



In: Hammwöhner, Rainer; Wolff, Christian; Womser-Hacker, Christa (Hg.): Information und Mobilität, Optimierung und Vermeidung von Mobilität durch Information. Proceedings des 8. Internationalen Symposiums für Informationswissenschaft (ISI 2002), Regensburg, 8. – 11. Oktober 2002. Konstanz: UVK Verlagsgesellschaft mbH, 2002. S. 289 – 302

Ein „Virtuelles Bibliotheksregal“ für die Informationswissenschaft als Anwendungsfall semantischer Heterogenität

Peter Hanke, Thomas Mandl, Christa Womser-Hacker

Informationswissenschaft - Universität Hildesheim
Marienburger Platz 22 - 31141 Hildesheim
hanke_peter@web.de
mandl@uni-hildesheim.de
womser@uni-hildesheim.de

Zusammenfassung

Das virtuelle Bibliotheksregal *MyShelf* soll den Browsing-Zugang zu heterogen repräsentierten Objekten erleichtern. Anwendungsfall sind die informationswissenschaftlichen Literaturbestände an der Universitätsbibliothek Hildesheim. *MyShelf* zielt darauf ab, die relevanten Bibliotheksbestände, Bestände anderer Bibliotheken, von Dozenten erstellte Lehrmaterialien und Quellen aus dem Internet zu integrieren. Der Zugang umfasst mehrere hierarchische Ordnungssystematiken, wobei vor allem bestehende reale Bibliothekssystematiken integriert werden. Die Systematik soll vom Benutzer ausgewählt werden können, woraufhin sich die Bücher nach der gewählten Systematik neu anordnen. Dazu müssen semantische Transferrelationen zwischen den Objekten oder Systematiken definiert werden. Der Beitrag berichtet über die konkreten Schritte zur Erstellung dieses Systems.

Abstract

The virtual library shelf *MyShelf* enhances the browsing access to heterogeneously represented objects. It is applied to the information science books in the library of the University of Hildesheim. *MyShelf* integrates the relevant library stock, other libraries and teaching material in the internet. The access is possible through various hierarchical cataloguing systems and especially library catalogues. The ordering system can be chosen by the user and the system reorganizes its content accordingly. For that purpose, transfer relations between the objects and categories need to be established. This article includes steps toward the implementation of *MyShelf*.



Dieses Dokument wird unter folgender [creative commons](http://creativecommons.org/licenses/by-nc-nd/2.0/de/) Lizenz veröffentlicht:
<http://creativecommons.org/licenses/by-nc-nd/2.0/de/>

1 Einleitung

Während die semantische Heterogenität im Information Retrieval für Suchsysteme inzwischen weitgehend als Problem erkannt wurde, wird dieser Aspekt beim Browsing-Zugang noch vernachlässigt. Dabei stellt gerade das Vorhandensein heterogener Systematiken, welche sich oft in unterschiedlichen hierarchischen Zugangssystemen äußern, eine sehr augenfällige Folge der Heterogenität dar. Dieser Artikel überträgt und adaptiert die existierenden Lösungsansätze für Suchsysteme auf Informationssysteme, die Browsing-Zugriff erlauben.

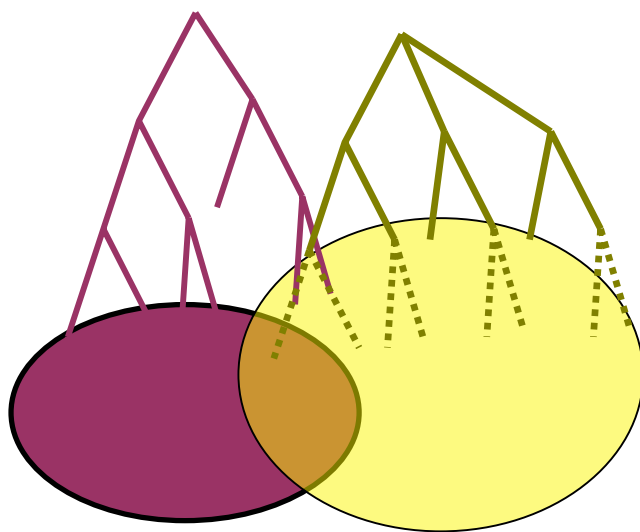


Abb. 1: Heterogene hierarchische Systematiken für sich überlappende Kollektionen

Dazu wird zunächst die Problematik der semantischen Heterogenität und ihrer Behandlung rekapituliert. Anschließend werden die konkreten Anforderungen und Lösungsansätze für das virtuelle Bibliotheksregal *MyShelf* eingeführt, das Browsing in heterogenen Bibliothekskollektionen ermöglicht. Kapitel vier und fünf zeigen im Detail, welche Schritte für die Erstellung einer Datenbasis für die Informationswissenschaft erforderlich waren. Das Fazit diskutiert Möglichkeiten, das System für den Alltagsbetrieb zu realisieren.

