62. Jahrgang, Nr. 1 – Januar/Februar 2011 – NADOAW 62 (1) 1-64 – ISSN 1434-4653 – D 11721

1/2011

Information Wissenschaft & Praxis

Bradfordizing als Re-Ranking-Ansatz

Kriterienkatalog zur Usability-Evaluation von Bibliothekswebsites

Erfolgsfaktoren kirchlicher Community-Angebote im Social Web

Benutzerforschung anhand von Log-Dateien

Interview mit Rafael Capurro zu Fragen der Informationswissenschaft

Einsatz von XML für bibliographische Metadaten

Tagungsbericht eSciDoc Days 2010

Aktuelles aus dem Patentbereich: AGM 2010

Programm DGI-Praxistage 2011

iwp Herausgeber: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e. V. DGI

Entscheidend mehr Wissen.

Für Sie als Informationsprofi ist es entscheidend **immer einen Schritt voraus** zu sein.

Sie müssen **zuverlässig und schnell** an Qualitätsinformationen gelangen. GENIOS steht Ihnen hierbei mit maßgeschneiderten Angeboten und **hochwertigen Informationen** aus Wirtschaft, Management, Unternehmen, Politik und Wissenschaft zur Seite.

Sichern Sie sich Ihren Informationsvorsprung auf www.genios.de

GENIOS

GBI-Genios Deutsche Wirtschaftsdatenbank GmbH

Ein Unternehmen der Frankfurter Allgemeinen Zeitung und der Verlagsgruppe Handelsblatt

Inhalt 1/2011

INFORMATIONSWISSENSCHAFT

3 Philipp *Mayr*
Bradfordizing als Re-Ranking-Ansatz in Literaturinformationssystemen

11 Thomas *Weinhold*, Sonja *Öttl*, Bernard *Bekavac*
BibEval – Ein webbasierter Kriterienkatalog zur Usability-Evaluation von Bibliothekswebsites

19 Natascha *Pauls*, Joachim *Griesbaum* und Thomas *Mandl*
Erfolgsfaktoren kirchlicher Community-Angebote im Social Web: Eine Analyse des Wikis „Evangelisch in Niedersachsen“

29 Thomas *Mandl* , Julia Maria *Schulz*, Nadine *Mahrholz* und Katrin *Werner*
Benutzerforschung anhand von Log-Dateien: Chancen, Grenzen und aktuelle Trends

37 Linda *Treude*
Information, Zeichen, Kompetenz. Fragen an Rafael Capurro zu aktuellen und grundsätzlichen Fragen der Informationswissenschaft

42 **PERSONALIE**

Prof. Dr. Jiri Peter Panyr †

INFORMATIONSPRAXIS

43 Dimitri *Busch*
Ein XML-basierter Ansatz für Austausch, Speicherung und Retrieval von bibliographischen Metadaten im Bereich Planen und Bauen

TAGUNGSBERICHTE

49 Vera *Münch*
eSciDoc Community sorgt für nachhaltige Nutzung der Projektergebnisse

53 Dieter *Geiß*
Aus der Praxis der Patentinformation. Die Entwicklung der elektronischen Medien und Dienstleistungen bei den Patentbehörden und Internetprovidern

BUCHBESPRECHUNGEN

36 Isaev, G.N.: Informacionnye sistemy v ėkonomike : učebnik (Informationssysteme in der Wirtschaft. Lehrbuch) Moskau : Omega, 2009. 462 S. : Tab., graph. Darst. In kyrill. Schr., russ. ISBN 978-5-370-00855-9 (Freytag, J.)

36 Isaev, G.N.: Modelirovanie informacionnych resursov: teorija i rešenie zadač : učebnoe posobie (Modellierung von Informationsquellen. Handreichung) Moskau : Al'fa-M . Infra-M, 2010. 223 S. : Tab., graph. Darst. In kyrill. Schr., russ. ISBN 978-5-98281-211-7 und 978-5-16-004232-9 (Freytag, J.)

