

Und sie bewegt sich doch!

Der Einsatz eines Thesaurus zur Unterstützung der Sacherschließung in einem Discoverysystem

Friederike Borchert, Petra Keidel

In der elektronischen Kommunikationsumgebung von Bibliotheken existiert eine Vielfalt von Informationsressourcen in jeglicher Form, die hochaktuell und dank Discoverysystemen größtenteils inhaltlich vollständig durchsuchbar sind. Vor diesem informationstechnischen Hintergrund zeigt sich die erstaunliche Stärke von tradierten Erschließungsinstrumenten. Allerdings endet der Mehrwert dieser sacherschließenden Metadaten dort, wo bibliothekarische Daten aufhören. Die intellektuell basierte Verschlagwortung kann aufgrund der stetig anwachsenden Datenmenge aus unterschiedlichen Informationsräumen nicht mehr bewältigt werden. Für den Einsatz eines kontrollierten Vokabulars, welches semantische Kontextinformationen transportieren kann und automatisierte Verfahren dank entsprechender Datenmodellierung unterstützt, ist ein Thesaurus das Mittel der Wahl. Der Artikel gibt einen Überblick über den Mehrwert von Thesauri in Discoverysystemen, beschreibt den Vorgang der Thesauruserstellung mit dem Schwerpunkt der Auswahl von Vokabularmanagementsystemen im Linked-Data-Kontext und stellt die verschiedenen Ansätze der Integration eines Thesaurus in das vor Ort verwendete Discoverysystem WILBERT vor.

In the electronic communication environment of libraries there are a multitude of different information resources in various forms, which are up-to-date and whose contents can be fully searched, thanks to discovery systems. Against this background in information technology, the astonishing strengths of traditional classification systems become evident. Yet, the value of these indexing metadata terminates at the point where library-related data end. As the amount of data from different information spheres is growing continuously, intellect-based indexing cannot be managed any longer. For the use of a controlled vocabulary which can transport semantic context-information and which supports highly automatized processes thanks to related data modeling, a thesaurus is the tool of choice. The article provides an overview of the additional benefit of Thesauri in discovery systems, it describes the process of compilation of a thesaurus centered upon the choice of vocabulary management systems in a linked-data context and presents the different approaches to integration of a thesaurus in the discovery system WILBERT, which is used in this location.

Einleitung

Beinahe vergessen sind die Zeiten der Zettelkataloge, der Magazinbestellungen und der veralteten Bestände: Dank des Einsatzes von Discoverysystemen in der modernen elektronischen Kommunikationsumgebung von Bibliotheken sehen wir heute zufriedene Nutzer vor den Bildschirmen, die, gemessen an der geringen Nachfrage nach Errungenschaften der tradierten Bibliotheksarbeit, in der Fülle der hochaktuellen und schnell verfügbaren Informationen augenscheinlich immer etwas finden. Unzufriedenheit entsteht bei den Informationssuchenden allenfalls, weil beim Durchwühlen von mehrseitigen Trefferlisten Zeit verloren geht.

Dem professionellen bibliothekarischen Blick stellen sich aber die übergroßen Treffermengen als Discoveryerlebnis nicht nur als Zeitverlust dar. Hinter der durchaus erwünschten Angebotsvielfalt an Informationsressourcen in Bibliotheken verbirgt sich eine Heterogenität auf technischer, struktureller und semantischer Ebene: Schließlich wird in Bibliotheken mit unterschiedlichen Systemen gearbeitet und mit diversen Datenformaten umgegangen.

In dieser heterogenen Arbeitsumgebung von Bibliotheken verhält sich der Einsatz von Discoverysystemen genauso paradox wie der Käse zum Loch: Je komplexer die Umgebung für Informationen ist (mehr Käse), desto mehr Informationen stehen bereit (mehr Löcher). Je mehr Informationen bereit stehen, desto mehr schränkt die Heterogenität den Zugang zu Informationen ein (je mehr Löcher, desto weniger Käse). Also: Je ausgebauter die Umgebung von Informationen ist, desto mehr schränkt sie den Zugang zu Informationen ein (je mehr Käse, desto weniger Käse). Es entsteht ein Informationsverlust, der das zielgerichtete Auffinden von relevanten Inhalten erschwert.

Um diesem Umstand Abhilfe zu leisten, sollte das Potential von normierten und standardisierten Vokabularen nicht auch noch vergessen werden und es gilt, diese über die Datenpflege des eigenen Bestandes hinausgehend mit größtmöglicher Wirkung einzusetzen.

