

# Kolibri – Entwicklung von Modellen und Werkzeugen für kooperative digitale Handbibliotheken

## Kolibri – Development of Models and Tools for Cooperative Digital Reference Libraries

Florian Matthes, Joachim W. Schmidt, Ulrike Steffens, Technische Universität Hamburg-Harburg

---

Dieser Beitrag beschreibt die Arbeitsergebnisse des Kolibri-Projektes, das sich die Unterstützung kooperativer Prozesse bei der Nutzung digitaler Handbibliotheken zur Aufgabe gemacht hat. Nach einer Diskussion der für dieses Anwendungsfeld spezifischen Anforderungen wird ein entsprechendes generisches Modell für digitale Handbibliotheken vorgestellt und zwei auf diesem Modell aufbauende Werkzeuge beschrieben.

This paper describes the results of the Kolibri project, which aims at supporting cooperative processes within digital reference libraries. After a discussion of the requirements specific for this application area, a corresponding generic model for digital reference libraries is introduced and two library tools relying on this model are described.

---

### 1 Einleitung

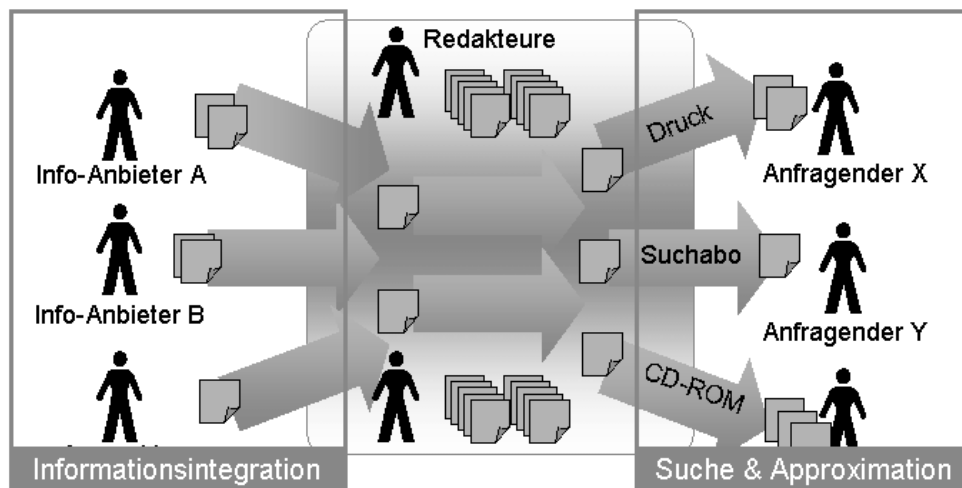
Eine Handbibliothek stellt ihrem Benutzer eine tief erschlossene und gut handhabbare Sammlung digitaler Dokumente zur Verfügung, die zumeist in einem kooperativen Kontext entstehen und verwendet werden, um die Arbeit innerhalb einer Organisation zu vereinfachen und voranzubringen [8]. Die Entwicklung und Bewertung von Modellen und Werkzeugen zur Unterstützung solcher kooperativen Handbibliotheken ist das Hauptziel des Projektes Kolibri am Arbeitsbereich Softwaresysteme der Technischen Universität Hamburg-Harburg.

Im Projektverlauf wurden zunächst die besonderen Gegebenheiten und Anforderungen bei der Nutzung digitaler Handbibliotheken identifiziert (vgl. Abschnitt 2). Auf dieser Grundlage wurde ein generisches Modell entwickelt, das sowohl die statische als auch die dynamische Semantik einer digitalen Handbibliothek repräsentiert (vgl. Abschnitt 3). Dieses Modell bildet die Basis für eine Reihe von Bibliothekswerkzeugen, von denen zwei während des Projektes implementiert wurden (vgl. Abschnitt 4). In Phase I des Projektes wurde deutlich, dass die Nutzung einer Handbibliothek mit einer Reihe von Lernprozessen sowohl von Seiten des Informations-

anfragenden als auch von Seiten des Informationsanbieters verbunden ist. Die Untersuchung und Unterstützung dieser Prozesse bestimmt die Ausrichtung der zweiten Projektphase von Kolibri (vgl. Abschnitt 5).

