Kurt Schneider

13 Jahre Kataloganreicherung in der Deutschen Nationalbibliothek

Neue Wege zur Erschließung der Buchbestände des 20. und 21. Jahrhunderts

Benutzerinnen und Benutzer von Bibliothekskatalogen wollen bei ihrer Suche nach Medien nicht ausschließlich bibliografische Daten nutzen. Sie möchten selbstverständlich darüber hinausgehend direkt aus dem Katalog heraus auf Zusatzinformationen über für sie interessante Werke zugreifen können. Von besonderer Relevanz sind dabei zum einen Informationen über ein Werk, wie sie heute etwa in Form von Cover-Abbildungen, Klappentexten oder Rezensionen bereitgestellt werden. Zum anderen, und in der Regel von noch größerer Bedeutung, ist der unmittelbare Zugriff auf das Werk selbst oder Teile davon, um sich beispielsweise mit einem Blick in das Inhaltsverzeichnis zu informieren oder auf der Basis einer Lese- oder Hörprobe einen ersten Eindruck zu verschaffen.

Nützliche Zusatzinformationen: Cover, Klappentext, Inhaltsverzeichnis

Das war nicht immer so. Bevor Zusatzinformationen dieser Art Eingang in Bibliothekskataloge fanden, waren es ab Mitte der 1990er-Jahre große Internetplayer wie Amazon und Google, die den Bedürfnissen ihrer Nutzerinnen und Nutzer nach mehr und tiefergehenden Informationen zu den gesuchten Medien entgegenkamen.

Kataloganreicherung in Bibliotheken

In der Bibliothekswelt wurden diese Veränderungen erst ein paar Jahre später wahrgenommen. Die Vorarlberger Landesbibliothek ermöglichte ab 2002 als erste Bibliothek im deutschsprachigen Raum ihren Benutzerinnen und Benutzern den direkten Zugriff etwa auf digitalisierte Inhaltsverzeichnisse[1]. Zu den Pionieren gehörte ab 2003 auch die Deutsche Nationalbibliothek (DNB), die damit begann, sogenannte Inhaltstexte in ihre Katalogdatenbank zu übernehmen. Dabei handelt es sich um mehr oder weniger lange Beschreibungen zum Inhalt eines Werkes ähnlich der Klappentexte in Büchern. Diese Inhaltstexte werden von den Verlagen an die Marketing- und Verlagsservice des Buchhandels GmbH (MVB) gemeldet, von der sie die Deutsche Nationalbibliothek seither laufend bezieht.

Angaben aus der Verlagsmeldung

„Unser ganzes Leben“ : Die Fans des BVB / von Ulrich Hesse, Gregor Schnittker

80.000 Besucher pro Spiel: Beim Zuschauerzuspruch hält Borussia Dortmund einen einsamen Rekord. Die Fans und ihre Kultur prägen wesentlich das Gesicht der Stadt.Dieses aufwendig recherchierte und gestaltete Buch erzählt die Geschichte der BVB-Fanbewegung von den (kleinen) Anfängen bis heute. Veteranen der fünfziger Jahre kommen ebenso zu Wort wie »Kutten-Fans« oder die heutigen Ultras. Man kann sich sattlesen an originellen Anekdoten und skurrilen Begebenheiten, nichtunterschlagen werden dabei Problembereiche wie das Wirken der rechtsradikalen »Borussen-Front«.
Dutzende junger und älterer Fans steuerten für das Buch private Fotoschätze bei, vom Leben auf den Kurven, von Auswärtsfahrten, von den großen Choreografien. Kurzum: Für Außenstehende ist das Buch interessant, für BVB-Fans das reinste Poesie-Album.

Typischer Inhaltstext

Anreicherung mit Inhaltstexten

Zunächst wurden die Inhaltstexte den Nutzerinnen und Nutzern des Katalogs jedoch nur indirekt über die Suche zugänglich gemacht. Sie wurden soweit technisch möglich in die Katalogdatenbank übernommen und indexiert, jedoch nicht über die Kataloganzeige angeboten. Man ging damals davon aus, dass das unmittelbare Nebeneinander von nationalbibliografisch gesicherten Informationen und den wenig normierten, teilweise werbenden Verlagsinformationen Irritationen auf Seiten der Nutzerinnen und Nutzer hervorrufen würde.