2 Semantische Heterogenität

Unerwünschte Heterogenität und die sich daraus ergebende Forderung nach Interoperabilität sind Aspekte der weltweiten Vernetzung von Informationssystemen. Die Heterogenität entsteht dabei v.a. auf drei Ebenen:

- Technische Heterogenität: Informationssysteme benutzen unterschiedliche Hardware, Zeichensätze, Datenformate, Kommunikationsprotokolle u.ä.

- **Metadaten-Heterogenität:** Informationssysteme basieren auf verschiedenen Standards für die Erfassung von Metadaten (Dublin-Core, MARC, usw.) oder unterschiedlichen Datenmodellen.
- **Semantische Heterogenität:** Die ersten beiden Ebenen sind gelöst, die inhaltlichen Repräsentation können technisch verglichen werden, aber die beschreibenden Elemente entstammen unterschiedlichen Ontologien oder Sinnzusammenhängen.

Alle Ebenen erschweren die Kommunikation und damit das Wissensmanagement für die Benutzer.

2.1 Ursachen der Heterogenität

Der Hauptgrund für die semantische Heterogenität liegt in der durchaus wünschenswerten virtuellen Integration von online bereitgestellten Wissensbeständen verschiedenster Anbieter. Die Trennung der Informationsangebote wird aus politischen und inhaltlichen Interessen bestehen bleiben, da viele Informationsanbieter sich durch ihr individuelles Angebot profilieren. Jeder Anbieter setzt für seine Daten seine optimierte Inhaltserschließung ein, ohne auf Kompatibilität mit anderen, möglicherweise konkurrierenden Informationsdienstleistern zu achten. Anbieter wie Fachinformationszentren entwickeln Thesauri für ihre Fachgebiete, die das Vokabular definieren und für die intellektuelle Indexierung verwendet werden. Anbieter mit kleinen Datenmengen greifen zunächst häufig auf eigens gebildete hierarchische Strukturen zu, ohne sich bewusst zu machen, dass hier ein neuer Thesaurus geschaffen wird.

Die semantische Heterogenität besitzt aber noch eine weitere Dimension. Missverständnisse treten nicht nur fachkulturübergreifend auf, sondern sind besonders zwischen verschiedenen nationalen oder ethnischen Kulturen folgenreich. Heterogenität vermindert gegenseitiges Verständnis in der privaten wie auch in der geschäftlichen Kommunikation.

Jede Einteilung der Welt in Begriffe - sei es in Form eines Thesaurus, einer Klassifikation oder einer sonstigen Ontologie - repräsentiert eine meist in sich sinnvolle Perspektive. Diese entspricht aber keineswegs immer der Sichtweise des Benutzers, besonders wenn dieser aus einer anderen Kultur oder auch Fachkultur stammt. Benutzer sind nicht an der globalen Einteilung des verfügbaren Wissensbestandes interessiert, sondern wollen mit den ihnen geläufigen Begriffen suchen.

Verschiedene Begriffssammlungen und Systematiken sind also in aller Regel nicht kompatibel. Es kommt auch oft vor, dass der gleiche Begriff eine andere Bedeutung hat, da er anders in den größeren Kontext eingebettet ist. Analysen solcher Probleme liegen für viele Fachgebiete und Anwendungsfälle vor (cf. z.B. BUCKLAND ET AL. 1999 für Bereiche der Ingenieurwissenschaft, CHEN ET AL. 1998 für die Biologie, MANDL & WOMSER-HACKER 2002 für Lehr- und Lernmaterialien, MANDL 2001 für einen Überblick, einen Überblick über die Rolle von Ontologien im Wissensmanagement bietet STAAB 2002.).

Im Information Retrieval Prozess führt Heterogenität dazu, dass jede Quelle mit den für sie spezifischen Begriffen befragt werden müsste. Aus Sicht des Benutzers stehen der von ihm intendierten Bedeutung des Anfrageterms potentiell mehrere unterschiedliche Bedeutungen gegenüber. Häufig weckt die semantische Heterogenität den Wunsch nach einem weltübergreifenden Ordnungssystem für Wissen. Zahlreiche Versuche in diese Richtung sind bereits unternommen worden und gescheitert. Solche Lösungen sind weder durchsetzbar, noch könnten sie die erforderliche Dynamik für die Integration neuen Wissens bieten. Ideen dieser Art stehen hinter dem sogenannten *Ontology Merging* (cf. NOY & MUSEN 1999).

Bei gemischten Korpora, die sowohl intellektuell als auch automatisch indexierte Dokumente enthalten, verschärft sich die semantische Heterogenität. Diese Situation liegt etwa im Internet vor, wo Suchmaschinen automatisch indexieren, aber auch intellektuelle Inhaltsanalysen in Form von Meta-Daten des Autors oder einer Zuordnung etwa in einem Subject Guide vorliegen können. Die Möglichkeiten der Virtualisierung stellen das Information Retrieval also vor die Herausforderung, Suchen in semantisch stark heterogenen Umgebungen zu ermöglichen.