### 2 Kooperative Nutzung digitaler Handbibliotheken

Digitale Handbibliotheken verbinden *Informationsanbieter* mit *Informationsanfragenden*, wobei ein Benutzer beide Rollen einnehmen kann (vgl. Bild 1). Charakteristisch ist hierbei, dass es sich bei der entstehenden Kommunikation nicht um eine 1:n-Massenkommunikation handelt. Ziel ist es vielmehr, so zwischen Anbieter und Anfragendem zu vermitteln, dass beide Seiten voneinander lernen und dadurch zu einer bestmöglichen Einigung zwischen Informationsbedürfnissen und Informationsangebot kommen. Aus dieser Motivation heraus erwächst ein Kooperationszenario, in dem Anbieter und Anfragende durch eine dritte Partei, den sog. Vermittler oder *Redakteur*, verbunden sind. Der Redakteur übernimmt hierbei zwei Aufgaben. Zum einen schafft er eine einheitliche, sowohl für den Anbieter als auch für



**Bild 1:** Kooperative Nutzung digitaler Handbibliotheken.

den Anfragenden verständliche Begriffswelt, in der beide miteinander interagieren können. Zum anderen unterstützt er beide Seiten jeweils mit dem Feedback, das sie benötigen, um in ihrem Kommunikationsverhalten auf das jeweilige Gegenüber sinnvoll reagieren zu können. Dieses Szenario entspricht den Prozessen einer digitalen Handbibliothek in besonderer Weise, da anders als in großen allgemeinen digitalen Bibliotheken Informationsanfragende und Informationsanbieter durch gemeinsame Ziele, Vorgaben o. ä. verbunden sind. Hieraus ergibt sich eine langfristige Beziehung zwischen beiden Seiten, die nur durch lernenden personalisierten Umgang miteinander effizient funktionieren kann.

Will man dieses Kooperationsszenario unterstützen, so ergeben sich die folgenden Anforderungen an die zugrundeliegenden Informationen und den Umgang mit ihnen.

- **Strukturierte Information:** Eine angemessene Auswertung und die Verständlichkeit der Informationen erfordern eine Strukturierung der Information.
- **Anpassbare Information:** Anbieter können den Wünschen der Anfragenden am besten entsprechen, wenn sie ihr jeweiliges Angebot, im Falle der Handbibliothek also die Information, an die konkreten Bedürfnisse anpassen können. Informationen sollten daher keinem starren Schema unterliegen, sondern auf viele unterschiedliche Arten darstellbar sein.
- **Flexibler Umgang:** Anfragende der Handbibliothek, ebenso wie Kunden in einem Geschäft, wissen nicht immer von vornherein, welche Information sie genau benötigen. Sie wollen ausdrücken, was sie in etwa suchen, dieses oder jenes anschauen und dadurch ihren Bedarf schrittweise konkretisieren, nicht ohne während des gesamten Vorgangs auf Beratung zurückgreifen zu können. Die vorhandene Information sollte so beschaffen sein, dass diese vage Herangehensweise möglich wird.

Der Begriff einer so beschaffenen digitalen Handbibliothek lässt sich nicht nur auf Dokumente in Biblio-

thekskatalogen anwenden, sondern ist auch für ähnliche Kollektionen anderer Organisationen gültig, wie z. B. für Produkt- oder Dienstleistungskataloge, Software-Bibliotheken oder Störfallbeseitigungsanleitungen.