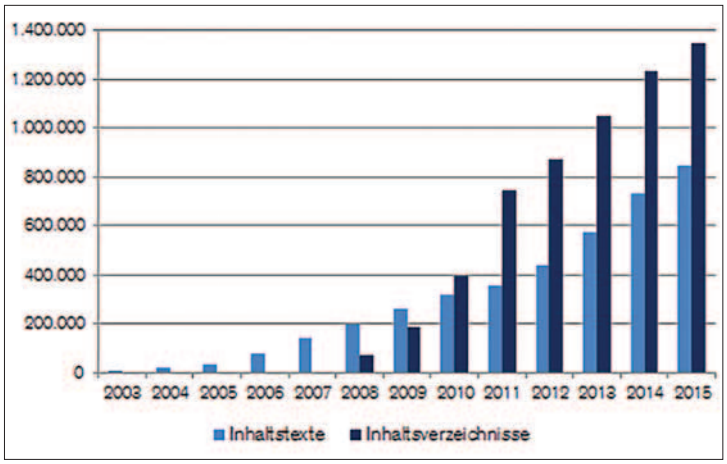
Erst ab dem Jahr 2005 änderte sich die Datenübernahme und -präsentation der Inhaltstexte grundlegend. Seither werden die Inhaltstexte aus den Metadatenlieferungen der MVB automatisch extrahiert, in eigene Textdateien auf einem separaten Server geschrieben und via Link im Titeldatensatz als Verlagsinformation zugänglich gemacht.

Datenlieferung via MVB-Meldung

Link zu diesem Datensatz	http://d-nb.info/1028365136
Titel/Bezeichnung	„Unser ganzes Leben" : die Fans des BVB / Ulrich Hesse ; Gregor Schnittker
Person(en)	Hesse, Ulrich Schnittker, Gregor
Verleger	Göttingen : Verl. Die Werkstatt
Erscheinungsjahr	2013
Umfang/Format	331 S. : zahlr. Ill. ; 28 cm
ISBN/Einband/Preis	978-3-7307-0014-3 Pp. : ca. EUR 24.90 (DE), ca. EUR 25.60 (AT), ca. sfr 35.50 (freier Pr.)
EAN	9783730700143
Sprache(n)	Deutsch (ger)
Schlagwörter	Borussia Dortmund ; Fußballfan ; Geschichte
DDC-Notation	796.33409435633 [DDC22ger]
Sachgruppe(n)	796 Sport
Weiterführende Informationen	Inhaltsverzeichnis Inhaltstext

Frankfurt	Signatur: 2013 B 13195 Bereitstellung in Frankfurt
Leipzig	Signatur: 2013 B 15533 Bereitstellung in Leipzig

Kataloganzeige mit Links zu Inhaltstext und Inhaltsverzeichnis



Inhaltstexte und Inhaltsverzeichnisse: Bestandsentwicklung 2003 bis 2015

Hohe Klickraten

Dass dieses Angebot für die Nutzerinnen und Nutzer des Katalogs seither von großer Bedeutung ist, zeigen die noch immer wachsenden Zugriffszahlen deutlich (siehe Abbildung Seite 22 oben rechts). Dass die Klickraten so hoch sind, liegt vor allem auch daran, dass die Titeldaten der Deutschen Nationalbibliothek inklusive der Links auf die Inhaltstexte in eine Vielzahl von Verbund- und Bibliothekskatalogen eingebunden sind und dadurch nicht nur aus dem DNB-Katalog heraus aufgerufen werden können, sondern tatsächlich von Nutzerinnen und Nutzern tausender Bibliothekskataloge weltweit.