48 Spink, A.: Information Behavior: An Evolutionary. Heidelberg [u.a.]: Springer, 2010. XXVII, 85 S. (Information Science and Knowledge Management. Vol.16) ISBN 978-3-642-11496-0 (Lewandowski, D.)

INFORMATIONEN

10 Neue Umschriftnormen aus dem DIN

36 Aufruf zur Kandidatur für den DGI-Vorstand

61 Qualifiziert in die Zukunft

61 GBI-Genios präsentiert das neue GENIOS 2011

64 CeBIT-Premiere des Elektronischen Lesesaals „MyBib eRoom“

INTERVIEW

63 Ronald *Kaiser* und Wolfgang *Ratzek*
Internetnutzung in Freiheit und Fairness. Fragen an Dorothee Bär

64 **IMPRESSUM**

U3 **TERMINKALENDER**

Die DGI-Geschäftsstelle ist zum 1. Dezember 2010 in neue Räume in die Nähe des Frankfurter Hauptbahnhofs Richtung Mainufer umgezogen. Die neue Hausanschrift lautet:
Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V.
Windmühlstraße 3 · 60329 Frankfurt am Main

Missing Link

Westerstrasse 114-116 | 28199 Bremen

Tel.: [0421] 504348 | Fax: [0421] 504316

Erwerbungspartner, mit denen Sie rechnen können

Flexibel

Erfahren

Innovativ

Konditionsstark

Serviceorientiert

Engagiert

Klar

www.missing-link.de

info@missing-link.de

Internationale Versandbuchhandlung

Bradfordizing als Re-Ranking-Ansatz in Literaturinformationssystemen

Philipp Mayr, Bonn

In diesem Artikel wird ein Re-Ranking-Ansatz für Suchsysteme vorgestellt, der die Recherche nach wissenschaftlicher Literatur messbar verbessern kann. Das nicht-textorientierte Rankingverfahren Bradfordizing wird eingeführt und anschließend im empirischen Teil des Artikels bzgl. der Effektivität für typische fachbezogene Recherche-Topics evaluiert. Dem Bradford Law of Scattering (BLS), auf dem Bradfordizing basiert, liegt zugrunde, dass sich die Literatur zu einem beliebigen Fachgebiet bzw. -thema in Zonen unterschiedlicher Dokumentenkonzentration verteilt. Dem Kernbereich mit hoher Konzentration der Literatur folgen Bereiche mit mittlerer und geringer Konzentration. Bradfordizing sortiert bzw. rankt eine Dokumentmenge damit nach den sogenannten Kernzeitschriften. Der Retrievaltest mit 164 intellektuell bewerteten Fragestellungen in Fachdatenbanken aus den Bereichen Sozial- und Politikwissenschaften, Wirtschaftswissenschaften, Psychologie und Medizin zeigt, dass die Dokumente der Kernzeitschriften signifikant häufiger relevant bewertet werden als Dokumente der zweiten Dokumentzone bzw. den Peripherie-Zeitschriften. Die Implementierung von Bradfordizing und weiteren Re-Rankingverfahren liefert unmittelbare Mehrwerte für den Nutzer.

Bradfordizing as re-ranking approach in abstracting and indexing databases

This article proposes a re-ranking approach for search systems which intents to improve the retrieval for scientific literature on a significant level. The non text-oriented ranking approach Bradfordizing will be introduced and then the effectiveness for typical subject specific retrieval topics will be evaluated. The underlying Bradford Law of Scattering (BLS) states that the literature on any subject or domain is distributed in different zones with typical concentrations in each zone. The core zone or core with a high concentration of literature is followed by zones with decreasing and low concentrations of papers. Bradfordizing sorts or ranks a document space by the so called core journals. The retrieval test with 164 intellectually assessed topics from different abstracting and indexing databases in the social and political sciences, economics, psychology and medical science shows that documents in the core journals are significantly more often relevant than documents from zone 2 journals and peripheral journals. The paper shows that the implementation of Bradfordizing and other re-ranking approaches generates an immediate value-added for users.

