuführen – einmal als Kombination aus dem Thema und dem (verbalisierten) Geographikum und einmal als Kombination aus dem Thema und dem entsprechenden Ländercode – und die Ergebnisse miteinander zu vergleichen. Die Testrecherchen wurden im (stichwortindexierten) Schlagwortindex ausgeführt, bei dem, wie bereits erwähnt, auch verwandte Begriffe und Oberbegriffe mit indexiert sind. Bei einem angenommenen Suchinteresse „Tourismus in Baden-Württemberg“ wurde also einmal die Suchsyntax „sw:tourismus baden-württemberg“ verwendet und einmal die Suchsyntax „sw:tourismus lcd:xa-de-bw“. Der Ländercode wurde ggf. trunziert, um alle im jeweiligen Geographikum liegenden Länder bzw. Bundesländer mit zu berücksichtigen. Beispielsweise wurde die Ländercode-Recherche beim Suchinteresse „Klimaänderung in Afrika“ mit „XC?“ durchgeführt und beim Suchinteresse „Sonnenenergie in Deutschland“ mit „XA-DE?“. Die Vermutung, dass sich der Recall bei Verwendung des Ländercodes z.T. erheblich erhöht, wurde durch die Testrecherchen eindrücklich bestätigt. Die Tabelle zeigt einige Beispiele:

Suchinteresse	mit Geographikum	mit Ländercode	Erhöhung Recall
Wirtschaftskrise in den USA	145	173	19,31 %
Muslime in Österreich	23	33	43,48 %
Umweltschutz in der Schweiz	274	399	45,62 %
Klimaänderung in Afrika	22	52	136,36 %
Tibetischer Buddhismus in Indien	35	90	157,14 %
Tourismus in Baden-Württemberg	51	179	250,98 %
Sonnenenergie in Deutschland	21	74	252,38 %
Rechtsphilosophie in Deutschland	90	447	396,67 %
Höhlen in Baden-Württemberg	8	92	1.050,00 %

Ergebnisse von Testrecherchen im SWB

Die Erhöhung des Recalls ergibt sich vor allem durch vier Faktoren. Erstens werden über den Ländercode auch Teile des betrachteten Geographikum berücksichtigt, die ja denselben Ländercode tragen: So werden beim Suchinteresse „Tourismus in Baden-Württemberg“ nicht nur Titel gefunden, die sich mit dem Tourismus in ganz Baden-Württemberg beschäftigen und deshalb mit *g. Baden-Württemberg* verschlagwortet sind (z.B. „Tourismuskonzept Baden-Württemberg“), sondern auch solche, die sich mit dem Tourismus beispielsweise in Heidelberg, dem Murgtal oder Ostwürttemberg beschäftigen (z.B. „Typisierung von Städ-

touristen am Beispiel von Heidelberger Besuchern“, „Das Murgtal als Fremdenverkehrsgebiet“, „Touristisches Entwicklungskonzept für die Region Ostwürttemberg“). Derartige, vermutlich durchaus relevante Titel werden bei einer Suche mit dem Geographikum „Baden-Württemberg“ nicht gefunden, da sie – gemäß dem Prinzip des engen Schlagworts – mit den entsprechenden spezielleren Schlagwörtern versehen sind (*g. Heidelberg, g. Murgtal, g. Ostwürttemberg*). Zweitens werden bei einer trunzierten Suche mit dem Ländercode auch diejenigen Geographika berücksichtigt, die auf einer untergeordneten Hierarchiestufe liegen: Beispielsweise werden beim Suchinteresse „Klimaänderung in Afrika“ auch Titel zu einzelnen afrikanischen Ländern gefunden (z.B. „Climate change and adaptation in Nigeria“). Auch in diesem Fall bietet die Ländercode-Recherche also einen Ausgleich zum Prinzip des „engen Schlagworts“, das sich in der Recherche häufig kontraproduktiv auswirkt.