2.2 Behandlung der Heterogenität

Die Behandlung von Heterogenität erfolgt über Transferrelationen zwischen verschiedenen Ontologien. Kuhlen spricht in diesem Zusammenhang von einer „postkoordinierenden Ordnung durch transformierende Anpassung“ (KUHLEN 1999:138). Zwei Familien von Methoden können dafür eingesetzt werden (cf. HELLWEG ET AL. 2001):

- Intellektuell definierte Übergänge zwischen Ordnungssystematiken
- Assoziative, vage Beziehungen, die durch maschinelles Lernen abgeleitet werden (cf. MANDL 2001)

Erstere führen zu sicheren Beziehungen, die etwa in Form von wenn-dann-Regeln formalisiert werden.

2.3 Maschinelles Lernen

Die Grundlage für vage Transfer-Relationen stellen maschinelle Lernverfahren dar. Maschinelles Lernen beschäftigt sich mit Software, welche aufgrund von Eingaben ihr Verhalten bzw. ihren Output verändern kann. Die Konstruktion eines Modells des Gegenstandsbereichs erlaubt es, Aufgaben besser zu erledigen. Das maschinelle Lernen unterscheidet verschiedene Lernstrategien:

- Lernen durch Deduktion
- Lernen durch Analogie
- Induktives Lernen (anhand von Beispielen)

Alle diese Strategien lassen sich auch für Transferrelationen einsetzen. Vor allem kommt der induktive Ansatz zur Anwendung, bei dem meist aus vielen Beispielen ein quantitatives Modell abgeleitet wird.

Vage Relationen lassen sich als vage Regeln darstellen, wie etwa in Form von Wahrscheinlichkeiten für bestimmte Beziehungen. Grundlage für die vagen Verfahren sind Objekte, bei denen die Beziehungen zu allen Systematiken bekannt sind bzw. schon intellektuell erschlossen wurden. Damit bildet gewissermaßen auch hier die menschliche Informationsarbeit die Basis. Daneben kommen aber auch Heuristiken für die Ableitung von Doppelkorpora zum Einsatz (cf. MANDL 1999). Maschinelle Lernverfahren extrapolieren die Beziehungen dann auf andere Objekte. Besonders wichtig sind in diesem Zusammenhang die Erfahrungen aus der Text-Kategorisierung (cf. JOACHIMS 2002). Da die Auswahl des für den Anwendungsfall besten maschinellen Lernverfahrens ein Problem darstellt, kommen zunehmend Kombinationen von mehreren Algorithmen zum Einsatz (cf. MANDL & WOMSER-HACKER 2001). Während die Erstellung intellektueller Übergänge sehr teuer ist, läuft sie bei vagen Relationen automatisch ab.

2.4 Anwendungen

Die zahlreichen Anwendungen von Systemen zur Heterogenitätsbehandlung ergeben sich aus den Anforderungen und finden sich in allen Bereichen, in denen integrierte Informationssysteme auf Objekte mit mehreren unterschiedlichen Repräsentationsgrundlagen zugreifen.

Wichtige Anwendungen zeichnen sich derzeit im Umfeld der Vision des semantic web ab, das durch intensives Markup von Web-Seiten datenbankähnli-

che Abfragen im Internet ermöglichen will (cf. BERNERS-LEE ET AL. 2001). Ein Beispiel bieten DOAN ET AL. 2002, die eine Abbildung zwischen den im Internet zugänglichen Kursen verschiedener Universitäten erstellen. Dies erfordert das Mapping heterogener Fachbezeichnungen, Niveaus und Inhalte aufeinander.

3 Das virtuelle Bibliotheksregal *MyShelf*

Semantische Heterogenität überwiegt trotz zahlreicher Standardisierungsbemühungen nach wie vor auch im Bibliotheksbereich.

3.1 Ausgangssituation

Die Präsenz-Bibliothek hat auch im Zeitalter Digitaler Bibliotheken ihre wichtige Rolle noch nicht verloren. Der Gang zum Regal repräsentiert gewissermaßen den Urahn aller Browsing-Systeme und stellt häufig eine adäquate Strategie bei der Literatursuche dar. Die Vorteile liegen neben dem direkten Zugriff in dem Ordnungssystem. Die dem Benutzer vertraute lokale Aufstellungssystematik liefert einen erheblichen Mehrwert. Der direkte Zugriff bietet weitere Vorteile. Manche Benutzer erkennen anhand der Gestaltung des Buches den Verlag und können eine erste Relevanzentscheidung treffen. Auch der schnell ersichtliche Umfang kann bei bestimmten Informationsbedürfnissen eine Rolle spielen.