1 Einleitung

Nutzer sehen sich je nach Art und Umfang der Literaturrecherche immer häufiger in der Situation, auf unterschiedliche Suchsysteme zugreifen zu müssen, um ihren Informationsbedarf zu befriedigen. Inzwischen führt die alleinige Nutzung z. B. eines Online-Katalogs (OPAC) der lokalen Bibliothek genauso wenig zu vollständigen Ergebnissen wie die Beschränkung bei der Suche auf eine einzelne Fachdatenbank. Vor allem wissenschaftliche Nutzer erwarten aus diesem Grund heute möglichst die Bündelung und Konzentration fachlich relevanter und qualitativer Information an einer Stelle.

Des Weiteren fordern Nutzer für datenbankübergreifende Literaturrecherchen mit großen Dokumentmengen einen möglichst hohen Anteil an relevanten und qualitativ hochwertigen Dokumenten in den vorderen Trefferergebnissen. Insbesondere die Reihenfolge und Struktur der gelisteten Ergebnisse (Ranking) spielt, neben dem direkten Volltextzugriff auf die Dokumente, inzwischen eine entscheidende Rolle beim Design von Suchsystemen. Abgegrenzt wird Ranking oder Relevance Ranking von sog. Sortierungen z. B. nach dem Erscheinungsjahr der Publikation, obwohl hier die Grenze zu „nach inhaltlicher Relevanz“ gerankten Listen konzeptuell nicht sauber zu ziehen ist.

Das Ranking von Dokumenten führt letztlich dazu, dass sich die Benutzer solcher Suchsysteme fokussiert mit den oberen Treffermengen (top-gerankt) eines Suchergebnisses beschäftigen. Der mittlere und untere Bereich eines Suchergebnisses wird häufig nicht mehr in Betracht bezogen. Bei großen Treffermengen ist eine vollständige Rezeption eines Suchergebnisses ohnehin nicht mehr möglich. Häufig ist die Konsequenz, dass Nutzer aus Bequemlichkeitsgründen ausschließlich mit kommerziellen Internetsuchmaschinen wie Google oder Yahoo recherchieren und die fachwissenschaftlich erschlossenen Literaturdatenbanken nicht mehr konsultiert werden.

Auf Grund der Vielzahl an relevanten und verfügbaren Informationsquellen ist es daher notwendig, Kernbereiche in den Suchräumen zu identifizieren und diese anschließend dem Nutzer hervorgehoben zu präsentieren. Die Kernbereichsbildung kann zum einen über unscharfe textstatistische Ranking-Verfahren (z. B. term frequency - inverse document frequency (tf-idf)) erfolgen, zum anderen aber auf Grundlage der den Suchräumen inhärenten Strukturen.

Das in diesem Artikel evaluierte Verfahren Bradfordizing setzt auf einer typischen bibliometrischen Regelhaftigkeit thematischer Dokumentenräume auf und nutzt diese Strukturmerkmale zum Ranking von Dokumenten (siehe dazu Mayr, 2010).

Motivation

Neben der Abgrenzung der Suchräume und Dokumenttypen erhält die Zugänglichkeit und Relevanz der Dokumente eine entscheidende Bedeutung für den Benutzer. Die neueren technologischen Entwicklungen des Web Information Retrieval (IR), wie sie die kommerziellen Internetsuchmaschinen implementieren, werten insbesondere frei zugängliche Dokumente mit ihrer gesamten Text- und Linkinformation automatisch aus. Diese Verfahren sind vor allem deshalb erfolgreich, weil sie Ergebnislisten nach Relevanz gerankt darstellen, einfach und