Anreicherung mit Inhaltsverzeichnissen

Nachdem urheberrechtliche Fragen im Kontext der Kataloganreicherung zwischen dem Deutschen Bibliotheksverband e. V., dem Börsenverein des Deutschen Buchhandels und der Deutschen Nationalbibliothek geklärt werden konnten,[2] startete die DNB im Februar 2008 ihren zweiten Kataloganreicherungsservice. Im Rahmen dieses Dienstes werden die Inhaltsverzeichnisse von Monografien und Zeitschriften-Stücktiteln systematisch und in Format und Design einheitlich digitalisiert und über den Katalog und die Datendienste der Deutschen Nationalbibliothek frei zugänglich gemacht.[3]

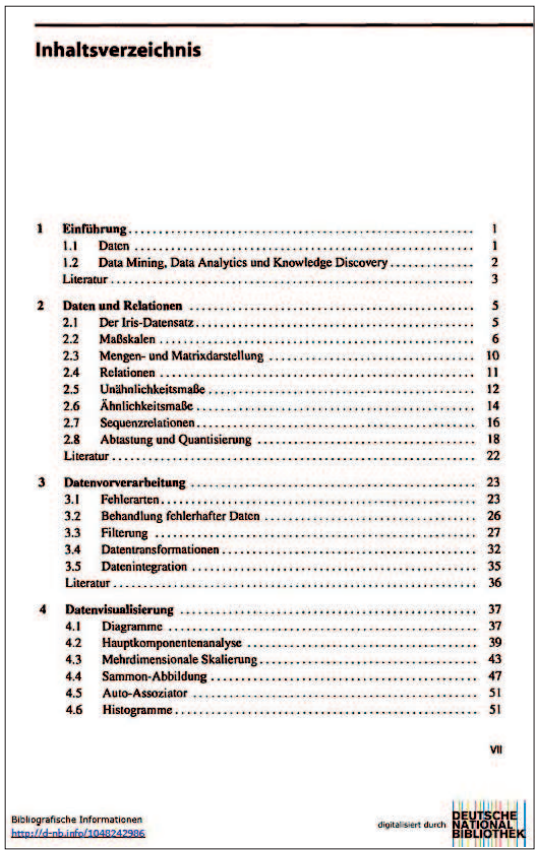
Sie sind neben den traditionellen bibliografischen Informationen und den Inhaltstexten die wohl wichtigste Quelle inhaltsbezogener Begriffe und damit für die Literatursuche von besonders hoher Relevanz.

Lag der Schwerpunkt dieses Anreicherungsservices zunächst noch bei der Digitalisierung der Inhaltsverzeichnisse von Publikationen des Verlagsbuchhandels (Reihe A der Deutschen Nationalbibliografie), wurden die Scan-Aktivitäten in den Folgejahren ausgeweitet und sukzessive auf weitere Bereiche bei den Neuerwerbungen wie auch retrospektiv auf bereits archivierte Bibliotheksbestände übertragen.

Beim Neuzugang wurden ab Juli 2011 auch diejenigen Publikationen, die außerhalb des Verlagsbuchhandels erscheinen (Reihe B), in den Scangeschäftsgang einbezogen; im August 2012 folgten die Dissertationen und Habilitationsschriften (Reihe H), im Januar 2013 die Publikationen des Auslandes einschließlich Germanica und Übersetzungen und ab Juli 2014 wurden dann auch die Inhaltsverzeichnisse von Musikdrucken (Reihe M) digitalisiert. Seither sind sämtliche Bücher sowie Zeitschriften-Stücktitel des Neuzugangs, die über ein Inhaltsverzeichnis verfügen, in den Anreicherungsservice einbezogen.