Der Nachteil hierarchischer Systematiken wie auch des Bibliotheksregals liegt in der Starrheit. Das Ordnungssystem repräsentiert immer nur eine Sichtweise von vielen möglichen, was auch zu einer Vielzahl von Bibliotheksklassifikationen geführt hat. Auf aktuelle Entwicklungen innerhalb der Universitäten oder der Wissenschaften kann die Bibliothek nur langsam reagieren. Dabei würde sie die angestrebte Konsistenz gefährden. Das Bibliotheksregal ist vom Umfang her beschränkt, es fehlen Titel aus den nicht frei zugänglichen Beständen, ebenso gibt es oft keinen Hinweis auf ausgeliehene Titel oder Bücher in Semesterapparaten (die wiederum eine neue Sichtweise auf die Dokumente einführen). Äußerst schwierig ist meist das Einfügen von Dokumenten, die nicht vom Typ „Buch“ sind. Video- und Audio-Dokumente sowie Datenträger werden oft getrennt aufbewahrt und das Einstellen von Unterlagen für Lehrveranstaltungen, die von Dozenten selbst erstellt wurden, ist meist nicht vorgesehen.

Die Literaturversorgung für das kürzlich eingerichtete Fach Informationswissenschaft an der Universität Hildesheim bietet einen Anwendungsfall, bei

dem zahlreiche Aspekte der semantischen Heterogenität auftreten. Die Literaturversorgung und Ausstattung der Bibliothek ist besonders aus studentischer Sicht nicht optimal, da keine Signatur für die Informationswissenschaft existiert und relevante Literatur an sehr unterschiedlichen Orten aufgestellt wird. Die Studierenden greifen aufgrund des notwendigerweise beschränkten Umfangs auf viele andere Quellen zu und werden dort mit anderen Systematiken konfrontiert.

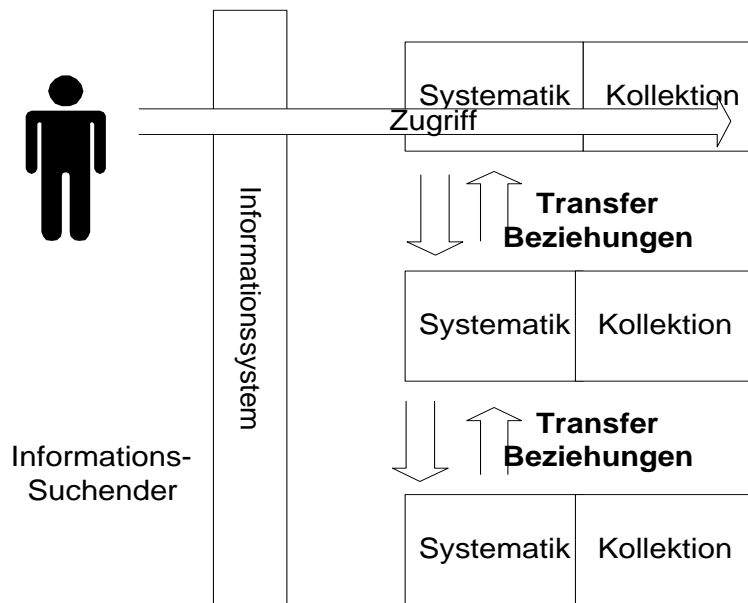


Abbildung 2: Die Lösung der Heterogenität in MyShelf: Transfer-Beziehungen erlauben den Zugriff auf mehrere Kollektionen über eine Ontologie

Die Informationswissenschaft ist stark interdisziplinär ausgerichtet, daher sind viele Perspektiven auf das Gebiet möglich, was sich in den Aufstellungssystematiken informationswissenschaftlicher Literatur an verschiedenen Universitäten in Deutschland widerspiegelt.

Quellen aus dem Internet und elektronische Kursmaterialien gewinnen zwar zunehmend an Bedeutung und Relevanz für die Ausbildung, sie sind aber nicht nach bibliothekarischen Gesichtspunkten erschlossen und nicht am gleichen Standort zu finden wie relevante Bücher. Wissenschaftler und Studierende wünschen sich, von ihrem Arbeitsplatz aus auf die relevanten Materialien zuzugreifen.

3.2 Lösungsansatz *MyShelf*

Das Ziel des virtuellen Bibliotheksregals *MyShelf* besteht in einer virtuellen Integration von mehreren Dokument-Beständen, bei der sich die semantische

Heterogenität nicht nachteilig für den Anwender auswirkt. Zunächst kann der Benutzer also mit *MyShelf* auf den maximal verfügbaren Korpus und damit auf alle Objekte zugreifen. Darüber hinaus bietet das System auch alle Systematiken an, welche die einzelnen Teilkorpora beschreiben. Im Anwendungsfall Informationswissenschaft sollen also möglichst viele relevante heterogene Bibliothekssystematiken eingehen. Damit hat jeder Benutzer die ihm vertraute Systematik zur Verfügung, er kann aber auch auf andere Sichtweisen ausweichen, um zu prüfen, ob er bei einem konkreten Suchproblem mit einer anderen Systematik eventuell erfolgreicher ist.

Der entscheidende Mehrwert besteht aber darin, dass immer die gesamte Kollektion nach der aktuell ausgewählten Systematik geordnet wird. Das virtuelle Bibliotheksregal kennt also mehrere Organisationsmöglichkeiten für seinen Inhalt und passt diese den Präferenzen des Benutzers an.

