

## **Information-Retrieval im Kontext Virtueller Fachbibliotheken**

**Vortrag von Franz Jürgen Götz am 04.06.2009 auf dem Bibliothekartag in Erfurt**

**Themenkreis 11 – Bibliothekskooperationen – alte und neue Partner**

**Sektion: Virtuelle Fachbibliotheken – Podiumsdiskussion mit Einführungsreferaten zu ViFa Problembereichen**

Blickt man heute auf die Möglichkeiten des Information-Retrieval in Virtuellen Fachbibliotheken, so möchte man nicht vermuten, dass es hier aktuell einen Diskussionbedarf gibt. Doch die vor kurzem erschienenen Meldungen zu den Weiterentwicklungen bei Google (Options: z.B. Timeline, Wonderweel) sowie zum Start der neuen Suchmaschinen WolframAlpha oder Bing, aber auch die jüngsten Entwicklungen in einigen Virtuellen Fachbibliotheken signalisieren, sich wieder auf eine der Kernaufgaben Virtueller Fachbibliotheken zu konzentrieren: die Gestaltung der Suche nach konsistenten, nutzerfreundlich aufbereiteten und kontrollierten Fachinformationen. Der Beitrag möchte – gerade im Zusammenhang mit der Verstetigung der Virtuellen Fachbibliotheken und ihren Entwicklungsmöglichkeiten in dieser Phase, diese Diskussion wieder anzuregen.

### **1. Blick auf die Situation des Information-Retrieval in Virtuellen Fachbibliotheken**

#### **a) Texte der Website als Informationsquelle und ihre Strukturierung**

Das Information-Retrieval in Virtuellen Fachbibliotheken – dies sollte in diesem Zusammenhang noch einmal kurz angesprochen werden – beginnt nicht erst mit dem Start der Metasuchmaschine. Die Website selbst und ihre Inhalte sowie ihre Struktur sind für zahlreiche Nutzer ein erster Einstieg in ihrer Suche nach Fachinformationen. Gerade bei Virtuellen Fachbibliotheken, die mehrere Fächer umfassen oder sich zu kooperativen Einheiten zusammengeschlossen haben wie z.B. das erst kürzlich gestartete [Portal Medien – Bühne – Film](#), oder die [Virtuelle Fachbibliothek für die Altertumswissenschaften Propylaeum](#) nutzen die Startseite, um den Nutzer über eine Bereichswahl (z.B. zur Filmwissenschaft, oder zur Archäologie) von Beginn an zu führen. Die Strukturierung der Startseite trägt, auch das ist nicht neu, wesentlich zur Lenkung des Nutzerinteresses bei. In der Regel werden hier bereits zentrale Recherchebereiche vorgestellt, so z.B. bei der [ViFaMath](#) oder dem [Wissenschaftsportal b2i](#). Im Zentrum steht oftmals ein Einstieg in die Metasuche, selten in so erweiterter Form wie bei der [Virtuellen Fachbibliothek Cibera](#). Bisweilen wirken die Startseiten auch ein wenig überfrachtet mit zahlreichen Möglichkeiten der Informationsgewinnung, z.B. mit verschiedenen themengebundenen Einstiegen. Es scheint im

Rahmen regelmäßiger Oberflächenneugestaltungen auf jeden Fall geboten, entsprechende Konzepte regelmäßig auf den Prüfstand zu stellen.

## **b) Metasuche**

Unser Hauptaugenmerk soll jedoch im Folgenden auf die Metasuche gelenkt sein. Hier scheint die Situation geklärt. Dem einzeiligen Suchschlitz auf der Startseite bzw. oft mitlaufend auch auf den Unterseiten, der eine googleartige Recherche suggerieren soll, steht für die „Profi-Sucher“ eine erweiterte Suchmaske mit mehreren Suchfeldern und Suchkombinationen zur Seite. Hier erfährt der Nutzer auch, in welchen Datenbeständen er recherchieren kann. Diese kann er dann entweder alle anwählen oder sich sein Rechercheset situationsbedingt zusammenstellen. Zumeist sind die einzelnen Datenquellen über Z39.50-Schnittstellen angebunden, was bei komplexeren Suchen oft zu längeren Antwortzeiten der Zielsysteme führt, evtl. können diese Daten auch nicht übermittelt werden. Eine weitere Eingrenzung der Suche nach dem Trefferüberblick ist darüberhinaus mit dieser Technik nicht möglich. Ein weiterer Punkt, der sich aus der Technik ergibt ist die Art der eingebundenen Datenpools – Bibliothekskataloge mit entsprechenden Schnittstellen stehen dabei im Vordergrund. Zahlreiche kleinere, aber evtl. interessante Datenbestände, die nicht über eine solche Schnittstelle verfügen, müssen dann entweder vom Partner mit einer entsprechenden Schnittstelle versehen werden, oder sie können in einer so strukturierten Metasuche nicht berücksichtigt werden. Zudem muss der Anbieter ein ständiges Angebot gewährleisten können, d.h. auch über die entsprechende Serverstruktur etc. verfügen.