Der Einsatz von Erschließungsvokabularen in der Umgebung von Discoverysystemen

Vor dem Hintergrund dieser immer komplexer werden den Heterogenität im bibliothekarischen Arbeitsumfeld tritt die Stärke von tradierten Erschließungsinstrumenten wie z.B. Normdateien (GND, inkl. SWD), und Klassifikationen (RVK, DDC) deutlich in den Vordergrund. Auch wenn sie als standardisierte und normierte Vokabulare nicht mehr in Form eines klassischen Schlagwortkatalogs angeboten werden bzw. als Aufstellungssystematik die Suche am Regal als primären Such Einstieg unterstützen, so unterstützen sie nun beim Einsatz von Discoverysystemen vor allem den iterativen Suchprozess. Sie ermöglichen, den Suchvorgang entsprechend des Informationsbedarfs schrittweise neu zu justieren. Dafür sind in Discoverysystemen die sacherschließenden Elemente als Einträge in Facetten abgebildet oder gemeinsam mit thematisch verwandten Wörtern in einer Wolke aufbereitet.¹

Nicht zuletzt aufgrund der beinahe flächendeckenden Vergabepraxis von Schlagwörtern und Notationen, die fest in die Geschäftsgänge von Bibliotheken integriert und mit fachspezifischer Expertise umgesetzt wird, besitzen diese sacherschließenden Elemente eine für die Recherche zuverlässige Größe. Sie gelten in der elektronischen Umgebung von Informationen weiterhin als qualitätssteigernde Inhaltskomprimata mit hoher Informationsdichte.

Der Mehrwert dieser beigefügten sacherschließenden Metadaten endet allerdings dort, wo bibliothekarisch erarbeitete Daten aufhören. Die intellektuell basierte Verschlagwortung kann im Zuge der genannten Problematik, die Vielzahl der Informationsquellen, nicht mehr geleistet werden. Aufgrund der stetig anwachsenden Datenmenge aus diversen Informationsräumen wird diese Praxis der Sacherschließung zunehmend durch maschinelle Verfahren ersetzt, die bestenfalls auf gut ausbalanciertem Textmining beruhen.² Durch das Abgleichen von Buchstabenfolgen werden dann bei der Facettenbildung Inhalte geclustert, deren Gemeinsamkeit vor allem auf einer morphologischen Ebene begründet ist. Die Folge ist, dass ein zunehmend größer werdender Anteil an Inhalten in Discoverysystemen lediglich durch die semantische Referenz des Wortes im Verständnis eines Lemmas präsentiert wird.

Es bedarf also eines Vokabulars, welches befähigt ist, in einer elektronischen Informationsumgebung über die Buchstabenfolge des Wortes hinausgehend auch semantische Kontextinformationen zu transportieren.

Die Wiederentdeckung des Thesaurus

Spätestens in den siebziger Jahren, als die Bundesregierung mit dem ersten IuD-Förderprogramm³ auf eine wachsende Informationsproduktion reagierte, entstanden zahlreiche fachlich differenzierte und hochkomplexe Thesauri.⁴

Diese bis dahin wenig beachtete Dokumentationspraxis erlebte so ihren Aufschwung zeitgleich mit dem Einzug der elektronischen Datenverarbeitung in die Informationspraxis und im Zuge der allgemein einsetzenden Normierungsprozesse im Umgang mit Informationen.⁵ Aber nicht zuletzt die Fähigkeit eines Thesaurus, den Inhalt von Dokumenten in natürlicher Sprache in eine Systemsprache zu übersetzen, die die Mittelbarkeit zwischen dem Autor, dem Indexierer (bzw. Fachreferenten) und dem Benutzer aufgrund der Verwendung desselben fachspezifischen und kontrollierten Vokabulars herstellt, begründete den praktischen Einsatz dieser mächtigen Erschließungsinstrumente in Zeiten eines Informationsüberflusses. Für den Aufbau eines Thesaurus wird mittels verschiedener Prozesse der Wortschatz eines Fachgebietes in ein normiertes Vokabular überführt, eine ein-eindeutige Beziehung zwischen Begriff und Benennung hergestellt und die Relationen zwischen den Benennungen nachgewiesen.

Der Thesaurus kennt hierbei drei Arten von Beziehungen. Die hierarchische Relation zwischen Benennungen beruht auf der klassifikatorischen Über- bzw. Unterordnung von Begriffsmerkmalen (z.B. der Teil-Ganzes-Beziehung des Gegenstandes (Tischbein - Tisch). Die Äquivalenzrelation greift die Unschärfe der natürlichen Sprache auf und führt Benennungen zusammen, die, obwohl sie Begriffe mit gleichen Merkmalen beinhalten, auf der Wortebene auseinander gefallen sind (z.B. (quasi-)synonyme Benennungen Hocker - Schemel). Die assoziative Relation zwischen Benennungen bildet schließlich eine Verwandtschaft zwischen Benennungen ab, die innerhalb des betrachteten Wortguts weder hierarchisch noch äquivalent

1 Das Angebot besteht bei der Anzeige der Treffermenge im E-LIB-Portal der SUB Bremen (<http://elib.suub.uni-bremen.de/>), das über das Frontend des Discoverysystems nicht nur ausschließlich Schlagwörter als sogenannte „Teilaspekte des Themas“ zusammenfasst.

2 Für die automatische Indexierung von Volltexten werden bereits auch erweiterte Verfahren zur Erkennung von Mehrwortgruppen eingesetzt, die gute Ergebnisse liefern (Bredack/Lepsky 2014). Eine Bewertung des semantischen Gehalts des gewonnenen Wortmaterials wird jedoch nicht von diesen Verfahren abgedeckt.