Die folgenden Beispiel-Anfragen an solche domänenspezifischen Handbibliotheken geben einen konkreten Eindruck der aufgeführten Anforderungen, die sich in den o. g. Anwendungsbereichen stellen:

*Digitale Bibliotheken:* „Eine TrueColor Abbildung zum Thema Mobilität, die auf DIN A1 skalierbar ist, zu Lizenzkosten von ca. 4000 DM mit einer Bearbeitungszeit von weniger als 48 h.“

*Produktkataloge (E-Commerce):* „Der billigste Videorecorder mit Stereo-Ton und Jog Shuttle.“

*Anzeigensuche:* „Ein guter Chinese in der Stadtmitte, der Master-Card akzeptiert.“

*Dienstleistungskataloge (E-Brokerage):* „Eine günstige Fahrzeugversicherung für einen BMW, die mehr als zwei Fahrer erlaubt.“

Digitale Handbibliotheken haben demnach folgende Aspekte gemein: Sie enthalten *heterogene, semistrukturierte Dokumente* [1], die häufig von verschiedenen Informationsanbietern stammen. Informationsanfragende wissen, insbesondere wenn sie auf große Kollektionen zugreifen, nicht, welche Werte in den Dokumenten tatsächlich enthalten sind. Wie sich bereits in den obigen Beispielen zeigt, drücken sie daher ihren Informationsbedarf in Form von *approximativen Anfragen* aus, die auf *unscharfen Kategorien* beruhen.

### 3 Ein generisches Modell für digitale Handbibliotheken

Die geschilderten Gegebenheiten digitaler Handbibliotheken finden ihre Berücksichtigung in den durch das Kolibri-Projekt erstellten Werkzeugen, die aus diesem Grunde auf einem generischen Modell für interaktive Produkt- und Dienstleistungskataloge basieren. Die Einträge eines solchen Kataloges wer-

den dabei als Dokumente einer digitalen Handbibliothek aufgefasst. Das verwendete Modell macht von zwei sich ergänzenden Formalismen Gebrauch:

Ein semistrukturiertes Dokument besteht aus einer beliebigen Anzahl von Eigenschaften in Form von Attribut-Wert-Paaren, die dynamisch anpassbar sind und keinem starren Datenschema unterliegen. Die Werte sind dabei nicht auf elementare Datentypen beschränkt, sondern können auch durch Instanzen diskreter oder kontinuierlicher Domänen gegeben sein. Semistrukturierte Dokumente bilden das Kernkonzept der statischen Semantik des Kolibri-Modells.

Die dynamische Semantik des Modells beruht auf der Theorie der Fuzzy-Mengen. Das Ziel, das hier beim Entwurf des Modells und der darauf basierenden Anwendungen verfolgt wurde, war es, zum einen die größtmögliche Freiheit in der Formulierung und Kombination von Anfrageprädikaten zuzulassen, jedoch zum anderen den gemeinsamen Nenner möglichst gering zu halten, der notwendig ist, um sinnvolle Benutzungsschnittstellen, eine effiziente Anfrageauswertung, eine angemessene Verwaltung von Benutzerprofilen und Persistenzanforderungen etc. zu gewährleisten. Aus diesem Grunde sind Anfragen in Kolibri auf eine Boolesche Kombination von atomaren Fuzzy-Prädikaten für die Dokumenteigenschaften beschränkt. Konkret sind hier  $n$ -äre gewichtete Disjunktionen aus  $n$ -ären gewichteten Konjunktionen erlaubt [4].