3.3 Benutzungsoberfläche

Die Präsentation und die Interaktion des Benutzers mit großen Mengen von Kategorien stellt nach wie vor ein Problem dar, das für jeden Anwendungsfall erneut optimiert werden muss. Allein die Existenz zahlreicher Alternativen für die Darstellung und Bedienung von hierarchischen Strukturen zeigt die Vielfalt der Gestaltungsmöglichkeiten und bei genauer Analyse zahlreiche Detailschwierigkeiten (cf. BEKAVAC 1999). Die für den *MyShelf*-Ansatz entscheidende Visualisierung mehrerer hierarchischer Strukturen wurde bisher selten realisiert. MUKHERJEA ET AL. 1995 präsentieren eine Anwendung in der für Hypertext-Strukturen mehrere Ursprungsknoten gewählt werden, die jeweils in eine andere Hierarchie münden. Die parallele Präsentation mehrerer Produktthesauri in einem wirtschaftlichen Informationssystem findet sich bei MANDL & STEMPFHUBER 1998.

Für die Darstellung nicht hierarchisch sondern assoziativ verbundener Konzepte entstehen zunehmend stärker visualisierte Darstellungsformen wie zweidimensionale Karten (cf. EIBL & MANDL 2002). Diese könnten sich besonders für die Darstellung von vagen Transferrelationen eignen, die durch Verfahren des maschinellen Lernens abgeleitet wurden. In diesem Beitrag steht die Visualisierung nicht im Fokus.

4 Bewertung von Bibliotheken

Der Zugriff auf informationswissenschaftliche Bücher in der Universitätsbibliothek Hildesheim stellt, wie bereits aufgezeigt, einen konkreten Anwendungsfall für das Problem der semantischen Heterogenität dar. Um Studierenden und Wissenschaftlern einen leichteren Zugang zur gesuchten Literatur zu ermöglichen, sollte exemplarisch eine einheitliche Sichtweise auf den informationswissenschaftlichen Bücherbestand der UB Hildesheim erarbeitet werden. Zu diesem Zweck musste die entsprechende Literatur erfasst werden, um sie anschließend einer neuen Ordnungssystematik zuzuführen.

Zur Erschließung des relevanten Buchbestandes in der UB Hildesheim wurde der Weg des Abgleichs mit dem Literaturangebot anderer Bibliotheken beschritten. Dabei boten sich Standorte informationswissenschaftlicher Studiengänge an. Für die Durchführung wurden die Bibliotheken der Universitäten Konstanz, Regensburg und Saarbrücken ausgewählt. Um Aussagen darüber treffen zu können, wie geeignet die einzelnen informationswissenschaftlichen Bestände für einen Abgleich mit den Titeln der Hildesheimer UB sind, erfolgte eine stichpunktartige Bewertung der genannten Bibliotheken.

Im Rahmen einer Literaturrecherche mit 200 verschiedenen Titeln der Informationswissenschaft sollte aufgezeigt werden, welche Überschneidungen zum Hildesheimer Bestand existieren. Jeweils 50 der zu recherchierenden Titel stammten dabei aus den Universitätsbibliotheken Konstanz, Regensburg, Saarbrücken bzw. Hildesheim. Sie wurden aus 20 unterschiedlichen Systemstellen der dortigen informationswissenschaftlichen Klassifikationen entnommen (abgesehen von Hildesheim, wo Schwerpunktbereiche der dortigen Lehre gewählt wurden).

Die Auswertung der Recherchen zeigte, dass Konstanz die größten Überlappungen mit den Hildesheimer Beständen aufweist (cf. HANKE 2002:36 ff.). Von den insgesamt 80 in Hildesheim gefundenen Titeln wurden 56 auch in Konstanz geführt. In Regensburg belief sich die Schnittmenge auf 48 Titel, von diesen waren jedoch nur fünf in der Regensburger Verbundklassifikation unter einer informationswissenschaftlichen Signatur erfasst oder mit einer entsprechenden Nebenstelle versehen; dies trifft ebenfalls auf nur 34 der insgesamt 110 gefundenen Regensburger Titel zu. In Konstanz hingegen besaßen von den 56 Titeln 41 eine Signatur der dortigen Ordnungssystematik „kid“ (Kybernetik, Informatik, Datenverarbeitung und Informationswissenschaft), so dass ein Abgleich mit den Daten einer Bestandsdatei des „kid“ die besten Ergebnisse versprach. In Saarbrücken fanden sich im Vergleich dazu nur 26

der 80 Hildesheimer Titel wieder, von denen bis auf fünf auch alle im „kid“ geführt werden. Die Gesamtbestände der in den einzelnen Bibliotheken unter einer informationswissenschaftlichen Signatur eingeordneten Bestände unterstützen die Ergebnisse der Studie: in Konstanz sind es mehr als 23.000, in Regensburg weniger als 1.000 und in Saarbrücken etwa 2.500 Titel (Stand: April 2002). Der im Folgenden beschriebene Abgleich ergab für die UB Hildesheim eine Menge von ca. 5.000 informationswissenschaftlichen Titeln.