3 Programm der Bundesregierung zur Förderung der Information und Dokumentation: (IuD-Programm); 1974 - 1977. Bonn, 1976.

4 Begründet auf einen Beschluss der EG von 1971 zur Förderung der wissenschaftlichen Zusammenarbeit, weist der 1985 erschienene Thesaurus Guide, nach einer dreijährigen Erarbeitungsphase, bereits 462 Einzelthesauri nach (Thesaurus Guide, Vorwort).

5 Die Erarbeitung und die Verbindlichkeit von Regelwerken im nationalen und internationalen Bibliothekswesen erfolgte ebenfalls in diesem Kontext.

begründet ist (z.B. Abfall - Abfallrecht). Diese Beziehungen werden innerhalb des ausgewählten Wortguts eines Faches den Benennungen mitgegeben, indem sie in normierter Schreibweise ausgewiesen werden (BROADER TERM, USE FOR/USED-FOR etc.).

Aufgrund der verbindlichen Festlegung, welche Benennung aus diesem Bedeutungsgeflecht zur Verschlagwortung verwendet werden darf, transportieren dann diese sogenannten Deskriptoren den komplexen semantischen Gehalt des Beziehungsgeflechts gemeinsam mit der Buchstabenfolge des Wortes.⁶

In dieser Ausgestaltung können Thesauri in einer heterogenen Metadatenumgebung die Gemeinsamkeit von Inhalten auf der Basis einer Semantik clustern, die auf einer logisch einwandfreien Modellierung von sprachlicher Bedeutung beruht.⁷

Für einige Fachgebiete existieren bereits etablierte, nachnutzbare Thesauri. Dazu gehören u.a. der Standardthesaurus Wirtschaft (STW), der Thesaurus Sozialwissenschaften (TheSoz) oder der Umweltthesaurus (UMTHES) und GEneral Multilingual Environmental Thesaurus (GEMET) für die Bereiche Umwelt und Ökologie.

Der frühen Ära der Entstehung zahlreicher Thesauri ist es aber geschuldet, dass viele Vokabulare, gerade mit sehr speziellen Ausrichtungen, oftmals nur in Printform vorliegen.⁸ Sollten sie tatsächlich in elektronischer Form erfasst sein, besitzen sie meistens eine unmodellerte Datenstrukturierung, die eine problemlose Integration in automatisierte Vorgänge verhindert.⁹

Es bedarf also einer Softwareumgebung, die es auf einfachem Wege ermöglicht, die teils bereits existierenden Wortschätze, teils aber auch neu entstehende Thesauri, als standardisierte Vokabulare in einer nachnutzbaren Weise zu modellieren, um sie für die Integration in die elektronische Kommunikationsumgebung (z.B. in der Umgebung von Discoverysystemen) vorzubereiten.¹⁰

6 Ein fachlich spezifizierter Thesaurus enthält ca. 3.000 Einträge, die für die Sacherschließung verwendet werden dürfen, weitere 10.000 Einträge dienen als Zugangsvokabular.

7 Vgl. Keller/Schneider/Volk 2014, S. 16.

8 Aber hier besteht gerade an Fachhochschulen mit ihrem Kanon an eher kleinen Studiengängen mit spezifischen Ausrichtungen (z.B. Telematik) das dringende Desiderat, über das fachlich differenzierte Vokabular zu verfügen, da die bereits nachnutzbaren Thesauri in Bezug zur Fachspezifikation nicht die gewünschte Tiefe besitzen.

9 Auch wenn diese Thesauri in der gängigen Erschließungspraxis z.B. zur Verschlagwortung vor Ort verwendet werden, so ist ihre Existenz außerhalb der lokalen Informationsumgebung häufig nicht bekannt. Einen ersten gegenwärtigen Ansatz zum Verzeichnen und Nachweisen von Thesauri bietet der Register of Thesauri, Ontologies & Classifications (<http://www.bartoc.org/de>).

10 Neben der Notwendigkeit, Thesauri inhaltlich weiterzupflegen, um sie der aktuellen Fachsprache stetig anzupassen, wird bereits

Das Management von Thesauri

An der Hochschulbibliothek der Technischen Hochschule Wildau wird im Rahmen eines mehrjährigen Projekts ein Workflow konzipiert und eine auf Open-Source-Komponenten beruhenden Umgebung geschaffen, der den Aufbau neuer und die Nachnutzung vorhandener Thesauri ermöglicht. In der ersten Phase wurden in einer Teststellung bereits standardisierte, onlinefähige und bilinguale Mikro-Thesauri für die Fachgebiete Logistik, Informatik, Wirtschaft und Verkehr aufgebaut.

Für die Erstellung dieses Fachvokabulars wurden Begriffe aus verfügbaren Terminologien, Klassifikationen, Glossaren oder Inhaltsverzeichnissen relevanter Fachbücher extrahiert und bearbeitet. Grundlage für die Bearbeitung des gesammelten Wortgutes bildeten einschlägige Normen und Standards wie ISO 25964-1, ISO 2788, sowie DIN 1463-1.