schnell zu recherchieren sind und direkt auf die Informationsquellen verweisen. Die qualitativen Verfahren der traditionellen Fachinformationsanbieter (z. B. Web of Science) hingegen zeigen genau bei diesen Punkten (Ranking, Einfachheit und Volltextzugriff) Schwächen, überzeugen aber vor allem durch ihre Stringenz, in diesem Fall die selektive Aufnahme von qualitätsgeprüften Dokumenten in das System und die inhaltliche Erschließung der Dokumente. Aufgrund der Beschaffenheit der Dokumente in klassischen Informationssystemen wie z. B. bibliographischen Fachdatenbanken greifen die Ranking-Verfahren des Web IR genauso wenig wie die herkömmlichen textstatistischen Information Retrieval-Verfahren. Alternative, nicht-textorientierte Ansätze, wie z. B. der Einsatz von Autorenzentralitätsmaßen beim Ranking (Mutschke, 2004), aber auch das in diesem Artikel vorgestellte Verfahren, zeigen, dass die spezifische Struktur der Literaturnachweise und die Qualität der Daten (i. d. R. Metadaten) für alternative Mehrwertdienste gewinnbringend eingesetzt werden können.

Das Mehrwertverfahren Bradfordizing zielt auf die Neu- bzw. Reorganisation eines vorliegenden Trefferergebnisses (sog. Re-Ranking) und setzt damit konzeptuell nach der Suche bzw. Suchraumerweiterung z. B. durch Heterogenitätsbehandlung (vgl. Krause et al, 2003; Krause, 2004; Mayr & Petras, 2008) an. In diesem Artikel soll überprüft werden, ob sich automatisch generierte Kernbereiche der Forschungsliteratur für den Nutzer zur Informationsrecherche eignen. Es wird angenommen, dass die Kernzonen/Kernzeitschriften nach Bradfordizing (hohe Artikelhäufigkeit zu einem Thema) positive Auswirkungen für den Nutzer zeigen (Hood & Wilson, 2001).

Motivation dieses Artikels ist es, empirisch zu untersuchen, ob das hier vorgestellte alternative Re-Rankingverfahren Bradfordizing auf Basis der Bradfordschen Gesetzmäßigkeit im Anwendungsbereich bibliographischer Datenbanken zum einen operabel ist und zum anderen voraussichtlich gewinnbringend in Informationssystemen eingesetzt und dem Nutzer angeboten werden kann.

3 Bradfordsche Gesetzmäßigkeit

Der folgende Abschnitt trägt die Grundlagen von Bradford's Law of Scattering (BLS) zusammen.

Der Begriff BLS stammt ursprünglich von Vickery (1948), der sich sehr früh mit den Grundlagen der Bradfordschen Gesetzmäßigkeit beschäftigt hat. Die Beobachtung des Bibliothekars Samuel Bradford aus den frühen 1930er Jahren kann als eine der Wurzeln der Informetrie (engl. Informetrics) bezeichnet werden. Die Informetrie ist wiederum eine Spezialdisziplin der Informationswissenschaft, die sich mit der mathematischen Beschreibung und Modellierung der Regularitäten bzw. sogenannten „Gesetzmäßigkeiten" der beobachtbaren Objekte im Bibliotheks- und Informationswesen beschäftigt (vgl. Nacke, 1979). Zu diesen Objekten zählen z. B. Autoren, Publikationen (Zeitschriften, Monographien usw.), Referenzen, Zitationszahlen usw. Wichtig bei der Beurteilung der informetrischen Gesetzmäßigkeiten ist, dass diese keine Gesetzmäßigkeiten in Sinne von Naturgesetzen wie z. B. der Fallgesetze oder des Gravitationsgesetzes darstellen, sondern empirische Gesetze (Wilkinson, 1972). Diese können zwar mathematisch über Formeln beschrieben werden, Abweichungen und Unschärfe bei der Formulierung und Beobachtung der informetrischen Zusammenhänge sind aber die Norm. Häufig können die informetrischen Modelle die Realität sehr gut abbilden, weshalb die Bezeichnung Gesetzmäßigkeit naheliegt.