## **2. Weiterentwicklung des Information-Retrieval im Hintergrund: Einbindung von Suchmaschinentechnologie in die Metasuche**

Die Nutzung der Suchmaschinentechnologie im Rahmen von Bibliothekskatalogen (z.B. OPACplus der Bayerischen Staatsbibliothek) bzw. Verbundkatalogen sowie das Entstehen neuerer Suchmaschinen für akademische Materialien wie z.B. der [Bielefeld Academic Search Engine \(BASE\)](#) zeigen eine auch für die Virtuellen Fachbibliotheken relevante neue Entwicklungsrichtung auf.

Die Integration von Suchmaschinen in die Metasuche virtueller Fachbibliotheken bringt mehrere Vorteile mit sich, insbesondere eine erhöhte Geschwindigkeit in der Trefferanzeige, die Stabilität der Suche durch die lokale Datenhaltung, Möglichkeiten der Datenanreicherung und der Indexierung von Volltexten (z.B. Inhaltsverzeichnissen), Verbalisierung von Abkürzungen, Ranking der Daten, Nachträgliche Datenfilterung über Navigatoren, etc..

Darüber hinaus bieten sich durch die Indexierung weitere Möglichkeiten, neue Services anzubieten wie z.B. RSS-Feeds für einzelne Suchanfragen.

Einige der neuen Features sollen hier kurz am Beispiel der vor kurzem neu gestarteten Virtuellen Fachbibliothek b2i der Bayerischen Staatsbibliothek vorgestellt werden. Einiges davon finden Sie aber auch in der ebenfalls vor nicht allzu langer Zeit online gegangenen Virtuellen Fachbibliothek medien bühne film, oder auch in den Virtuellen Fachbibliotheken Propylaeum und ViFaRom:

- **Datenfilterung:**

Eine der hervorstechenden Möglichkeiten der Metasuche mit Suchmaschinentechologie ist die Möglichkeit, die Daten nach bestimmten Kriterien zu filtern, so z.B. nach Erscheinungsjahr oder auch Publikationsform. Damit wird es möglich, sehr allgemein gehaltene Eingaben, die bei einer sehr vagen und weiträumigen Herangehensweise an ein Thema oft dominieren, Stück für Stück zu konkretisieren.

- **Datenanreicherung**

Der Einsatz von Filtern ist auf Anhieb natürlich nicht bei jedem Datenbestand möglich, da evtl. keine entsprechenden Daten festgehalten sind. Im Falle der Wolfenbütteler Bibliographie WBB wurde deshalb ein Sprachanalysetool eingesetzt, um die Sprachbezeichnungen in den Datensätzen zu ergänzen und im Anschluss zu verbalisieren. Die Anreicherung der Daten über PDF/OCR-Beigaben ist mit der Möglichkeit der Indexierung von integrierten Volltexten ebenfalls möglich.

- **RSS-Feeds zu Suchanfragen:**

Ein weiterer Bereich, der mit Hilfe der Datenindexierung bei b2i realisiert werden konnte ist die Überführung von Suchanfragen in RSS-Feeds, so dass man sich über Neuerwerbungen/Neuaufnahmen zu bestimmten Teilgebieten ohne eine wiederholte Eingabe der entsprechenden Abfrage auf dem Laufenden halten kann.

- **Mehrsprachige Tastatur für die Eingabe**

Ein Feature, das bei b2i noch nicht eingesetzt wird, jedoch im Portal Propylaeum oder auch in der ViFaOst ist die Einbindung einer mehrsprachigen Tastatur für verschiedene weitere Schriften, wie z.B. Griechisch oder Kyrillisch. Damit kann die Suche in originalschriftlichen wie transliterierten Datenbeständen auf der Basis von Unicode weitgehend zusammengeführt werden.

