

Die Bereinigung der Systematik an der Bibliothek der Universität Konstanz

Anna Kasprzik

Abstract

Die in ihren Grundzügen hierarchisch angelegte Aufstellungssystematik der Bibliothek der Universität Konstanz ist im Laufe der Zeit durch zahlreiche nicht-hierarchische Strukturen ergänzt worden, so dass ein heterogenes Gebilde entstanden ist. Außerdem kam der Wunsch nach Austauschbarkeit der Konstanzer Klassifikation mit anderen bibliothekarischen Einrichtungen auf. Aus diesen Gründen sollte die Systematik nun bereinigt werden. Dazu wird als zentrales Werkzeug ein Diagnosealgorithmus benötigt, welcher sichtbar macht, wie weit die aktuelle Systematik von einem vorher zu definierenden, möglichst baumförmigen Idealzustand entfernt ist. Im Rahmen eines Referendariats ist in diesem Projekt ein solcher Algorithmus entwickelt und implementiert worden.

Ausgangssituation und Vorhaben

Eine der ersten öffentlichen Beschreibungen der Aufstellungssystematik der Bibliothek der Universität Konstanz, also der "Regeln für den Aufbau von Buchsignaturen", findet sich 1969 im Sonderheft 1 der im selben Jahr erstmals erschienenen "Bibliothek aktuell"¹, drei Jahre nach Gründung der Universität. Die Systematik ist ein Eigenprodukt und ein Unikat und als solches ein wertvolles Stück Bibliotheksgeschichte.

Vier Jahrzehnte später stellt sich die Ausgangssituation für das vorliegende Projekt folgendermaßen dar: Die im Ansatz hierarchisch angelegten Strukturen der Aufstellungssystematik der Bibliothek sind im Laufe der Zeit durch zahlreiche nicht-hierarchische Elemente ergänzt worden, so dass ein heterogenes Gebilde entstanden ist. Außerdem besteht der Wunsch, die Bibliothek der Universität Konstanz langfristig aus der Isolation einer individuellen Haussystematik herauszuführen. Ziel ist es, verbale oder klassifikatorische Sacherschließungsarbeit anderer Bibliotheken in Konstanz nutzen zu können und umgekehrt andere von der Konstanzer Arbeit auf diesem Gebiet profitieren zu lassen. Zur Vorbereitung auf die Kompatibilität der Konstanzer Systematik zu

anderen Sacherschließungen soll die Systematik nun bereinigt werden. Als zentrales Diagnosewerkzeug benötigen wir dazu zunächst einen Algorithmus, der den Istzustand (d.h., die Menge aller vergebenen und möglichen Systemstellen in Konstanz) nach einem bestimmten formalen Schema hierarchisiert (in Baumform überführt), um damit Diskrepanzen zur inhaltlich intendierten Hierarchie aufzuzeigen, mit dem langfristigen Ziel das formale Schema und die inhaltlich intendierte Hierarchie zur Deckung zu bringen, mit allen Folgen, evtl. auch Umsignierung. Das hier beschriebene Projekt wurde von der Autorin realisiert im Rahmen ihres Referendariats für den höheren Dienst an der Bibliothek der Universität Konstanz, unter der Ausbildungsleitung von Oliver Kohl-Frey. Betreut wurde es auf programmierter Seite von Bernd Schelling, dem Leiter der IT-Abteilung der Bibliothek, und Mitarbeitern der IT-Abteilung, namentlich Sebastian Danisch und Sorin Gheorghiu. Auf inhaltlicher Seite beraten wurde die Autorin von Ralph Hafner und Uwe Jochum, Fachreferenten an der Bibliothek. Als zeitlicher Rahmen waren ursprünglich sechs volle Wochen angesetzt, letztendlich sind es dann circa 10–12 Wochen geworden. Dieses Projekt ist eingebunden in ein größeres Projekt mit dem Ziel der Rundumerneuerung des "Schlagwort-Information-Systems" (SIS) der Bibliothek, welches im nächsten Abschnitt kurz umrissen werden soll.