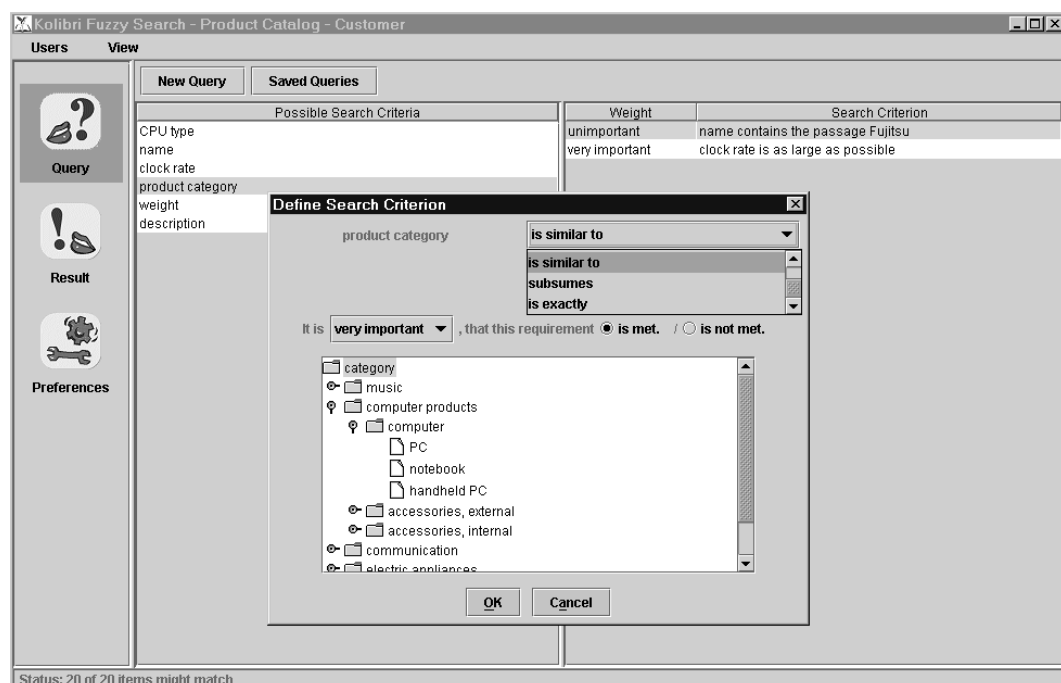
## 4 Kolibri-Werkzeuge

Das beschriebene Modell ist unter Verwendung der Programmiersprache Java implementiert worden und bietet unter anderem eine stabile Grundlage für die

Benutzungsschnittstelle von *PIA* (*Personal Information Assistant*), einem interaktiven Werkzeug, das Informationsanfragenden die Möglichkeit bietet, anhand einer Kombination von numerischen, taxonomischen [6] und Volltext-Prädikaten semistrukturierte Dokumente in einer digitalen Handbibliothek aufzufinden.

Bild 2 vermittelt einen Einblick in die Arbeitsweise von PIA: Bereits während er seine Anfrage formuliert, werden dem Informationsanfragenden Informationen über die Kataloginhalte angeboten. In der linken Tabelle erscheint eine Auflistung aller momentan im Katalog verfügbaren Attribute. Dadurch erhält der Anfragende einen ersten Überblick über die möglichen Anfrageformulierungen. Der aktuelle Zustand der Anfrage ist aus dem rechten Teil des Fensters ersichtlich, wo die einzelnen Fuzzy-Prädikate mit Parametern und Gewichtungen aufgeführt sind. Die Auswertung einer Anfrage kann über die entsprechende Schaltfläche auf der linken Seite des Fensters ausgelöst werden. Sie berechnet eine Liste von Dokumenten und zeigt diese geordnet nach ihrer Relevanz für die Anfrage an. Jedes Dokument ist mit einem Hyperlink verbunden, der zu einer detaillierten Beschreibung führt. Auf diese Art kann der Anfragende auf direktem Weg Entscheidungen über die Qualität der Ergebnisdokumente treffen.