Seit 2014 vollständige Erfassung des Neuzugangs

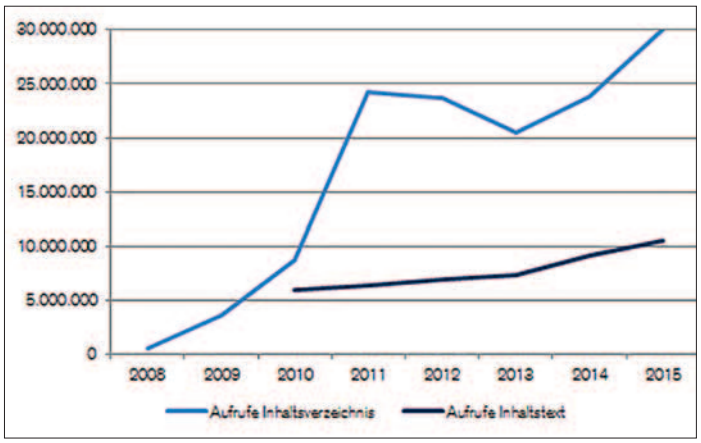
Retrospektiv wurden Anreicherungsprojekte von September 2008 bis Ende 2014 durchgeführt. Im Rahmen dieser Projekte wurden die monografischen Bestände der Zugangsjahre 1913 bis 1922 und 1983 bis 1990 am Standort Leipzig sowie der monografische Gesamtbestand der Exilsammlun-



Inhaltsverzeichnis

VII

Erste Seite eines Inhaltsverzeichnisses



Inhaltstexte und Inhaltsverzeichnisse: Nutzungsentwicklung 2008 bis 2015

gen in Frankfurt am Main und Leipzig bearbeitet. Konkret bedeutete dies, dass insgesamt 660.000 Bände systematisch am Regal gesichtet, hinsichtlich ihres Katalogstatus überprüft, gegebenenfalls katalogseitig nachbearbeitet, mit Barcodes ausgestattet und die darin enthaltenen 410.000 Inhaltsverzeichnisse digitalisiert und mit den Titeldaten verlinkt wurden.

Darüber hinaus gelang es in den Jahren 2010 und 2011, mehr als 310.000 digitalisierte Inhaltsverzeichnisse von Bibliotheksverbünden in Deutschland und Österreich auf der Basis bestehender Kooperationsbeziehungen zu übernehmen. Der eigene Bestand an digitalen Inhaltsverzeichnissen konnte dadurch erheblich ausgebaut und insbesondere für Monografien der Zugangsjahre 1985 bis 2007 wesentlich ergänzt werden.

Insgesamt verfügt die Deutsche Nationalbibliothek derzeit über mehr als 1,3 Millionen digitalisierte Inhaltsverzeichnisse mit Fokus auf die Jahre 1913 bis 1922 und 1983 bis heute.

Retrospektive Anreicherung und Datenübernahmen

Dass dieses Informationsangebot von Informationssuchenden stark nachgefragt wird, belegen die wachsenden Zugriffszahlen auch in diesem Fall (siehe Abbildung oben rechts). Ein Teil dieser sehr positiven Entwicklung ist sicherlich auch darauf zurückzuführen, dass die Inhaltsverzeichnisse zunehmend von großen Suchmaschinenbetreibern wie Google indexiert werden.

Was wurde bislang erreicht?

Die mit der Kataloganreicherung erreichten Ziele sind vielfältig[4] und gehen über den im engeren Sinne bibliografisch-bibliothekarischen Nutzen weit hinaus:

- Durch die Möglichkeit, Inhaltstext und Inhaltsverzeichnis unmittelbar bei der Recherche lesen zu können, wird die Beurteilung der Relevanz eines Titels für die Benutzerinnen und Benutzer erleichtert.
- In dem Maße, wie sich die Auswahlsicherheit bereits bei der Bestellung der Bücher aus den Magazinen der Bibliothek verbessert, können Fehlbestellungen reduziert werden.
- Durch die Einbindung der Anreicherungselemente in einen eigenen Suchindex wird die Qualität der Informationsrecherche wesentlich verbessert und die Wahrscheinlichkeit, relevante Treffer zu finden, entscheidend erhöht.
- Durch die retrospektive Digitalisierung von Inhaltsverzeichnissen wurden die Vorteile der Kata-

loganreicherung auch auf ältere Bestandsgruppen ausgeweitet.
- Außerdem konnte die Sichtbarkeit von Buchbeständen des 20. und 21. Jahrhunderts gerade im Web erhöht und damit die Möglichkeiten zur Entdeckung der in den Magazinen der Deutschen Nationalbibliothek lagernden, für Wissenschaft, Gesellschaft und Kultur relevanten Bestände verbessert werden.
- Durch die Integration der mit Zusatzinformationen angereicherten bibliografischen Daten der Deutschen Nationalbibliothek in eine Vielzahl von Verbund- und Bibliothekskatalogen werden die positiven Effekte verstärkt.