Beschreibung SIS-Gesamtprojekt

Das Schlagwort-Information-System – seinerseits ebenfalls eine Konstanzer Eigenentwicklung – wurde 2001 auf einem Oracle-Datenbanksystem basierend in Betrieb genommen².

Ein paar Jahre später zeichnete sich ab, dass das SIS aus Modernisierungsgründen auf ein neues System umgezogen werden musste. Damit muss auch eine neue Oberfläche für das SIS programmiert werden. Die Ziele für das neue SIS samt zugehöriger Oberfläche wurden folgendermaßen formuliert (frei nach einer Auflistung im bibliotheksinternen Wiki, abzurufen³:

Für Fachreferenten:

- Bessere Übersicht über die eigene Systematik, und damit Erleichterung bei der Sacherschließungsarbeit. Die Informationen zur Systematik sind derzeit auf das alte SIS und zum Teil noch maschinenschriftlich verfasste sogenannte "Kladden" verteilt. Diese Informationen sollen im neuen SIS zusammengeführt und damit für eine beliebige Anzahl von Personen gleichzeitig sichtbar und recherchierbar sein. Die "Kladden" werden dann nicht mehr benötigt.
- Keine lokale Pflege der Verweise mehr, sondern Nutzung der Verweisungen in der Gemeinsamen Normdatei (GND).

Für Nutzer:

- Attraktive Möglichkeit zur Recherche, zum Browsen und zur thematischen Orientierung im Bestand.

Für die Bibliothek:

- Ende der Isolation der lokalen Klassifikation, Vernetzung mit der Sacherschließungsarbeit anderer Bibliotheken. Unter Verwendung von Semantic-Web-Technologien können fremde Sacherschließungsdaten (etwa aus der GND, der DDC oder der RVK) durch das gemeinsame Format RDF für die Erschließungsarbeit mit der Konstanzer Systematik nutzbar gemacht werden.

Die Projektplanung² sah zu dem Zeitpunkt Folgendes vor:

- Umzug der Daten auf eine PostgreSQL-Datenbank
- Entwicklung eines Viewers zur Sichtbarmachung der (angenommenen) Hierarchiestrukturen in der Systematik
- Ausbau der Recherchemöglichkeiten in diesem Viewer
- Einbinden der GND in die Recherche
- Entwicklung eines Bearbeitungstools für Fachreferenten
- Nutzung der Sacherschließungsarbeit anderer Bibliotheken mit Hilfe von Semantic-Web-Technologien für eine (halb-)automatisierte Sacherschließung nach Konstanzer Systematik.

Der Umzug der Datenbank ist mittlerweile erfolgt, ein Viewer, der das Browsen in der Systematik und eine einfache Suche erlaubt, wird aktuell von der IT-Abteilung der Bibliothek entwickelt. Vor der Inbetriebnahme eines unterstützenden Bearbeitungstools für Fachreferenten stand aber noch ein wesentlicher Schritt: Eine Analyse, inwieweit die ständig aktualisierte Systematik noch eine klare Struktur besitzt – wo sind Stellen, die relativ leicht

zu bereinigen sind, und wo ist kein durchgängiges Ordnungsprinzip mehr erkennbar? Der entstehende SIS-Viewer kann nur Hierarchiestrukturen anzeigen, wenn dahinter ein Algorithmus steht, der diese Strukturen aus der Systematik heraus analysiert, und für einen solchen Algorithmus war zum Beginn des vorliegenden Unterprojekts nur ein sehr rudimentäres, ebenfalls von der IT-Abteilung gestelltes Grundgerüst gegeben.

Aufbau von Systemstellen

Für eine Beschreibung des Aufbaus von Systemstellen in Konstanz siehe [1]. In dieser Darstellung nicht explizit erwähnt sind sogenannte Erstreckungen, also Bereichsangaben innerhalb der Systematik, z.B. 'psy 410 - psy 415' für "Kulturpsychologie". Diese wurden aber wohl schon von Beginn an mehr oder weniger konsequent ins SIS eingetragen, und am Anfang des vorliegenden Projektes stand die Idee, diese Erstreckungen auf Zahlenebene als Hierarchisierungsprinzip heranzuziehen². Aus Sicht der formalen Sprachentheorie liegt dies nahe, da zwischen korrekt geklammerten oder geschachtelten Strukturen und Baumstrukturen ein unmittelbarer Zusammenhang besteht.