Um die Nachnutzbarkeit des Thesaurus und die Interoperabilität mit anderen Informationssystemen zu gewährleisten, wird für die Modellierung der W3C-Standard SKOS (Simple Knowledge Organization System) verwendet. Mit diesem Modell lassen sich kontrollierte Vokabulare wie Thesauri oder Taxonomien als maschineninterpretierbare Informationen von Konzepten und ihrem Umfeld abbilden. Dazu gehören lexikalische Informationen (Konzeptbenennungen und deren alternative Schreibweisen), Notationen, dokumentarische Informationen (Definitionen, Beispiele, Veränderungsanmerkungen etc.), semantische Beziehungen innerhalb eines kontrollierten Vokabulars (Ober-, Unter- und verwandte Begriffe) sowie das Abbilden von Crosskonkordanzen zu externen Vokabularen.

Das Erstellen, Bearbeiten und Publizieren von Konzepten erfordert die Auswahl eines geeigneten Vokabularmanagementsystems. Das Angebot an SKOS-fähigen Vokabularmanagementsystemen ist überschaubar. Neben wenigen modernen Open-Source-Lösungen haben sich auch kommerzielle Produkte wie PoolParty¹¹ auf dem Markt etabliert.

Für das Thesaurus-Projekt der Hochschulbibliothek wurden eine Reihe projekt- und toolspezifischer Auswahlkriterien aufgestellt. Projektspezifische Kriterien betrafen die Open-Source-Verfügbarkeit, eine aktive Nutzer-Community, die Unterstützung von Mehrsprachigkeit (deutsch/englisch) und ein hohes Maß an Anwenderfreundlichkeit. Konkrete toolspezifische Anforderungen basieren auf Morshed & Duttas er-

1985 von Gernot Wersig die Änderung im Informationssystem neben fünf weiteren Gründen genannt, um die Fortschreibung von Thesauri zu forcieren (Wersig 1985, S. 274 f.).

11 <http://www.poolparty.biz/> [05.09.2014]

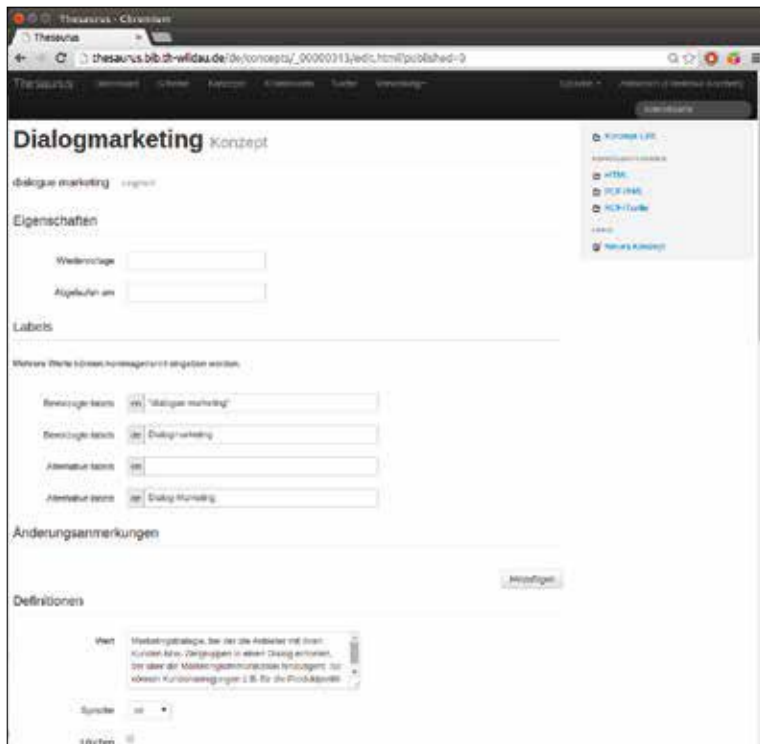


Abbildung 1:
Bearbeitungs-
modus von
iQvoc

stellten Auswahlkriterien für Vokabularmanagementsysteme im Linked-Data-Kontext.¹²

Neben Kriterien, die sich auf das Erstellen von kontrollierten Vokabularen konzentrieren (Anlegen und Löschen von Konzepten, Erstellen von hierarchischen und verwandten Beziehungen zwischen Konzepten und die Sichtbarkeit dieser Beziehungen sowie das Rückgängigmachen und Wiederherstellen von Aktionen), wird das Ausmaß, in dem ein Tool die Konsistenz der erstellten Konzepte überprüft, untersucht. Auch die Komplexität eines Tools wird betrachtet. Darunter fallen die intuitive Bedienbarkeit, die Verfügbarkeit einer benutzerfreundlichen Oberfläche und die Einfachheit der Installation. Für den Linked-Data-Kontext wird überprüft, inwieweit die Unterstützung von RDF bzw. des SKOS-Standards, die Abfrage der Daten mit SPARQL, die Verwendung von Content-Negotiation-Mechanismen, um über eine URI unterschiedliche Repräsentationen eines Konzepts auszuliefern oder die Nachnutzbarkeit externer Datenquellen unterstützt wird. Weitere Kriterien sind die Verfügbarkeit bezogen auf Kosten und Onlinefähigkeit, Browsing- und Abfragefunktionalitäten, Konfigurationsmöglichkeiten, Mehrsprachigkeit, Nutzerauthentifizierung und die Möglichkeit für nutzerspezifische Anpassungen des Tools.