Letztlich steigert die Kataloganreicherung nicht nur den Wert des Kataloges, sondern vor allem auch den des Bestandes, dessen »verborgene Schätze« zum Teil überhaupt erst gefunden und hinsichtlich ihrer Relevanz in einem ersten Schritt und ohne die Bibliothek aufzusuchen beurteilt werden können. Dass dieser Mehrwert im Verhältnis zur intellektuellen Erschließungsleistung in Bibliotheken zu erheblich niedrigeren Kosten zu haben ist, macht diesen Dienst umso attraktiver.

Hebung verborgener Schätze

Perspektiven

Aufgrund der Verstetigung der Kataloganreicherung im Bereich des Neuzugangs ist in den kommenden Jahren mit einem Zuwachs von jährlich mehr als 100.000 Inhaltsverzeichnissen zu rechnen. Hinzu kommen weitere Übernahmen aus Bibliotheksverbünden in Höhe von bis zu 100.000 Inhaltsverzeichnissen, die für 2016 geplant sind.

Darüber hinaus wird auch die Wiederaufnahme der retrospektiven Kataloganreicherung angestrebt. Sie verbessert nicht nur die bereits beschriebenen Recherchebedingungen und die Möglichkeiten zum Auffinden benötigter Literatur in älteren Bestandssegmenten. Durch die in den Workflow integrierte Katalogbereinigung und Barcodeausstattung der Bücher werden darüber hinaus die Voraussetzungen dafür geschaffen, dass Maßnahmen wie die Massenentsäuerung oder die Massendigitalisierung von Büchern teilautomatisiert und damit wirtschaftlich und effizient durchgeführt werden können.

In der Zwischenzeit ist der Bestand an digitalisierten Inhaltsverzeichnissen auch für Text- und Data-Mining von Interesse: Er ist für maschinelle Analysen groß genug und wächst täglich weiter; er kann beliebig nachgenutzt werden, da keine urheber- oder sonstigen rechtlichen Einschränkungen bestehen; er besteht aus weitgehend einheitlichen und einfach strukturierten Dokumenten, die maschinell leicht weiterverarbeitet werden können; er bildet aufgrund des Sammelprofils der Deutschen Nationalbibliothek und ihrer auf Vollständigkeit zielenden Sammlung die nationale und auf Deutschland bezogene Buchproduktion umfassend ab. Und: Er wird vermutlich noch lange Zeit die wichtigste Quelle für maschinelle Analysen großer Teile der Buchproduktion Deutschlands im 20. und 21. Jahrhundert bleiben. Aufgrund urheberrechtlicher und finanzieller Grenzen ist in den nächsten Jahrzehnten nicht damit zu rechnen, dass das gedruckte Erbe Deutschlands dieser Zeit in entsprechend hoher Dichte und Vollständigkeit digital vorliegen wird.

Relevanz für Text- und Datamining

Trotz des Vorhandenseins eines nahezu idealen Datenbestandes für Text- und Data-Mining gibt es Wünsche: Zum einen mangelt es derzeit noch an Auswertungs- und Präsentationsinstrumenten, die speziell auf diesen kultur- und geistesgeschichtlich relevanten Datenbestand ausgerichtet sind. Dabei interessieren Programme, mit denen einerseits begriffliche Häufigkeitsverteilungen abgefragt und bislang unentdeckte semantische Beziehungsstrukturen aufgedeckt und andererseits die Ergebnisse für unterschiedliche Anwendungszwecke aufbereitet und visualisiert werden können. Hier sind Wissenschaft und Forschung gefragt und aufgerufen, das Potenzial des Datenbestandes für ihre Fragestellungen und Zwecke zu nutzen.