Desweiteren stellen Fachgebiete, die durch den dritten Buchstaben schon in Teilgebiete untergliedert werden, natürlicherweise Erstreckungen auf Buchstabenebene dar, etwa 'jua - juz' ("Rechtswissenschaft"), und können weitere Untererstreckungen enthalten, z.B. 'jue - jur' ("Deutschland <Bundesrepublik> / Recht").

Schon in der Beschreibung der Inbetriebnahme des SIS in der "Bibliothek aktuell" von 2001³ wird davon gesprochen, dass Erstreckungen im SIS nun "aufklappbar" seien, de facto besteht diese Funktionalität lediglich darin, dass bei einem Klick auf eine im SIS eingetragene Erstreckung ein Sprung zum Anfang einer Liste der in dieser Erstreckung enthaltenen Einzelstellen erfolgt, dies ist jedoch noch keine Implementierung einer Baumstruktur im engeren Sinne.

Projektplan

Der Plan für das vorliegende Unterprojekt sah Folgendes vor:

- Formuliere anhand einiger weniger möglichst klarer Grundprinzipien einen Algorithmus, der die Daten aus der SIS-Datenbank in ihrem aktuellen Zustand einliest und diese in eine hierarchische Struktur überführt.
Beobachte,
(a) was für eine Struktur sich daraus ergibt, bzw.

(b) wie viele Stellen sich den Prinzipien des zuvor entwickelten Algorithmus' nicht unterordnen und

(c) welcher Art sie sind.

Die Entwicklung des Algorithmus' sollte unterstützt werden durch Recherchen in den diversen Dokumentationen zur Entstehung und zum Aufbau der Systematik und des SIS, und im Anschluss sollte er dann auch implementiert werden, aufbauend auf dem oben erwähnten, von der IT-Abteilung gestellten Programmstumpf.

Warum eine Baumstruktur?

Eine Systematik dient der Erschließung und der Recherche, also dem Einordnen und dem Wiederfinden. Um jedoch diese Funktionen einer systematischen Anordnung voll zu entfalten, benötigt man eine gut implementierte Suchmöglichkeit mit in verständlicher Form angezeigten

Ergebnissen (siehe [2]). Wie bereits erwähnt, ist das im bestehenden SIS nur bedingt gegeben, da die Ergebnisanzeige im Grunde auf einer Listenstruktur basiert, wobei einzelne Systemstellen und Erstreckungen ununterschieden (besser nicht unterscheidbar?) nebeneinander stehen.

Eine echte Baumstruktur hingegen ermöglicht eine weit aussagekräftigere Anordnung der Themenbereiche in einer Gliederungshierarchie. Dies wiederum hat Vorteile bei der Recherche: Jedes Objekt im Baum hat nur ein unmittelbar übergeordnetes Objekt und damit nur einen Pfad nach oben zur Wurzel, also zum Ausgangspunkt der Suche. Während man in einer linearen Anordnung nur durch erschöpfendes Blättern suchen kann und in einem Netz die Möglichkeiten zur Verzweigung zu vielfältig sind, ist ein Baum die ideale Struktur, um mit einigen wenigen Auswahlentscheidungen schnell an die gesuchte Information zu gelangen.

Der Algorithmus - Hierarchisierungsprinzipien

Der in dem vorliegenden Projekt entwickelte Algorithmus stützt sich auf zwei Grundprinzipien für die Hierarchisierung:

- Auflösung der durch die Erstreckungen in der Systematik gegebenen Klammerstruktur, und zwar sowohl auf Buchstaben- als auch auf Zahlen- als auch auf Schlüsselebene, soweit eingetragen.