Die optimalsten Ergebnisse wurden während einer

12 Vgl. Morshed, Ahsan/ Dutta, Ritaban: Machine Learning based Vocabulary Management Tool Assessment for the Linked Open Data, in: International Journal of Computer Applications 60(9) (2012), S. 51-58.

kurzen Testphase mit iQvoc¹³ erzielt, einem webbasierten Open-Source-Tool, das von der Firma innoQ Deutschland GmbH bereitgestellt und weiterentwickelt wird. Die Ruby-on-Rails-Anwendung für das Erstellen mehrsprachiger, kontrollierter Vokabulare unterstützt die wesentlichen Elemente des SKOS-Modells. Modelliert werden können neben Kollektionen für das thematische Gruppieren von Konzepten auch lexikalische und dokumentarische Informationen (Abb. 1) sowie semantische Beziehungen innerhalb eines Thesaurus sowie zu externen Vokabularen über die SKOS-Mapping-Eigenschaften `exactMatch`, `closeMatch`, `relatedMatch`, `broadMatch` und `narrowMatch`. Ein Consistency Checker überprüft grundlegende Fehler bei der Erstellung des Vokabulars. So ist es beispielsweise nicht möglich, die Vorzugsbenennung eines Begriffs in einer Sprache für mehrere Konzepte zu vergeben. Mit der von iQvoc bereitgestellten Linked-Data-Infrastruktur wird jedes Konzept durch eine URI repräsentiert, die ein HTTP Redirect nach RDF/XML, Turtle (Terse RDF Triple Language) und HTML ermöglicht. Für SPARQL-Abfragen kann ein externer Triple Store angebunden werden. Das erstellte und publizierte Vokabular ist nach der Anmeldung am System als Administrator vollständig als RDF in unterschiedlichen Serialisierungen exportierbar. Bei dem von iQvoc verfolgten „Cool-URLs-don't-Change“-Ansatz¹⁴ ist das Löschen bereits publizierter Konzepte unzulässig.¹⁵ Über die Nutzerverwaltung werden Rollen mit unterschiedlichem Rechtemumfang zugeteilt. Gäste haben lesenden Zugriff auf das Dashboard, in dem angelegte, aber noch nicht publizierte Konzepte verzeichnet sind. Editoren können neue Konzepte anlegen und bestehende bearbeiten, diese jedoch nicht publizieren. Diese Aufgabe wird von sog. Publishers übernommen. Administratoren können darüber hinaus eine elementare Systemkonfiguration vornehmen und übernehmen die Rechte- und Nutzerverwaltung. Die Darstellung publizierter Konzepte erfolgt in einer hierarchischen Baumstruktur bzw. als alphabetische Auflistung.

Die Integration des mit iQvoc erstellten Thesaurus in ein Discoverysystem wird gegenwärtig in der lokal verwendeten Library Search Engine (LSE) erprobt.

13 <http://iqvoc.net/> [05.09.2014]

14 Vgl. Berners-Lee, Tim: Cool URIs don't change (1998), <http://www.w3.org/Provider/Style/URI.html> [26.08.2014].

15 Vgl. Bandholtz, Thomas/ Schulte-Coerne, Till/ Glaser, Robert/ Fock, Joachim/ Keller, Tim: iQvoc - Open Source SKOS(XL) Maintenance and Publishing Tool, in: Proceedings of Scripting and Development for the Semantic Web Workshop at the ESWC, Heraklion, Greece, May 31, 2010, CEUR Workshop Proceedings (2010), <http://ceur-ws.org/Vol-699/Paper2.pdf>, S. 6 [05.09.2014].