5 Extrahieren des informationswissenschaftlichen Buchbestandes

Um den informationswissenschaftlichen Bücherbestand der Universitätsbibliothek möglichst umfassend zu bestimmen, sollte ein automatischer Abgleich mit einer geeigneten Bestandsliste von Titeln durchgeführt werden. Aus den beschriebenen Gründen wurde dafür die Bestandsdatei des „kid“ ausgewählt. Das entsprechende Gegenstück, eine Datei mit relevanten Hildesheimer Titeln, musste erstellt werden. Eine Erfassung des Gesamtbestandes der Hildesheimer UB kam aus technischen Gründen nicht in Betracht.

Somit wurde alternativ die Ermittlung einer relevanten Teilmenge anvisiert, welche anhand einer Heuristik erarbeitet wurde. Dazu wurde eine Liste informationswissenschaftlicher Begriffe erstellt, die sich am so genannten „alphabetischen Register“ der Universitäts- und Landesbibliothek Düsseldorf orientierte¹. Eingang fanden auch Bezeichnungen von Systemstellen der Bibliotheksklassifikationen in Konstanz, Regensburg und Saarbrücken, und Klassifikationsbeschreibungen der UB Hildesheim, sowie Begriffe der Basisklassifikation.

Diese Schlüsselbegriffe dienten der Bestimmung von Notationen in der UB Hildesheim, die informationswissenschaftliche Aspekte beinhalten. Dafür wurden sämtliche Klassifikationsbeschreibungen der UB Hildesheim nach den Begriffen durchsucht. Nachdem nun diese Notationen bekannt waren, konnten die zugehörigen Titel per OPAC-Suchanfragen nach den entsprechenden Systemstellen heruntergeladen werden. Die Gesamtmenge der so gefundenen Titel repräsentierte die Bestandsdatei der informationswissenschaftlichen Titel der UB Hildesheim. Denkbar wäre auch gewesen, direkt mit den Schlüsselbegriffen Suchanfragen an den OPAC zu richten, ohne den Umweg über die Notationen zu wählen. Es hatte sich aber im Verlauf der Untersuchungen gezeigt, dass Titelstichwort- und Schlagwortsuchanfragen keine

¹ <http://www.uni-duesseldorf.de/WWW/ulb/iudreg.html>.

zufriedenstellenden Ergebnisse erwarten lassen. Dies liegt sowohl an der teils mangelnden Aussagekraft von Buchtiteln, als auch an der nicht vollständigen bzw. dem Benutzer nicht vertrauten Verschlagwortung (cf. HANKE 2002:24 ff.).