Beispiele:

'juh' ("Strafrecht") sollte sich der Erstreckung 'jue - jur' ("Deutschland <Bundesrepublik> / Recht") unterordnen, und diese sollte sich wiederum der

Erstreckung

'jua - juz' ("Rechtswissenschaft") unterordnen.

'psy 412' ("Kulturpsychologie / Asien") sollte sich der Erstreckung

'psy 410 - psy 415' ("Kulturpsychologie")

unterordnen, welche sich wiederum der Erstreckung

'psy 400 - psy 427' ("Kultur- und Völkerpsychologie") unterordnen sollte.

Die schlüsselhaltige Signatur

'spr 76:e' ("Europa / Sprachkontakt / Gesamtdarstellung") sollte sich der Schlüsselerstreckung

'spr 76:d - spr 76:f' ("Europa / Sprachkontakt") unterordnen.

- Ausnutzung der Präfixrelation zwischen den Zeichenketten, aus denen Signaturen bestehen.

Beispiele:

'psy 418.50' ("Pastoralpsychologie") sollte sich

'psy 418' ("Religionspsychologie") unterordnen,

'kid 219:f' ("Datenverdichtung / Kongress")

sollte sich

'kid 219' ("Datenverdichtung") unterordnen, und

'kid 259.75:fa' ("Fourth International Conference on Logical Programming") sollte sich

'kid 259.75:f' ("Logische Programmierung / Kongress") unterordnen.

Das erste Prinzip leuchtet unmittelbar ein, das zweite ist für gewisse Fälle etwas strittiger, nämlich wenn die Erweiterung einer Stelle nicht zwingend mit einer Verengung des Themenbereichs einhergeht.

Als Folge der oben genannten Prinzipien ergibt sich, dass in der Systematik nicht eingetragene übergeordnete Stellen (das entspricht Mutterknoten im Baum) gegebenenfalls künstlich nachgeneriert werden müssen.

Konkrete Beispiele:

Wenn es 'kid 259.50' gibt, so muss es auch

'kid 259' geben, wenn es

'kid 220.60:k' gibt, so muss es auch

'kid 220.60' geben, und wenn es

'spr 12:a - spr 12:z' gibt, muss es auch

'spr 12' geben.

Für das vorliegende Projekt bot sich als Programmiersprache Ruby an aufgrund seiner Erweiterbarkeit in Richtung der Skriptsprache Javascript, welche wiederum besonders gut geeignet ist für die ange-

strebte Darstellung aufklappbarer Bäume im SIS-Viewer, der parallel von der IT-Abteilung der Bibliothek weiterentwickelt wurde. Einen Screenshot des Viewers zeigt Abbildung 1. Links befindet sich die Auswahlleiste nach Fachgebieten und -erstreckungen, mit dem Suchschlitz darunter können einzelne Systemstellen gesucht werden. In der Mitte entfaltet sich die Baumstruktur und rechts davon werden Informationen zu der jeweils angeklickten Stelle angezeigt. Unter dem zweiten Reiter ist eine Schlagwortsuche realisiert.

Die SIS-Datenbank wird abgefragt mittels einer PostgreSQL-Verbindung. Einträge in dieser Datenbank (also Einzelsystemstellen und Erstreckungen in der Systematik) werden von dem hier implementierten Programm in Knoten einer Baumstruktur umgewandelt und mitsamt der Mutter-Kind-Beziehungen zwischen diesen über die Zugriffsarchitektur Active Record in eine neue Datenbank gespeichert.

führt werden müssten, relativ leicht zu erkennen. An zahlreichen Stellen kommt es durch die strikte Befolgung der gesetzten Hierarchisierungsprinzipien zu kuriosen Unterordnungen, meist ausgelöst durch Flüchtigkeitsfehler, die man ansonsten eventuell übersehen hätte.

Desweiteren lässt sich feststellen: Trotz der schönen Struktur, die sich meist durch die Einhaltung des Präfixprinzips ergibt, gibt es Stellen, wo die Aufrechterhaltung dieses Prinzips strittig ist, da sich hier inhaltlich unerwünschte Unterordnungen ergeben. Besonders häufig ergeben sich solche unerwünschten Unterordnungen an Stellen, wo alphabetisch vercuttert wurde und ein Cutter-Code mehrfach belegt war, so dass er durch Buchstaben erweitert werden musste – die Erweiterungen ordnen sich dann natürlich unter dem ursprünglichen Code ein.