Neben der Anfrageschnittstelle bietet PIA auch einen Zugang für Informationsanbieter, über den diese ihre angebotenen Dokumente in den Katalog laden und sie dort pflegen können, sowie eine Schnittstelle für Redakteure, über die diese die Struktur eines Kataloges bearbeiten können, die sich z. B. über die vorhandenen Attribute und deren Domänen dar-



**Bild 2:** Erstellen einer Anfrage in PIA.

Felder	1. Eintrag	2. Eintrag	neuer Eintrag
BibTeX_Key	-> Buck-Emden96	-> Buck-Emden96	Buck-Emden96
BibTeX_Typ	-> book	-> book	book
address	-> Bonn Paris	-> Bonn Paris	Bonn Paris
author	-> Rüdiger Buck-Emden and Jürgen Galimow	-> Rüdiger Buck-Emden and Jürgen Galimow	Rüdiger Buck-Emden and
edition	-> dritte	-> dritte	dritte
isbn	-> 3-89319-764-8	-> 3-89319-764-8	3-89319-764-8
publisher	-> Addison Wesley	-> Addison Wesley Verlag	
title	-> Die Client/Server-Technologie des Systems R/3	-> Die Client/Server-Technologie des Systems R/3	Die Client/Server-Tech
year	-> 1996	-> 1996	1996

**Bild 3:** Vergleich von Katalogeinträgen.

stellt. [7] enthält eine genauere Beschreibung dieser Schnittstellen, die unter [2] auch als Online-Demonstration betrachtet werden können.

Das vorgestellte generische Modell ist jedoch nicht nur die Grundlage von Anfrageschnittstellen wie PIA, sondern bietet darüber hinaus das Potential für weitere Werkzeuge im Bereich der digitalen Handbibliotheken, wie z. B. das *RMT (Reference Merging Tool)*, das ebenfalls auf diesem Modell basiert. Auch RMT ist ein interaktives Werkzeug, das zwei unabhängige Handbibliotheken miteinander vergleicht und daraus eine dritte Bibliothek generiert, die Dokumente der beiden Ausgangsbestände umfasst. Diese dritte Bibliothek kann sich dann unabhängig von den zugrundeliegenden Kollektionen weiterentwickeln.

RMT wurde im Rahmen des Kolibri-Projektes zunächst als Prototyp entwickelt, dessen Benutzungsschnittstelle in Bild 3 zu sehen ist: Hier sind zwei Dokumente, die aus unterschiedlichen autonomen BibTeX-Dateien stammen und von RMT als mögliche Duplikate identifiziert worden sind, einander gegenübergestellt. RMT hat außerdem automatisch einen Vorschlag generiert, wie die beiden Dokumente zu einem dritten verbunden werden können, das dann in eine gemeinsame Bibliothek eingefügt wird. Der Benutzer kann das neue Dokument akzeptieren, verwerfen oder die angezeigten Attribut-Wert-Paare einzeln bearbeiten.

## 5 Ausblick

Als Ergebnis der theoretischen, praktischen und industriellen Auseinandersetzung mit den in Kolibri-Phase I entwickelten Techniken, insbesondere hinsichtlich des Anfragewerkzeuges PIA, wurde deutlich, dass die im Rahmen des dort ablaufenden interaktiven Prozesses anfallenden Informationen einen wertvollen Feedback liefern, der u. U. in einem strukturierten und durch Computer unterstützbaren Lern-

prozess innerhalb der Organisation genutzt werden kann (Organizational Learning [9]). Es entstehen somit aus der rohen Datenmenge zahlreicher Interaktionsprotokolle nützliche kunden- und marktbezogene Informationen, die als Handlungsgrundlage für einen kundenorientierten Verbesserungsprozess der der Handbibliothek zugrundeliegenden Artefakte und Dienste verwendet werden können.

Konkret sollen daher in Kolibri-Phase II Informatik-Modelle, Techniken und Systemkomponenten evaluiert und entwickelt werden, mit deren Hilfe die Suchparameter der Nutzer, ihre Aktionen, sowie ihre expliziten Fragen oder Annotationen zu den Suchergebnissen analysiert, aggregiert und in strukturierter Form für die zielgerichtete Verbesserung der Kollektionen und der darauf aufbauenden Dienste verwendet werden können. Das Vorgehen orientiert sich dabei zum einen stark an den Ergebnissen von Kolibri-Phase I und berücksichtigt zum anderen aktuelle Forschungsergebnisse angrenzender Teildisziplinen der Informatik, wie z. B. des fallbasierten Schließens [5], des Information Retrieval [3] oder personalisierbarer Informationsdienste [10].