Drittens werden bei der Ländercode-Recherche auch Individualbegriffe (Instanzen) im jeweiligen geographischen Raum berücksichtigt. Beispielsweise wird beim Suchinteresse „Höhlen in Baden-Württemberg“ nicht nur die allgemeine Literatur gefunden, sondern auch Literatur zu einzelnen Höhlen wie der Falkensteiner Hö-

le (*g. Falkensteiner Höhle*) oder dem Geißenklösterle (*g. Geißenklösterle*; das Stichwort „Höhle“ ist hier nur im Oberbegriff enthalten) – denn die Schlagwörter tragen ebenfalls den Ländercode für Baden-Württemberg. Dieser Typ führt zu einer besonders großen Erhöhung des Recalls; die Trefferlisten wachsen leicht auf das Zehn- bis Zwanzigfache an.

Viertens werden über den Ländercode auch Personen berücksichtigt: Beim Suchinteresse „Rechtsphilosophie in Deutschland“ erscheinen in der Trefferliste u.a. auch Treffer zu individuellen deutschen Rechtsphilosophen wie Alfred Verdross und Gustav

Radbruch. Auch dies ist ein sehr positiver Effekt, da viele Themen auch sehr stark von Personen geprägt ist – entsprechend finden sich in den Schlagwortfolgen dann keine geographischen Schlagwörter, sondern Personenschlagwörter. Bereits Ursula Schulz hob diesen Aspekt besonders hervor.¹⁵ Allerdings kann der Ländercode natürlich nur dann weiterhelfen, wenn auch ein entsprechendes Sachschlagwort vorhanden ist – nicht hingegen, wenn die Schlagwortfolge z.B. nur aus dem Personenschlagwort besteht. Auch werden Ländercodes bei Personen, wie bereits erwähnt, nur auf der zweiten Hierarchieebene vergeben, so dass eine Suche mit einem Ländercode der dritten Stufe grundsätzlich keine Personentref-fer ergeben kann.

Körperschaftsschlagwörter mit dem entsprechenden Ländercode tragen nur gelegentlich zur Erhöhung der Treffermenge um relevante Titel bei, können aber in manchen Fällen zu Ballast führen: Beispielsweise wurde beim Suchinteresse „Umweltschutz in der Schweiz“ u.a. der Titel „The WTO and international environmental law“ ausgegeben, der sich inhaltlich nicht mit der Schweiz beschäftigt. Auslöser ist hier das Schlagwort *k. World Trade Organization* mit dem Ländercode für die Schweiz denn die WTO hat ihren Sitz in Genf. Das Problem ist jedoch leicht lösbar, denn solche Körperschaften besitzen außerdem den Code „XP“ für internationale Organisationen und können über dieses Kriterium ausgefiltert werden.

Davon abgesehen, wurden bei den den Testrecherchen kaum Precision-Probleme festgestellt. Gelegentlich kam es zu Ballast aufgrund von Fehlverknüpfungen: Beispielsweise werden beim Suchinteresse „Muslime in Österreich“ mehrere Festschriften gefunden, die sich nur ganz allgemein mit dem Islam beschäftigen, in denen aber eine österreichische Persönlichkeit gefeiert wird (z.B. „Islam and its cultural divergence : studies in honor of Gustave E. von Grunebaum“). Das Sachthema kommt aus der ersten Schlagwortfolge (*s. Islam* ist verwandter Begriff zu *s. Muslim*), der Ländercode aus dem Personenschlagwort in der zweiten Schlagwortfolge:

s. Islam ; s. Kultur ; z. Geschichte ; f. Aufsatzsammlung
p. Von Grunebaum, Gustave E. ; f. Bibliographie

¹⁵ Vgl. Schulz (wie Anm. 9), S. 319f.: „Jeder Kollege/ jede Kollegin, der/die eine geisteswissenschaftliche Disziplin betreut, möge sich überlegen, wie viel Prozent der zu erschließenden Dokumente sich auf eine Person beziehen und somit durch ein Personen-SW, nicht aber durch ein Geographikum erschlossen werden. In den Philologien mag dieser Anteil über 50 % liegen. Betroffen sind aber alle geisteswissenschaftlichen Disziplinen, deren Gegenstände zu großen Teilen geistige Produkte von Personen sind (...).“

Dies hat freilich nichts mit der Ländercode-Recherche zu tun, sondern ist ein grundsätzliches Problem, das auch in anderen Zusammenhängen zu Ballast führt und das in technischer Hinsicht durchaus lösbar wäre.¹⁶ Insgesamt ergaben die Testrecherchen also, dass eine Nutzung der Ländercodes den Recall deutlich erhöht, ohne die Precision zu verschlechtern. Zugleich würde dadurch das Input-Output-Verhältnis der Sacherschließung merklich verbessert, indem bisher ungenützte Informationen in den Normdaten endlich bei den Benutzern ankommen würden.