Dem BLS liegt zugrunde, dass sich die Literatur zu einem beliebigen Fachgebiet bzw. -thema, z. B. in einer Bibliographie, in Bereiche mit unterschiedlichen Dokumentenkonzentrationen unterteilen lässt. So besteht zwischen den Zeitschriften eines Forschungsthemas und den Artikeln in diesen Zeitschriften eine quantifizierbare Relation, die Bradford als erstes beschrieben hat. Dem Kernbereich mit hoher Konzentration der Literatur folgen Bereiche mit zunächst mittlerer und geringer Konzentration, die jeweils die gleiche Menge an Zeitschriftenartikeln beinhalten wie der Kernbereich. Bradford beschreibt diese Beobachtung mit:

> "The whole range of periodicals thus acts as a family of successive generations of diminishing kinship, each generation being greater in number than the preceding, and each constituent of a generation inversely according to its degree of remoteness." *(Bradford, 1934)*

Bradford hat das nach ihm benannte informetrische Gesetz wie folgt formuliert:

> "... if scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the numbers of periodicals in the nucleus and succeeding zones will be as $1 : n : n^2$ ..." *(Bradford, 1934, S. 116)*

Bradford selber hat in seinem Buch „Documentation" ein Kapitel „The Documentary Chaos" geschrieben und in diesem sein Verteilungsgesetz als Argument dafür genommen, Bibliographien künftig nach Zeitschriften auszuwerten und nicht nur nach rein fachlichen Gesichtspunkten aufzubauen. Bradford beschreibt damit sehr früh die Verfahren, mit denen noch heute Fachdatenbanken und Bibliographien aufgebaut werden. Bereits 1934 formuliert Bradford:

> "It follows that the only way to glean all the articles on these subjects would be to scrutinise continually several thousands of journals, the bulk of which would yield only occasional references or none at all. And because this work is altogether impracticable, a large portion of the total number of references on a given subject is regularly missed by the abstracting journals. ... In methods of abstracting and indexing, a radical change is needed. Periodical literature must be abstracted by source, and not by subject, as hitherto." *(Bradford, 1934, S. 179-180)*

Die Besonderheit der Bradfordschen Beobachtung liegt vor allem in der Betonung der Rangfolge bzw. dem Ranking der Zeitschriften nach der Anzahl der Artikel zu dem untersuchten Thema. Das gleiche Verfahren wird bei den verwandten Gesetzmäßigkeiten Lotka's Law für Publikationen eines Autors sowie Worthäufigkeiten in Texten in Zipf's Law angewendet. BLS weist der produktivsten Zeitschrift den Rank 1 zu, der zweitproduktivsten Zeitschrift den Rank 2 usw. Auf dem letzten Rank befinden sich die Zeitschriften, die die geringste Anzahl an Artikeln zu dem Thema produzieren, das heißt, i. d. R. mindestens einen Artikel beinhalten.

Egghe und Rousseau (1990) nennen diese spezifischen Verteilungen „rank-order distributions". Es werden in der Literatur aber auch Begriffe wie „long tail distributions", „extremely skewed", „law of the vital few" oder auch „schiefe Verteilung" verwendet. Power laws bzw. Potenzgesetze folgen im Übrigen der gleichen Verteilung (vgl. Newman, 2005).

Im Prinzip stellt das BLS lediglich eine Präzisierung der 80 : 20-Regel für den Bereich der Zeitschriftenliteratur dar. 80 Prozent der Nachfrage an Literatur können mit etwa 20 Prozent des Bestandes in einer Bibliothek abgedeckt werden (vgl. Umstätter, 2005; Wagner-Döbler, 1997). Anders formuliert: 80 Prozent der Zeitschriftenartikel zu einem Thema finden sich in 20 Prozent der Zeitschriften, die zu diesem Thema publizieren. BLS hat damit Parallelen zu Paretos Beobachtungen auf dem Gebiet der Einkommensverteilungen.