## References

- [1] *S. Abiteboul*: Querying Semi-Structured Data. In F. Afrati and P. Kolaitis, editors, Proceedings ICDT '97, 6th International Conference, Delphi, Greece, Lecture Notes in Computer Science, pages 1–18. Springer-Verlag, 1997.
- [2] Technische Universität Hamburg-Harburg Arbeitsbereich Softwaresysteme. Homepage des Kolibri-Projektes, 2000. <http://www.sts.tu-harburg.de/projects/Kolibri/entry.html>.
- [3] *D. Haines* and *W. B. Croft*: Relevance Feedback and Inference Networks. In Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 2–11, 1993.
- [4] *R. Fagin*: Fuzzy Queries in Multimedia Database Systems. In J. Paredaens, editor, Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Seattle, Washington, pages 1–10. ACM Press, 1998.

- [5] *J. Kolodner*: Case-based reasoning. Morgan Kaufmann, 1993.
- [6] *H. L. Larsen* and *R. R. Yager*: The Use of Fuzzy Relational Thesauri for Classificatory Problem Solving in Information Retrieval and Expert Systems. IEEE Transactions on Systems, Man, and Cybernetics, 23(1): 31–41, 1993.
- [7] *F. Matthes* and *U. Steffens*: PIA – A Generic Model and System for Interactive Product and Service Catalogs. In S. Abiteboul and A. Vercoustre, editors, Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries, Paris, ECDL'99, September 1999.
- [8] *J. W. Schmidt*, *G. Schröder*, *C. Niederée* and *F. Matthes*: Linguistic and Architectural Requirements for Personalized Digital Libraries. International Journal of Digital Libraries, 1(1), 1997.
- [9] *P. M. Senge*: The Fifth Discipline, The Art and Practice of the Learning Organization. Currency Doubleday, New York, NY, 1990.
- [10] *K. Tochtermann*, *D. Hicks* and *A. Kussmaul*: Support for Customization and Personalization on the Web. In Proceedings of Webnet99, World Conference on the WWW and Internet, 1999.



**Prof. Dr. Florian Matthes** ist Professor für Informatik am Arbeitsbereich Softwaresysteme der Technischen Universität Hamburg-Harburg. Seine Forschungsinteressen liegen schwerpunktmäßig im Bereich der Softwarearchitekturen und Programmiermodelle für kooperative Informationssysteme.

Adresse: Technische Universität Hamburg-Harburg, Arbeitsbereich 4-02, Softwaresysteme, D-21071 Hamburg, E-Mail: f.matthes@tuhh.de



**Prof. Dr. Joachim W. Schmidt** ist Professor für Informatik und Leiter des Arbeitsbereichs Softwaresysteme der Technischen Universität Hamburg-Harburg. Seine Forschungsinteressen liegen schwerpunktmäßig in den Bereichen Datenbanken, Informationssysteme, Programmiersprachen und transaktionsorientierte Systeme.

Adresse: Technische Universität Hamburg-Harburg, Arbeitsbereich 4-02, Softwaresysteme, D-21071 Hamburg, E-Mail: j.w.schmidt@tuhh.de



**Dipl. Inform. Ulrike Steffens** ist wissenschaftliche Mitarbeiterin am Arbeitsbereich Softwaresysteme der Technischen Universität Hamburg-Harburg. Ihre Forschungsinteressen liegen schwerpunktmäßig in der Konzeption benutzergerechter Dienste für kooperative digitale Handbibliotheken.

Adresse: Technische Universität Hamburg-Harburg, Arbeitsbereich 4-02, Softwaresysteme, D-21071 Hamburg, E-Mail: ul.steffens@tuhh.de