Einrichtung einer Geofacette

Die Kataloge der UB Heidelberg und der UB Mannheim arbeiten beide mit Suchmaschinentechnologie: Der HEIDI-Katalog¹⁷ ist eine Eigenentwicklung, die ursprünglich auf Lucene basierte und mittlerweile auf SOLR umgestellt wurde. In Mannheim wird die suchmaschinen-basierte Software „Primo“ von ex Libris eingesetzt.¹⁸ Die Suchmaschinentechnologie erlaubt es, große Treffermengen durch ein Drill-down nach unterschiedlichen Kriterien zu verkleinern. Dafür existieren in beiden Katalogen verschiedene Facetten, z.B. für Erscheinungsjahr oder Sprache. Es bot sich an, den geographischen Bezug in einem ersten Schritt ebenfalls als eine solche Drill-down-Facette zu implementieren.

Über die neue, als „Geographischer Bezug“ bezeichnete Geofacette in den beiden Katalogen lässt sich eine (auf beliebige Weise entstandene) Treffermenge geographisch gemäß der Ländercodes eingrenzen. Diese erscheinen in der Facette in verbalisierter Form und stehen sowohl auf Deutsch als auch auf Englisch zur Verfügung. In beiden Systemen war es möglich, dabei auch die Hierarchien der Ländercodes abzubilden. Wie bei Drill-down-Menüs üblich, werden anfangs nur die ersten Einträge der Geofacette angezeigt; über einen Link kann dann die gesamte Länderliste ausgeklappt werden. Basis dafür ist eine entsprechende Aufbereitung der Ländercodes. Dabei ist zu beachten, dass es bei der Datenhaltung von bibliographischen Daten einen grundsätzlichen Unterschied zwischen klassischen Bibliothekssystemen und Suchmaschinen gibt: Während Suchmaschinen erwarten, alle zu indexierenden Informationen in den einzelnen Dokumenten zu finden,

Geographischer Bezug

Europa (39)
 - Österreich (1)
 - - Niederösterreich (1)
 - Belgien (1)
 - Deutschland (19)
 - - Baden-Württemberg (2)
 - - Bayern (1)
 - - Hessen (1)
 - - Nordrhein-Westfalen (2)
 - - Rheinland-Pfalz (1)
 - - Schleswig-Holstein (1)
 - Frankreich (1)
 - Portugal (1)
 - Russland (1)
Asien (2)
 - China (1)
 - Südkorea (1)
Afrika (2)
 - Angola (1)
 - Namibia (1)
 - Südafrika (Staat) (1)
Amerika (3)
 - USA (2)
 - Venezuela (1)
Atlantischer Ozean (1)
Internationale Staatengemeinschaften, internationale Organisationen (3)
Gesamte Welt, Übrige Welt (1)
Römisches Reich (1)
Ausblenden der letzten 23 ▲

Abbildung 1

sind in Bibliothekssystemen die Informationen in der Regel verteilt. Dort werden Titel und Normdaten in getrennten Tabellen gespeichert und über Identifikatoren verknüpft. Bei der Aufbereitung der Daten für das Laden in der Suchmaschine werden für jeden Titel all diese Verknüpfungen nachverfolgt und die Informationen in den Titeldatensatz expandiert. Für die technische Umsetzung der Geofacette musste diese Expansion bei den Schlagwortnormdaten um das Feld für den Ländercode erweitert werden. Ist ein Titel beispielsweise mit dem Schlagwort *g. Siebenbürgen* verknüpft, so sieht dies im MAB-Titeldatensatz des Mannheimer Aleph-Lokalsystems zunächst so aus:

```
902 L $$gSiebenbürgen$$9MA0166199
```