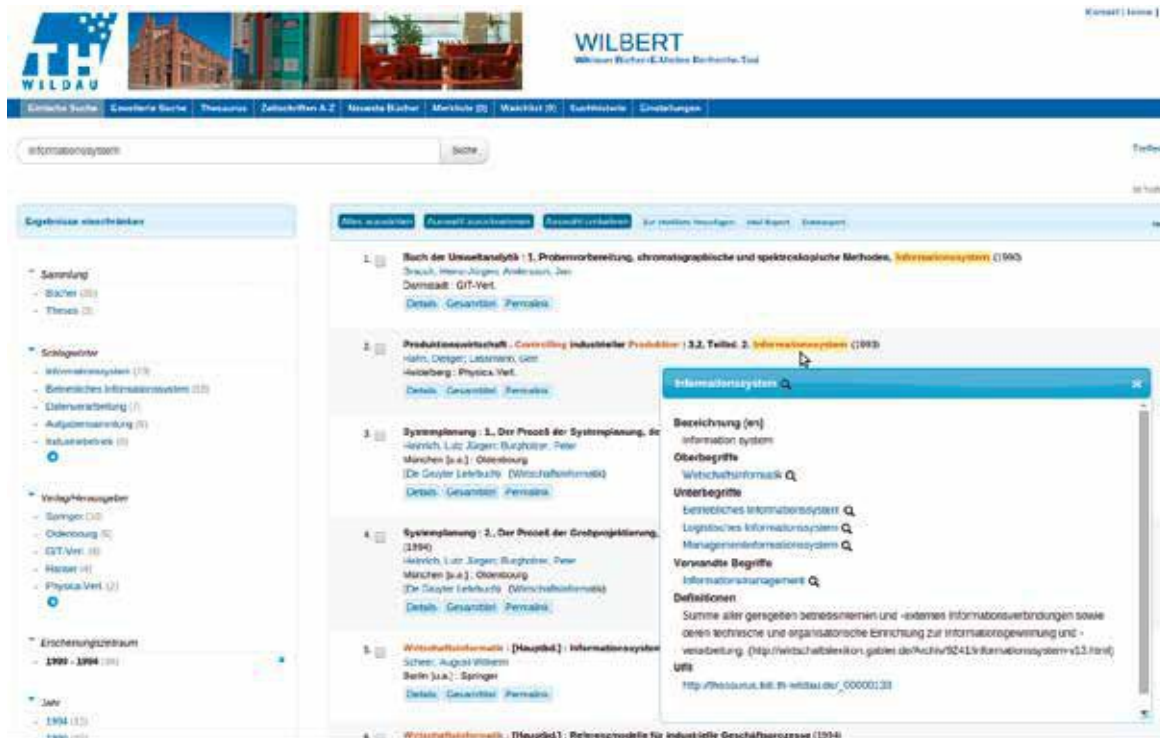


Abbildung 2:
Hervorhebung
von Thesaurus-
einträgen in
WILBERTs
Trefferliste

WILBERT kam, sah und siegte

An der TH Wildau wird die LSE WILBERT (Wildauer Buch + E-Medien Recherchetool) eingesetzt. Sie bietet den Nutzern einen im Internetzeitalter beinahe schon klassisch zu nennenden Sucheinstieg bzgl. der gedruckten bzw. elektronisch verfügbaren Beständen der Hochschulbibliothek: Mittels der Eingabe von Suchbegriffen in einem Suchschlitz durchsucht WILBERT parallel zu dem Datenbestand des lokalen Bibliothekskatalogs auch weitere, speziell ausgewählte Informationsquellen (fachspezifische Dokumentenserver, Zeitschriftenverzeichnisse, Table-of-Contents-Dienste etc.). Aus über sechs Millionen indexierten und täglich aktualisierten Datensätzen wird anschließend eine Trefferliste generiert, welche mithilfe von sechs angebotenen Facetten iterativ eingeschränkt werden kann.

Die LSE-Technologie von WILBERT basiert auf verschiedenen Open-Source-Komponenten.¹⁶ Die vom KOBV modifizierte Suchmaschinentechologie, ALBERT genannt, besteht als Hosting-Angebot im SaaS-Modell. Während der KOBV für alle technischen Belange zuständig ist, u.a. für die Bereitstellung der IT-Infrastruktur, Wartung der Software, Indexierung der Daten und Weiterentwicklungen, behält die Bibliothek die Hoheit über ihre Daten. Sie ist verantwortlich für deren Qualität, Pflege und Bereitstellung sowie für die

Auswahl externer Informationsräume, die indexiert werden sollen.

Während die ALBERT-Technologie bereits seit 2007 bzw. 2012 auf zwei Instanzen im Forschungskontext erfolgreich im Einsatz ist¹⁷, ist WILBERT in dieser „ALBERT-Familie“ die erste Instanz deutschlandweit, die im Hochschulbereich und für diese Umgebung spezifiziert, 2013 in den Praxiseinsatz überführt worden ist.

Die unterschiedlichen Anforderungen eines heterogenen Nutzerkreises im Hochschulkontext (Studierende, Lehrpersonal, externe Besucher) prädestiniert die Wildauer Instanz, die Integration eines Thesaurus in ein Discoverysystem exemplarisch zu erproben.

Die Integration des Thesaurus in WILBERT

Für die Thesaurusintegration in WILBERT wurden zwei verschiedene Ansätze entwickelt und umgesetzt.

Der erste Ansatz verfolgt die Suche nach Thesaurusbegriffen als separate, in WILBERT integrierte Suche. Für die Umsetzung wurde in WILBERT der neue Navigationspunkt „Thesaurus“ geschaffen, über den nach Deskriptoren und Nichtdeskriptoren gesucht werden kann. Suchtreffer sind hier also keine bibliographischen Titeldaten, sondern Einträge aus dem Thesaurus. Trifft die Suchanfrage auf eine alternative

¹⁶ Als Retrievaltechnologie wird die Software Solr/Lucene eingesetzt. Sie wurde von der Zentrale des Bibliotheksverbundes KOBV modifiziert und an bibliothekarische Bedürfnisse angepasst (vgl. Szott/Bertelmann 2012).