Eines der großen Grundprobleme einer Metasuche bleibt jedoch auch im Rahmen einer Indexierung bestehen: die oft höchst unterschiedlichen Datenstrukturen und Datenqualitäten der Datenquellen. Im Rahmen der Indexierung bei b2i wird die Vereinheitlichung über ein Mapping der Originalstrukturen in ein XML-Schema, das sogenannte VIFA-Application-Profile erzielt, ähnlich dem Vorgehen bei VASCODA.

Die Datenheterogenität zeigt sich nicht so sehr auf der Anzeigenseite, sondern verringert immer mehr die Anzahl der möglichen Suchfelder. Im Falle von b2i verbleiben neben der Freien Suche die bekannten Felder Autor, Titel, Jahr, ISBN und Schlagwort. Dies unterscheidet die Suche nicht von einem auf Z39.50-Anbindungen basierendem System. Durch die bereits angesprochene Datenanreicherung kann evtl. in Zukunft in bestimmten Bereichen ein Mehrwert erzielt werden. Der spezialisierte Nutzer sollte daher auf die Möglichkeiten, gezielt in der Originaldatenbank zu recherchieren, immer hingewiesen werden, um mögliche Nachteile der Metasuche bei Spezialrecherchen auszugleichen.

### **3. Perspektiven**

Die Virtuellen Fachbibliotheken werden künftig beim Information Retrieval verstärkt an den neuen Suchmöglichkeiten der großen allgemeinen Suchmaschinen und den dort präsentierten neuen Recherchemöglichkeiten gemessen. Im Rahmen dieser Erweiterungen der allgemeinen Suchmaschinen wird mittelfristig die Auswahl oder Anzahl der Datenquellen kein ausschließliches Alleinstellungsmerkmal für die Virtuellen Fachbibliotheken mehr sein. Das Information Retrieval muss im Kontext Virtueller Fachbibliotheken modernisiert werden, wobei sich die Integration von Suchmaschinentechnologie als gangbarer Weg anbietet. Einige Portale wie das Wissenschaftsportal b2i haben diesen Weg bereits beschritten.

Insgesamt erscheint es jedoch notwendig, ein Information Retrieval in den Virtuellen Fachbibliotheken stärker auf mehrere Kanäle zu verteilen, so z.B. durch die Integration von Navigations- und Schulungsoberflächen wie z.B. [LOTSE](#), das ja bei manchen Portalen bereits fester Bestandteil ist und einen ganz bestimmten Suchkanal öffnet.

Daneben sollte das Browsing in den Datenbeständen mehr im Vordergrund stehen. Im Rahmen von b2i ist z.B. geplant, einen auf DDC basierenden klassifikatorischen Zugang einzubauen. Hinter diesem Zugang wird ein Mapping aller vorhandenen Klassifikationen in den Datenbeständen stehen.

Die Möglichkeit, Recherchen auf der Basis von Begriffsrelationen zu nutzen, wird sicher in den nächsten Jahren eine große Rolle spielen, ebenso wie eine möglichst sprachübergreifende Suche, die durch das Hinterlegen von Wörterbüchern ermöglicht werden kann. So könnte z.B.

eine Implementierung der Ergebnisse z.B. aus dem von der FH Köln und der DNB durchgeführten [DFG-Projekts CrissCross](#) einen erheblichen Mehrwert erbringen. Zudem sollten weitere Elemente der Interaktion mit dem Nutzer für intelligente Suchanfragen sollten aufgrund steigender Datenmengen in den Virtuellen Fachbibliotheken stärker ins Blickfeld gerückt werden. In diesem Zusammenhang sollte natürlich darauf verwiesen werden, den Nutzer noch stärker in die Informationsgewinnung mit einzubeziehen, im Sinne der Möglichkeiten, die unter dem Begriff „User generated content“ zusammengefasst werden. Als weiteren Aspekt einer verbesserten Recherche kann eine erweiterte Visualisierung (z.B. in Richtung Dreidimensionalität) angeführt werden. Gleichzeitig sollte ein die Visualisierung ergänzender Zugang über eine Einbindung von Audio-Elementen Impulse geben. Insgesamt ist im Rahmen des Information Retrieval durch eine anzustrebende Mehrdimensionalität des Zugangs mit einer Erhöhung der Komplexität zu rechnen. Diese Komplexitätserhöhung muss zur Sicherstellung neuer und erweiterter Recherchefunktionalitäten nutzerseitig von einer Komplexitätsreduzierung begleitet sein.