Bei der Expansion werden nun sowohl die Verweisungsformen als auch der Ländercode aus dem zugehörigen Normdatensatz in den Titeldatensatz kopiert, so dass der Suchmaschine die folgende, erweiterte Datenbasis zur Verfügung steht:

```
902-1 L $$gSiebenbürgen$$aSiebenbürgen
```

```
952-1 L $$aXA-RO
```

```
952-1 L $$aSiebenbürgen <Fürstentum>
```

```
952-1 L $$aTranssilvanien
```

```
952-1 L $$aTransylvanien
```

```
952-1 L $$aTransilvania
```

```
952-1 L $$aTransylvania
```

```
952-1 L $$aArdeal
```

```
952-1 L $$aErdély
```

Bei der Umsetzung der Geofacette traten einige interessante Aspekte auf: Diskutiert wurde u.a., welche Reihenfolge sich für die Anzeige der Ländercodes in der Facette am besten eignet. Eine spannende Frage ist auch, ob man in der Facette einen Unterschied zwischen denjenigen Titeln machen soll, die sich genuin auf ein bestimmtes Geographikum beziehen, und solchen, bei denen sich dieser Bezug nur indirekt über die Hierarchie ableiten lässt: Soll es also beispielsweise einen getrennten Eintrag für Titel mit dem Ländercode XA-DE geben, die sich auf das gesamte Deutschland beziehen? Oder genügt ein übergreifender Eintrag „Deutschland“, der solche Titel mit denjenigen zusammenfasst, welche Ländercodes für die einzelnen deutschen Bundesländer (XA-DE-BW etc.) tragen? In Mannheim und Heidelberg sind hier z.T. unterschiedliche Entscheidungen getroffen worden, weshalb es gemäß dem derzeitigen Stand einige kleinere Unterschiede in den Implementationen gibt.

Implementierung an der UB Mannheim

Da die Menge der Codes vorab bekannt ist, wird die Geofacette im Primo-Katalog als statische Facette implementiert.¹⁹ Bei dieser kann die Reihenfolge in der Anzeige festgelegt sowie eine Übersetzung für unterschiedliche Sprachen hinterlegt werden. Die gewählte Reihenfolge folgt alphabetisch den Codes, bildet die Hierarchie durch Einrückung der Ebenen ab und berücksichtigt auch die besonderen Codes wie z.B. „XT“ für das Römische Reich. Abb. 1 zeigt eine beispielhafte Geofacette in voll ausgeklappter Form.

Um die in den geographischen Codes enthaltene Hierarchie auszunutzen und dem Benutzer einen mehrstufigen Drill-down anbieten zu können, wird die Facette mit allen direkt verknüpften sowie deren hierarchisch übergeordneten Codes befüllt. Ein mit „Baden-Württemberg“ (oder einem anderen Schlagwort, das den Ländercode „XA-DE-BW“ trägt) beschlagworteter Titel erhält also die Einträge „XA-DE-BW“, „XA-DE“ und „XA“. In Mannheim hat man sich bewusst dafür entschieden, keine Unterscheidung zwischen Titeln zu treffen, die direkt mit einem Code der ersten

¹⁹ Die Geofacette ist nur im Reiter „Katalog“ aktiv.

¹⁶ Vgl. dazu Heidrun Wiesenmüller: Schlagwortketten in Online-Katalogen: Realität und Potenzial. In: Bibliotheksdienst 42 (2008) 10, S. 1033-1059, hier S. 1051-1053.

¹⁷ URL: <<http://www.ub.uni-heidelberg.de/helios/kataloge/heidi.html>>

¹⁸ URL: <<http://www.bib.uni-mannheim.de/133.html>>



Abbildung 2

oder zweiten Ebene verknüpft sind, und solchen, die diese Verknüpfung aus der Hierarchie erhalten. Man erhält durch einen Klick auf „Deutschland“ also sowohl die Titel mit „XA-DE“ als auch solche, die sich auf ein deutsches Bundesland beziehen. Die Befürchtung war, dass die für eine solche Unterscheidung erforderlichen zusätzlichen Einträge der Form „Deutschland (übergreifend)“ die Länderliste zu sehr vergrößert hätten und für die Benutzer schwer verständlich gewesen wären. Zusätzlich zur Verwendung in der Geofacette werden die Ländercodes im Primo-Katalog auch als solche indexiert und sind damit in der Suche nach „Schlagwort“ und „alle Felder“ zu finden.