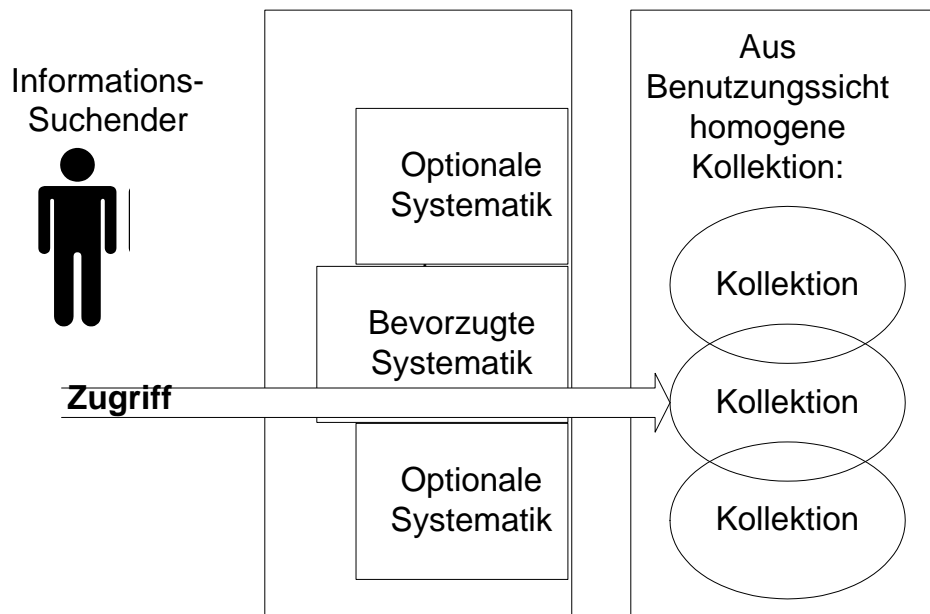


Abbildung 3: Benutzersicht auf MyShelf

Genau wie die vorliegende Bestandsdatei des „kid“ wurden die Hildesheimer Titel geparkt, so dass beide Dateien am Ende eine einheitliche Struktur aufwiesen: pro Titel eine Zeile mit dem Autor, einem Kurztitel, sowie in der Hildesheimer Datei der Signatur und zugeordneten Nebenstellen (Notationen). Die so bearbeiteten Dateien wurden automatisch abgeglichen. Um die Identität von Büchern zu erkennen, mussten Heuristiken entwickelt werden. Ergänzend dazu wurden, um beim Abgleich nicht erfasste Titel zu ermitteln, Literaturschaffungen von Dozenten informationswissenschaftlicher Lehrveranstaltungen berücksichtigt. Auch Suchanfragen mit Bezeichnungen einer neu erstellten „virtuellen“ Ordnungssystematik für Hildesheimer Bücher der Informationswissenschaft erweiterten den Kreis der Titel. Diese neue Systematik für Hildesheim basiert im Wesentlichen auf den Inhalten der Klassifikationen der Bibliotheken in Konstanz, Regensburg und Saarbrücken, sowie auf der vom Hochschulverband für Informationswissenschaft (HI e.V.) aufgestellten Gliederung für informationswissenschaftliche Lehrveranstaltungen. Die Organisation der neuen Systematik orientiert sich an qualitativen Analysen der anderen Bibliothekssystematiken und versucht, den Zugang vor allem für Studierende zu erleichtern.

Der Ordnungssystematik wurden die beim Abgleich ermittelten Hildesheimer Titel teils halbautomatisch, über die Liste der Schlüsselbegriffe mit zugeordneten Notationen, teils intellektuell zugeordnet.

Die Hildesheimer Studierenden sollen so die Möglichkeit erhalten, unabhängig von der bestehenden lokalen Aufstellungssystematik, über diese neue Ontologie auf informationswissenschaftliche Bücher zugreifen zu können. Durch die Berücksichtigung weiterer informationswissenschaftlicher Systematiken wird die Erstellung intellektuell definierter Transferrelationen stark erleichtert.

Ein weiteres Ziel wäre, über den Einsatz von lernenden Verfahren zu erreichen, dass automatisch erstellte Transferrelationen zwischen den bestehenden Klassifikationen der verschiedenen Bibliotheken erstellt werden. Die im Abgleich zwischen den Bibliotheken gefundene Datenmenge ließe sich für diese Aufgabe anpassen. Sie bilden einen Doppelkorpus, da darin für die gleichen Bücher je eine Repräsentation in zwei heterogenen Systematiken vorliegt. Damit lässt sich ein virtuelles Bibliotheksregal realisieren, das sich dadurch auszeichnet, dass der Bestand mehrerer Bibliotheken zur Verfügung steht, dass nach wie vor alle Systematiken eingesetzt werden und der Benutzer jederzeit zwischen Systematiken wechseln kann.

6 Ausblick

Damit die virtuelle Ordnungssystematik und deren Inhalte zur Anwendung gebracht werden können, müssen Realisierungskonzepte mit der Universitätsbibliothek erarbeitet werden. Dabei wäre vorstellbar, dass die beim Abgleich ermittelten Titel über ihre Signaturen den einzelnen Systemstellen der neuen Klassifikation zugeordnet werden. Beim Anwählen eines Teilgebietes der Klassifikation würde dann eine Suchanfrage nach den entsprechenden Signaturen gestartet. Neue Titel würden automatisiert eingestellt, indem bei bestimmten Suchanfragen (die genauer zu bestimmen wären) die Signaturen der Treffer automatisch einer Systemstelle zugeordnet würden. Auf diese Weise wären keine Einträge in die Datenbank notwendig.

Alternativ dazu könnten neue Bücher bei der Einstellung durch die Bibliothekare eine informationswissenschaftliche Nebenstelle (Notation) zugewiesen bekommen, so dass sie dann über diese bei einer Abfrage auffindig gemacht werden könnten. Dabei müssten die Bibliothekare jedes neue Buch, das sie in die Klassifikation einordnen, mit der passenden Notation versorgen. Zu die-

sem Zweck wäre es notwendig, vorab die gesamte neue Ordnungssystematik mit einer zu bestimmenden Notation auszuzeichnen. Die informationswissenschaftliche Klassifikation würde somit zum Bestandteil der bestehenden lokalen Systematik der UB Hildesheim.

Eine Realisierung außerhalb des Bibliothekssystems in einer eigenen Schicht erlaubt auch die spätere Integration von Internet-Dokumenten. In einem Prototyp von MyShelf werden die neue Systematik und das virtuelle Bibliothekssystem mit Studierenden evaluiert.