¹⁷ Deutsches Geoforschungszentrum Potsdam (GFZ), Deutsche Krebsforschungszentrum Heidelberg (DKFZ), die ALBERT-Instanz Ilse des Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik Kiel (IPN) ging am 28.8.2014 in Produktion; eine weitere Instanz startet in Kürze (GeoMar Kiel).

Schreibweise, wird bereits in der autocomplete-Anzeige automatisch auf den Deskriptor verwiesen.

Die Auswahl eines einzelnen Deskriptors bei den autocomplete-Vorschlägen führt direkt zu dessen Kurzansicht mit der Vorzugsbenennung in beiden Sprachen. Wird kein einzelner Deskriptor ausgewählt, wird eine Trefferliste erzeugt. In der Detailansicht eines Treffers werden zusätzlich Nichtdeskriptoren, Beispiele, Definitionen, hierarchische und verwandte Beziehungen sowie die URI zur Anzeige gebracht. Mit Ober-, Unter- und verwandten Begriffen kann eine Suche in WILBERTs Datenbestand ausgelöst werden. Beim zweiten Ansatz gibt der Nutzer einen beliebigen Suchbegriff über WILBERTs einfache oder erweiterte Suche ein. In der Trefferliste werden jene Begriffe farblich mit Hypertextfunktionalität hervorgehoben, die auch als Deskriptor oder Nichtdeskriptor im Thesaurus vorkommen (Abb. 2). Beim Klick auf einen dieser Begriffe öffnet sich ein Modalfenster mit weiteren Informationen zum Begriff: je nach Verfügbarkeit werden Nichtdeskriptoren, Ober-, Unter- und verwandte Begriffe sowie Beispiele und Definitionen angezeigt.

Wie auch in der Suche nach Thesaurusbegriffen kann mit Ober-, Unter- und verwandten Begriffen aus dem Modalfenster heraus eine neue Suche ausgelöst werden, um die vorherige Suche einzugrenzen oder auszuweiten.

Ausblick

Ein weiteres kurzfristiges Ziel im Thesaurus-Projekt ist die Ausweitung des Matchings beim oben erwähnten zweiten Ansatz. Zum jetzigen Zeitpunkt werden ausschließlich exakt übereinstimmende Begriffe in der Applikation hervorgehoben. Um auch andere grammatikalische Formen der Deskriptoren einzu beziehen, wird in der nächsten Thesaurus-Version zusätzlich das Element `skos:hiddenLabel` verwendet. `skos:hiddenLabel` erlaubt das Verzeichnen lexikalischer Informationen, die in Informationssystemen für die textbasierte Suche zum Einsatz kommen, ohne jedoch für den Nutzer sichtbar zu sein. In der Praxis wird diese `skos`-Eigenschaft verwendet, um Begriffe mit gängigen Rechtschreibfehlern bei der Eingabe zu berücksichtigen. Im Thesaurus-Projekt werden über `skos:hiddenLabel` weitere grammatikalische Formen der Deskriptoren und Nichtdeskriptoren hinterlegt. Mittelfristig ist eine Automatisierung dieses Prozesses mit entsprechenden Textmining-Verfahren geplant.

Als mittelfristiges Projektziel wird ein zusätzlicher Einstieg für die Thesaurussuche realisiert. Parallel zur textbasierten Suche nach Deskriptoren wird das

Browsen durch die hierarchische Baumstruktur des Thesaurus angeboten. Insbesondere Studierende in den ersten Semestern wird auf diesem Wege ein intuitiver Überblick über die Struktur ihres Studienfachs gegeben.

Während diese Weiterentwicklungen der Applikationen aus den ersten Erfahrungen der gegenwärtigen Testphase resultieren, ist zukünftig auch eine systematische Qualitätsauswertung der Thesaurus-unterstützten Recherche vorgesehen. Vorrangig wird hier die Expertise unserer Hochschulöffentlichkeit als Bewertungsmaßstab berücksichtigt und, sobald eine kritische Größe des Thesaurus erreicht ist, in einem nach Nutzergruppen spezifizierten Rahmen abgefragt werden hinsichtlich der Benutzerfreundlichkeit, der Steigerung der Recherchequalität und des gewünschten inhaltlichen Ausbaus des Thesaurus.

Längerfristig wird auf dieser Umfragebasis die Ausgestaltung der Navigationsmöglichkeiten innerhalb des Thesaurus und die Anbindung an die Oberflächengestaltung des Discoverysystems weiterentwickelt werden.