Zum anderen gilt es, bislang noch nicht bearbeitete Bestandsbereiche retrospektiv zu erschließen. Insbesondere bestehen derzeit noch Lücken bei den Monografien der Bestandssegmente von 1923 bis 1982 (rund 5 Millionen Werke) und bei den Hochschulschriften der Bestandssegmente von 1913 bis 1992 (rund 1 Million Werke). Diese gilt es sukzessive zu schließen. Damit würden nicht nur die vielfachen, bereits erwähnten Vorteile der Kataloganreicherung auch für diese Bestandsbereiche mo-

Desiderat retrospektive Anreicherung

bilisiert. Es würde darüber hinaus eine Datenquelle für Wissenschaft, Forschung und Kultur entstehen, die im Umfang und im nationalen Vergleich einmalig wäre, insbesondere für das im digitalen Schatten liegende 20. Jahrhundert.

Dass ein derartiges Großprojekt innerhalb von weniger als zehn Jahren zu verhältnismäßig geringen Kosten realisiert werden könnte, wissen wir. Allein, es fehlen derzeit die dafür benötigten finanziellen Mittel. Kooperationspartner und Sponsoren müssen deshalb gesucht werden, um die großen Schätze in den Magazinen der Deutschen Nationalbibliothek für Wissenschaft, Forschung und Kultur weiter zu heben.

Anmerkungen

1 Rädler, Karl: In Bibliothekskatalogen »googlen«: Integration von Inhaltsverzeichnissen, Volltexten und Web-Ressourcen in Bibliothekskataloge. In: Bibliotheksdienst 38 (2004) 7/8, S. 927–939.
Hinweise zur Geschichte der frühen Kataloganreicherung im Bibliothekswesen findet man auch bei Dietmar Haubfleisch und Irmgard Siebert: Catalogue Enrichment in Nordrhein-Westfalen – Geschichte, Ergebnisse, Perspektiven, in: Bibliotheksdienst, Jg. 42, 2008, H. 4, S. 384–391; ebenso Manfred Hauer und Reiner Diedrichs: Zwischenbilanz Collaborative Catalog Enrichment, in: Mitteilungen der VÖB 62 (2009) Nr. 3, S. 64–72 <http://fiz1.fh-potsdam.de/volltext/voeb/09540.pdf> und Matthias Groß: Kataloganreicherung – auf dem Weg zur kritischen Masse, in: Bibliotheksforum Bayern 01 (2007), S. 222–225.

2 Schreiben des Börsenvereins des Deutschen Buchhandels zur Kataloganreicherung vom 7. Juli 2007: <http://www.bibliotheksverband.de/fileadmin/user_upload/DBV/vereinbarungen/Boersenverein_110707_Kataloganreicherung.pdf>

3 Digitalisiert wird mit 300 dpi bitonal; bereitgestellt werden PDF-Dateien mit embedded text, die auf der ersten Seite ergänzt sind mit dem Logo der Deutschen Nationalbibliothek als digitalisierender Einrichtung und dem Link zum Datensatz der Deutschen Nationalbibliografie. Auch alle Inhaltsverzeichnisse in Frakturschrift enthalten Text und sind durchsuchbar.

4 Außer den hier aufgezeigten Vorteilen hat die Kataloganreicherung auch für andere Geschäftsbereiche der Deutschen Nationalbibliothek Relevanz: So werden die digitalisierten Inhaltsverzeichnisse bereits seit 2012 routinemäßig im Kontext der maschinellen Erschließung genutzt, um die lernbasierten Softwareverfahren zu trainieren. Außerdem werden im Geschäftsgang zur Kataloganreicherung auch Titelblätter von Hochschulschriften digitalisiert, die für die Automatisierung der (Formal-)Erschließung von Universitätsdissertationen benötigt werden (siehe Sandra Hamm und Kurt Schneider: Automatische Erschließung von Universitätsdissertationen, in: Dialog mit Bibliotheken, 2015, H. 1, S. 18–23).