Abb. 1: SIS-Viewer

The screenshot shows the SIS-Viewer interface. At the top, there is a header for 'Bibliothek der Universität Konstanz' and 'Schlagwort-Informationssystem / Viewer'. Below this, there are tabs for 'Viewer' and 'Suchen'. On the left, there is a sidebar with a list of subject areas (Fachgebiete) and their extensions (Erstreckungen), including 'mus', 'mwa - mwz', 'ndl', 'ngr', 'ori', 'pha - phx', 'phy', 'pol', 'psy', 'rom', 'rum', 'sci', 'ska - sks', 'sla - slx', 'soz', 'spa', 'spk', 'spp', 'spr', 'ssl', 'sta', 'stq', 'swa', 'tec', 'tff', 'tha - thp', 'vok', and 'wra - wru'. Below this list is a 'Systemstellenfilter' input field. The main area displays a hierarchical tree structure of system locations, starting with 'tha - thp = Theologie'. The tree includes sub-entries like 'tha = Theologie / Religionswissenschaft / Allgemeines', 'thb - thc = Religionswissenschaft / Religion / Geschichte', 'thg - thj = Bibelwissenschaft', 'thh = Altes Testament', 'thi = Judentum', and 'thj = Neues Testament'. The right sidebar contains 'Informationen' for the selected entry, showing details like 'Systemstelle: thi 351', 'Standort: G 3', 'Erfassungsdatum: 18.07.2000', and 'Korrekturdatum: 18.07.2000'. There is also a 'Schlagwortkette' section showing 'Europa / Judentum'.

Ergebnisse der Diagnose

Im Zuge der Visualisierung der Systematik mit Hilfe des implementierten Algorithmus' kommen zunächst eine Reihe offensichtlicher, leicht zu korrigierender Fehler zum Vorschein. Am prominentesten sind sowohl im SIS bereits angelegte, als auch für den Aufbau der Baumstruktur künstlich erzeugte, im SIS noch anzulegende Stellen ohne Benennung (in beiden Fällen fehlt die Schlagwortkette). Auch sind manche Autorenstellen, die zusammenge-

Weiteres Vorgehen

Die Fachreferenten bekamen zunächst PDF-Dokumente mit den mit Hilfe des Algorithmus' erstellten Baumstrukturen für ihre jeweiligen Fachgebiete, in

welchen problematische Stellen markiert und an kritischen Stellen kommentiert waren, zur Durchsicht. Damit sollten offensichtliche Fehler (z.B. fehlende Benennungen) gleich korrigiert und ansonsten der Aufwand für eine Bereinigung abgeschätzt werden.