Aus bibliothekarischer Sicht ist die Einschätzung der Recherchequalität in Bezug zur bisherigen Sacherschließung von großem Interesse. Beim Vergleich von unterschiedlich erzeugten Treffermengen (z.B. mittels freier Ein-Wort-Suche, Schlagwortsuche, Deskriptorensuche, klassifikatorischer Einstieg) entstehen folgende Fragestellungen: Ist die Größe der Treffermenge mittels der Deskriptorensuche angemessen? Ist die Qualität der Treffer in Bezug zum Informationsbedarf angemessen? Wie groß ist die Kongruenz unterschiedlich erzeugter Treffermengen? An welchen Stellen ist die lexikalische Nähe zwischen den Deskriptoren und den anderen sacherschließenden Vokabularen zu groß und sollte differenziert bzw. ineinander überführt werden? In welcher Form kann eine Überführung technisch realisiert werden?

Unter der Berücksichtigung verschiedener Faktoren, die auf die Trefferzusammenstellung Einfluss nehmen, z.B. die Größe des Datenbestandes in den jeweiligen Fachgebieten und die Komplexität des Thesaurus in den jeweiligen Fachdisziplinen, erwarten wir von der bereits zeitlich in die Nähe gerückten Bewertungsphase kurzfristig eine praktische Richtungsanweisung hinsichtlich der weiteren Granulierung des Wortschatzes an bestimmten Knotenpunkten.

Auf der Basis der Trefferanalyse erwarten wir aber ebenso die Erarbeitung eines sogenannten Kandidatenvokabulars, welches in der nächsten Wortgutsitzung als Vorschlag der Fachgemeinschaft vorgelegt werden kann, um den Thesaurus inhaltlich und fachbezogen weiterzuentwickeln.

Schließlich besitzen Thesauri ein sehr großes sacherschließendes Potential, welches in verschiedene wissensorganisatorische Szenarien eingesetzt werden kann. Dank eines geeigneten Vokabularmanagementsystems in SKOS-modellierter Thesaurus kann als Linked Data nachgenutzt und problemlos in andere Informationsräume integriert werden.

In einigen Fachdisziplinen ist ein Thesaurus aber auch häufig das erste fachspezifische, kontrollierte Vokabular überhaupt, welches den Wissensstand des Faches abbildet und als normierende Diskussionsgrundlage für die Fachgemeinschaft verwendet werden kann.

So wird zukünftig aus dem vorliegenden Thesaurus der Bereich Logistik in Kooperation mit dem in Poznan ansässigen Instytut Logistyki i Magazynowania¹⁸ um das Themenfeld RFID ausgebaut und vor Ort in der Lehre eingesetzt werden. ■

Literatur

Bandholtz, Thomas/ Schulte-Coerne, Till/ Glaser, Robert/ Fock, Joachim/ Keller, Tim: iQvoc – Open Source SKOS(XL) Maintenance and Publishing Tool, in: Proceedings of Scripting and Development for the Semantic Web Workshop at the ESWC, Heraklion, Greece, May 31, 2010, CEUR Workshop Proceedings (2010), <http://ceur-ws.org/Vol-699/Paper2.pdf> [26.08.2014].

Berners-Lee, Tim: Cool URIs don't change (1998), <http://www.w3.org/Provider/Style/URI.html> [26.08.2014].

Bertelmann, Roland / Szott, Sascha / Höhnnow, Tobias: Discovery jenseits von "all you can eat" und "one size fits all", in: Bibliothek Forschung und Praxis 3(2012). S. 369-376.

Bredack, Juliane/ Lepsky, Klaus: Automatische Extraktion von Fachterminologie aus Volltexten. In: ABI-Technik 34(1) (2014), S. 2-8.

Keller, Stefan Andreas (Hg.): Wissensorganisation und -repräsentation mit digitalen Technologien, Berlin 2014. (Bibliotheks- und Informationspraxis, 55).

Morshed, Ahsan/ Dutta, Ritaban: Machine Learning based Vocabulary Management Tool Assessment for the Linked Open Data, in: International Journal of Computer Applications 60(9) (2012), S. 51-58.

Thesaurus Guide: analytical directory of selected vocabularies for information retrieval, Amsterdam 1985.

Wersig, Gernot: Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis, 2., erg. Aufl., München 1985.

¹⁸ <http://www.ilim.poznan.pl/> [05.09.2014]



Petra Keidel

Mag. Art, Management
elektronische Ressourcen an der
Technische Hochschule Wildau
keidel@th-wildau.de



Friederike Borchert

Projektmitarbeiterin an der
Technische Hochschule Wildau
fborchert@th-wildau.de

... WERTE BEWAHREN

PROFI-BUCHSCHUTZ- FOLIEN VON NESCHEN



- erhöhen die Lebensdauer von Büchern
- schützen vor Verschmutzung
- zur besseren Repositionierung auch mit reduzierter Anfangslebkraft erhältlich
- lösungsmittelfreie alterungsbeständige Acrylatkleber

Fragen? Sie erreichen uns unter Telefon 05722 207-0 oder Sie senden uns eine E-Mail an info@neschen.de.

NESCHEN AG

Hans-Neschen-Straße 1
31675 Bückeburg
T +49 5722 207-0
E info@neschen.de

**SOFORT
GRATISMUSTER
BESTELLEN!**



NESCHEN

WWW.NESCHEN.COM