7 Literaturverzeichnis

- Bekavac, Bernard (1999): Suche und Orientierung im WWW. Verbesserung bisheriger Verfahren durch Einbindung hypertextspezifischer Informationen. Konstanz: Universitätsverlag [Schriften zur Informationswissenschaft Bd. 37].
- Berners-Lee, Tim; Hendler, James; Lassila, Ora (2001): The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In: Scientific American. 17. Mai 2001.
<http://www.sciam.com/issue.cfm?issueDate=May-01>
- Buckland, Michael; Gey, Fred; et al. (1999): Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies. In: D-Lib Magazine vol. 5(1).
<http://www.dlib.org/dlib/january99/buckland/01buckland.html>
- Chen, Hsinchun; Martinez, Joanne; Kirchhoff, Amy; Ng, Tobun; Schatz, Bruce (1998): Alleviating Search Uncertainty through Concept Associations: Automatic Indexing, Co-Occurrence Analysis, and Parallel Computing. In: Journal of the American Society for Information Science. JASIS 49(3). S. 206-216.
- Doan, AnHai; Madhavan, Jayant; Domingos, Pedro; Haelvy, Alon (2002): Learning to Map between Ontologies on the Semantic Web. In: Proceedings of the (WWW2002)
<http://www2002.org/CDROM/refereed/232/index.html>
- Eibl, Maximilian; Mandl, Thomas (2002): Including User Strategies in the Evaluation of Graphic Design Interfaces for Browsing Documents. In: Skala, Vaclav (ed.): Journal of WSCG. Special issue. The 10th Intl Conf in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2002). Pilsen. vol 1. S. 163-169.
http://wscg.zcu.cz/wscg2002/Papers_2002/B89.pdf
- Hanke, Peter (2002): Neue Chancen und Möglichkeiten für Ordnungssystematiken durch Virtualisierung: Anwendung am Beispiel der Erfassung und Klassifizierung des informationswissenschaftlichen Bücherbestands der Universität Hildesheim. Masterarbeit.
- Hellweg, Heiko; Krause, Jürgen; Mandl, Thomas; Marx, Jutta; Müller, Matthias N.O.; Mutschke, Peter; Strötgen, Robert (2001): Treatment of Semantic Heterogeneity in Information Retrieval. IZ-Arbeitsbericht Nr. 23, IZ Sozialwissenschaften, Bonn.
http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/index.htm#ab23
- Joachims, T. (2002): Learning Text Classifiers with Support Vector Machines. Kluwer.

- Kuhlen, Rainer (1999): Die Konsequenzen von Informationsassistenten: Was bedeutet informationelle Autonomie oder wie kann Vertrauen in elektronische Dienste in offenen Informationsmärkten gesichert werden? Frankfurt a.M.
- Mandl, Thomas (1999): Effiziente Implementierung von statistischen Assoziationen im Text-Retrieval. In: Ockenfeld, Marlies; Mantwill, Gerhard (eds.): Information und Region. 51. Jahrestagung der Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis e.V. (DGI). Hamburg, 21.-23.9.1999. S. 159-172.
- Mandl, Thomas (2001): Tolerantes Information Retrieval: Neuronale Netze zur Erhöhung der Adaptivität und Flexibilität bei der Informationssuche. Konstanz: Universitätsverlag [Schriften zur Informationswissenschaft Bd. 39].
- Mandl, Thomas; Womser-Hacker, Christa (2001): Fusion Approaches for Mappings Between Heterogeneous Ontologies. In: Constantopoulos, Panos; Sølvberg, Ingeborg (Hrsg.): Research and Advanced Technology for Digital Libraries: 5th European Conference (ECDL 2001) Darmstadt 4.-8.9.2001. Springer [LNCS 2163]. S. 83-94.
- Mandl, Thomas; Stempfhuber, Maximilian (1998): Softwareergonomische Gestaltung von Wirtschaftsinformationssystemen am Beispiel von ELVIRA. In: Ockenfeld, Marlies; Mantwill, Gerhard (eds.): Information und Märkte. Kongreß der Deutschen Gesellschaft für Dokumentation (DGD). Bonn, 22.-24.9.98. Frankfurt am Main. S. 145-157.
- Mandl, Thomas; Womser-Hacker, Christa (2002): Anforderungen an Information Retrieval Systeme in Bildungsmärkten. In: Schmidt, Ralph (Hrsg.): Content in Context: Perspektiven der Informationsdienstleistung. 24. Online-Tagung der DGI. Frankfurt am Main. 4-6. Juni 2002. S. 67-78.
- Mukherjea, Sougata; Foley, James; Hudson, Scott (1995): Visualizing Complex Hypermedia Networks through Multiple Hierarchical Views. In: Proc ACM Conference on Human Factors and Computing Systems (SIGCHI 1995) Denver. Colorado. http://www.acm.org/sigchi/chi95/Electronic/documnts/papers/sm_bdy.htm
- Noy N.; Musen, M. (1999): SMART: Automated Support for Ontology Merging and Alignment. In: Twelfth Workshop on Knowledge Acquisition, Modeling and Management. Banff, Alberta, Canada. 16.-21. Oktober, 1999. <http://sern.ucalgary.ca/KSI/KAW/KAW99/papers/Fridman1/NoyMusen.pdf>
- Staab, Steffen (2002): Wissensmanagement mit Ontologien und Meata Daten. In: Informatik Spektrum 25(3). S. 194-209.