Implementierung an der UB Heidelberg

An der UB Heidelberg hat man sich dafür entschieden, für die Facette einen Unterschied zwischen den Titeln zu machen, die direkt mit einem Code der ersten oder zwei-

ten Ebene verknüpft sind, und solchen, die diese Verknüpfung aus der Hierarchie erhalten. Entsprechend fällt die Indexierung komplexer aus:

■ Ländercodes der ersten Hierarchie-Ebene: Indexiert wird der angegebene Ländercode selbst (z.B. XA). Handelt es sich um einen Ländercode, für den auch Unter-codes definiert wird, wird zusätzlich der Ländercode mit angehängtem „-00“ indexiert (z.B. XA-00 für „Europa (gesamt)“). Dieser erweiterte Code legt fest, dass sich der Titel explizit auf den gesamten Raum bezieht und nicht nur auf eine Teilregion. Er wird wie ein Code der zweiten Hierarchie-Ebene behandelt: „XA-00“ ist damit ebenso ein Untercode von „XA“ wie beispielsweise „XA-DE“.

■ Ländercodes der zweiten Hierarchie-Ebene: Indexiert wird der angegebene Ländercode selbst (z.B. XA-DE) sowie der übergeordnete Ländercode (z.B. XA). Handelt es sich um einen Ländercode, für den auch Unter-codes (auf der dritten Ebene) definiert wird, wird auch hier zusätzlich der Ländercode mit angehängtem „-00“ indexiert (z.B. XA-DE-00 für „Deutschland (gesamt)“).

■ Ländercodes der dritten Hierarchie-Ebene: Indexiert wird der angegebene Ländercode selbst (z.B. XA-DE-BW) sowie die übergeordneten Ländercodes auf der ersten und zweiten Hierarchie-Ebene (z.B. XA und XA-DE).

In der Geofacette von HEIDI (Abb. 2) stellt sich dies so dar, dass es z.B. einen Eintrag „Deutschland (gesamt)“ gibt, der auf derselben Ebene wie die Einträge für die deutschen Bundesländer angesiedelt ist. Über diesen können gezielt diejenigen Titel angewählt werden, die sich originär auf ganz Deutschland beziehen. Der eine Hierarchiestufe höher liegende Eintrag „Deutschland“ vereint diese Titel mit denen, die sich auf die Bundesländer beziehen.

Gesondert behandelt werden Schlagwörter mit dem Ländercode XP: Aufgrund der oben beschriebenen Problematik bei Schlagwörtern für internationale Organisationen werden in diesen Fällen die weiteren Ländercodes des Schlagworts ignoriert.

In der Heidelberger Facette gibt es außerdem einen eigenen Eintrag für diejenigen Treffer, bei denen sich keine Aussage über den behandelten geographischen Raum machen lässt („keine Angabe“); typischerweise sind dies Titel ohne Beschlagwortung. Ebenfalls extra angezeigt werden die Treffer, die zwar beschlagwortet sind, jedoch kein Schlagwort mit einem Ländercode besitzen – hier handelt es sich also um allgemeine Titel, die sich nicht auf einen bestimmten geographischen Raum beziehen („räuml. nicht eingeschränkt“). Im ersten Fall wird beim Titel der Pseudo-Code



„Sind Ihre Etiketten wirklich so alterungsbeständig, wie beschrieben?“

Ja.

Sogar deutlich über 100 Jahre.

Warum?

Das erfahren Sie im Internet.

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

pleuser.de

„null“ indexiert, im zweiten Fall der Pseudo-Code „00“.

Der HEIDI zugrundeliegende SOLR-Index liefert die Ländercodes und Pseudo-Codes absteigend nach Auftrittshäufigkeit in der Treffermenge an das Frontend. Dort erfolgt eine Umsortierung: Bei der Ausgabe der Facettenwerte im Rahmen einer Recherche werden zunächst (soweit vorhanden) die Pseudo-Codes „00“ und „null“ ausgegeben. Anschließend werden die Codes der ersten Hierarchie-Ebene und nach diesen (eingerrückt) jeweils die untergeordnete Codes ausgegeben. Auf jeder Hierarchie-Ebene bleibt die Sortierung nach Häufigkeit erhalten.