▼ **ori 278:a16 = Abu-'l-Hasan as-Suri**

ori 278:a16:k = Abu-'l-Hasan as-Suri / Einzelwerk

▶ **ori 278:a16a = Abu-Ali al-Qali, Ismail Ibn-al-Qasim**

▼ **ori 278:a16b = Abu-Duaib**

ori 278:a16b:k = Abu-Duaib / Einzelwerk

▶ **ori 278:a16bl = Abu-Dulaf Misar Ibn-al-Muhalhil al-Hazrabq**

▶ **ori 278:a16c = Abu-'l-Atahiya Ismail Ibn-al-Qasim**

▶ **ori 278:a16d = Abu-'l-Aswad ad-Duali, Zalim Ibn-Amr**

Da es für die Bibliothek von großem Interesse ist, möglichst wenige Medien umsignieren zu müssen, sollten die Fachreferenten auch angeben, wie groß die Anzahl der davon betroffenen Medien sein würde. Auch sollte die Zahl der unerwünschten Unterordnungen abgeschätzt werden, damit entschieden werden konnte, ob sich diese durch Umstrukturieren (inklusive Umsignierung) eliminieren ließen oder ob eine kritische Masse zusammenkam, so dass in der Folge an den klaren Prinzipien des Algorithmus' eventuell Abstriche gemacht werden mussten. Die Priorität wurde zunächst auf die Bereinigung der Grobstruktur auf Zahlenerstreckungsebene gelegt, um eine optimale Navigation in den Systematikbäumen zu erzielen. Nach reiflicher Diskussion haben sich die Fachreferenten letztendlich dafür entschieden, das Präfixprinzip aufgrund seiner unerwünschten Nebenwirkungen auszuhebeln, eine algorithmisch weniger durchstrukturierte Version zu verwenden und stattdessen die gewünschte Struktur durch von Hand eingezogene Erstreckungen zu erwirken. Die verschiedenen Versionen des Algorithmus' finden sich auch im bibliotheksinternen Wiki unter der Seite „SIS“.

Mittlerweile sind alle Versionen des Algorithmus' mit dem neuen Viewer kombiniert worden und stehen den Bibliotheksmitarbeitern zu Verfügung. Wie schon seit 2001 das SIS laufen nun auch die Algorithmusversionen täglich einmal durch, so dass Änderungen am Folgetag zu sehen sind. Die Hoffnung der Projektteilnehmer, dass die Bereinigungsarbeit mit dem Viewer signifikant schneller und für die Fachreferenten auch um Einiges attraktiver wird aufgrund der "sinnlichen Erfahrbarkeit der Baumstruktur" (Zitat Uwe Jochum, Fachreferent an der Bibliothek), die der Umgang mit dem Viewer transportiert, hat sich bestätigt.

Literatur:

1 Bösing, Laurenz, Joachim Stoltzenburg und Barbara Thomashoff: Regeln für den Aufbau von Buchsignaturen. Bibliothek aktuell, Sonderheft 1, 1969.

2 Hafner, Ralph und Uwe Jochum: SIS. Online unter: <http://w7.ub.uni-konstanz.de/wiki/index.php/SIS>, 2007 (Abgerufen am: 22.01.2013).

3 Wilkens, Karsten: . . . und was ist neu am neuen Schlagwortregister? Bibliothek aktuell, 75, 2001.

Fazit und Ausblick

Jedes Fachgebiet hat charakteristische strukturelle Eigenheiten, die wohl zurückzuführen sind auf die Praktiken der jeweils (aktuell oder ehemals) zuständigen Fachreferenten. Charakteristisch ist auch die höchst unterschiedliche Größe eines Fachgebietes (d.h., die Anzahl der Systemstellen, die es umfasst), die sich rein aus dem Ausnutzungsgrad des Zahlenraums von 1 bis 999 samt der gegebenen Möglichkeiten zur Unterteilung ergibt. Dadurch, dass wir diese Ausdifferenzierung nun in einer Baum- statt in einer Listenstruktur abbilden, haben wir Breite in Tiefe übersetzt.

Die Systematik lässt sich also in weiten Teilen in eine Baumform bringen, aber das Präfixprinzip ist sowohl auf Schlüssel- als auch auf Punktzahlenebene wohl zu stark, hier geben die Daten zu häufig keine Baumstruktur her.

Der in diesem Projekt entwickelte Algorithmus in Kombination mit dem neuen Viewer ist ein erster Schritt hin zu dem gewünschten Online-Werkzeug, mit dem Fachreferenten, Mitarbeiter der Bearbeitungsabteilung und auch Hilfskräfte standortunabhängig und eventuell sogar halbautomatisch systematisieren können, wobei jegliche dazu benötigte Information im System verfügbar sein soll, d.h., ohne auf individuell gepflegte, unzugängliche und überholte Kladden zurückgreifen zu müssen. Der nächste Schritt besteht in der Bereinigung der aufgedeckten Unregelmäßigkeiten in der Systematik, welche eventuell ebenfalls in Teilen durch Hilfskräfte erfolgen kann. Wiederum der nächste Schritt führt von der Struktur- auf die semantische Ebene mit dem Fernziel einer Anbindung an die GND.