Eine Besonderheit der HEIDI-Implementa-tion ist es schließlich, dass – anders als sonst in Drill-down-Menüs üblich – auch mehrere Einträge der Facette zugleich ausgewählt

bzw. ausgeschlossen werden können. Gerade für eine geographische Einschränkung ist dies eine sehr sinnvolle Option.

Denkbare Verbesserungen

Da die Geofacetten auf den Ländercodes der SWD beruhen, greifen sie naturgegebenmaßen nur bei beschlagworteten Titeln. Dies ist ein weiteres Argument dafür, wie wichtig es ist, den Anteil beschlagworteter Titel in unseren Katalogen zu erhöhen. Eine enorme Verbesserung könnte allein dadurch erreicht werden, dass die vorhandenen Sacherschließungsinformationen über Verbundgrenzen hinweg sozusagen „gepooht“ würden.²⁰ Zu prüfen ist außerdem, ob die nicht beschlagworteten Titel auch auf anderem Weg mit Ländercodes versehen werden könnten – entweder auf

der Basis von Klassifikationen (RVK, DDC) oder über eine maschinelle Auswertung ihrer Sachtitel.

Ist erst einmal eine breite Basis an Titeln vorhanden, denen Ländercodes zugeordnet werden können, so könnte man zusätzlich zum Drill-down auch eine direkte geographische Suche implementieren – z.B. über eine zusätzliche Auswahlbox in der gefilterten Suche.

Vor dem Hintergrund der neuen Anwendungsmöglichkeiten ist außerdem die Vergabe der Ländercodes neu zu überdenken. So sollte man einem Schlagwort künftig auch mehr als drei Ländercodes zuweisen können, damit Schlagwörter wie *g. Alpen* in differenzierter Weise für die Geofacette ausgewertet werden können. Wünschenswert wäre außerdem die Einführung einer dritten Hierarchie-Ebene bei den USA, um eine Einschränkung auf Ebene der amerikanischen Bundesstaaten zu ermöglichen.

Schon jetzt aber stellen die Geofacetten eine äußerst nützliche Funktionalität dar, die den Benutzern einen echten Mehrwert bringt. Zugleich zeigen sie exemplarisch, wie man bisher ungenutzte Elemente in den Normdaten mit verhältnismäßig geringem Aufwand für eine komfortable Nutzung aufbereiten kann.

AUTOREN

LEONHARD MAYLEIN

Informationstechnologie und
DV-Anwendungen Universitätsbibliothek
Plöck 107-109
69117 Heidelberg
maylein@ub.uni-heidelberg.de

PROF. MAGNUS PFEFFER

Studiengang Bibliotheks- und Informationsmanagement
Hochschule der Medien
Wolframstraße 32
70191 Stuttgart
pfeffer@hdm-stuttgart.de

PROF. HEIDRUN WIESENMÜLLER

Studiengang Bibliotheks- und Informationsmanagement
Hochschule der Medien
Wolframstraße 32
70191 Stuttgart
wiesenmueller@hdm-stuttgart.de

Mit der weltgrößten
Messe für Technische
Kommunikation!

Bei Vorlage dieser Anzeige erhalten
Sie eine kostenlose Eintrittskarte
für die Messe.

Jahrestagung

mit **tcworld**
conference 2011

Rhein-Main-Hallen, Wiesbaden
18.–20. Oktober 2011

PDF Synergien
Mobile XLIFF
Dokumentation
Social Web
Customer Experience
Content Strategies
XML
Saas
Compliance
International Management
Networking

www.tekom.de/tagung

fachbuchjournal www.b-i-t-online.de

²⁰ Dazu könnte ein von Magnus Pfeffer entwickeltes Matching-Verfahren zum Einsatz kommen, vgl. dazu Self-Classification (Blog), URL: <<http://blog.bib.uni-mannheim.de/Classification/>>.