Folgendes Beispiel soll zur Illustration der Bradfordschen Gesetzmäßigkeit dienen. Angenommen, es werden 3.000 Artikel zu einem bestimmten Thema in insgesamt 800 Zeitschriften gefunden. Dann ergibt sich in der oben beschriebenen Reihung nach Bradford bei drei gleich großen Gruppen von jeweils 1.000 Artikeln ein Verhältnis von 42 : 158 : 600 Zeitschriften. Näherungsweise $1 : 3{,}8 : 3{,}8^2$

Der Core oder Kern der Zeitschriftenmenge besteht demnach aus insgesamt 42 Zeitschriften. Diese 42 Zeitschriften beinhalten die ersten 1.000 Artikel. In der

zweiten Zone (Zone 2) befinden sich 158 Zeitschriften, die wiederum 1.000 Artikel beinhalten. Die 600 Zeitschriften aus der dritten Zone (Zone 3) gehören zu den Peripheriezeitschriften, die nur selten, aber wenn dann mindestens einen thematisch relevanten Artikel beinhalten. Die Konzentration der Zeitschriftenartikel in den Zeitschriften nimmt somit von Zone zu Zone konstant ab. Charakteristisch und allen Bradford-Verteilungen gemeinsam ist die auftretende Streuung (Ungleichverteilung) bzgl. der Artikelproduktivität einzelner Zeitschriften zwischen den Zonen, die sich annähernd wie ein Power law verhält.

Die Einteilung der beobachteten Objekte (Zeitschriften) in drei Zonen bei Bradford ist eine pragmatische Entscheidung, die keinen Zusammenhang mit der eigentlichen Gesetzmäßigkeit hat. Obwohl Bradford nur drei Zonen bzw. Gruppen zugrundelegt, sind aufgrund der „Selbstähnlichkeit" (self-similarity) dieser Verteilungen auch mehr Zonen möglich. Egghe und Rousseau (1990) zeigen dies an einem Beispiel.

Die umfangreiche Literatur zu BLS (siehe dazu Mayr, 2010) lässt sich in unterschiedliche Gruppen einteilen: a) die Bestätigung der Bradford-Gesetzmäßigkeit in unterschiedlichen Fachgebieten und Disziplinen, b) die Diskussion über die mathematisch geeignete und korrekte Formulierung der Gesetzmäßigkeit sowie die Beziehung von BLS zu anderen informetrischen Gesetzmäßigkeiten und c) die Diskussion der theoretischen und praktischen Implikationen des Gesetzes.

Bisher sind in der Literatur unterschiedliche Einsatzmöglichkeiten des BLS beschrieben, von denen hier nur einige genannt werden sollen. BLS dient dabei:

- zum Aufbau und Management sowie zur Kontrolle von Kollektionen,
- zur Auswahl von Zeitschriften für eine Fachdatenbank (z. B. bei Thomson Scientific; Garfield 1990;1996),
- zur Messung der Abdeckung von Bibliographien und
- zur Rechercheunterstützung (vor allem Bates, 2002, White, 1981).

Insbesondere die Einsatzmöglichkeiten des BLS für die Informationssuche bei Bates (Bates, 1989, Bates, 2002) sind relevant für diesen Artikel: "… the key point is that the distribution tells us that information is neither randomly scattered, nor handily concentrated in a single location. Instead, information scatters in a characteristic pattern, a pattern that should have obvious implications for how that information can most successfully and efficiently be sought." (Bates, 2002)

Es lassen sich unterschiedliche Szenarien denken, in denen ein Nutzer die Suche nur auf die Kernjournale eines Fachgebiets einschränken will (z. B. Überblicksrecherche). Im Sinne der Rechercheunterstützung kann davon ausgegangen werden, dass das Verteilungsmuster des BLS (Kernzone mit hoher Artikelhäufigkeit) für den Benutzer (vermutlich insbesondere ungeübte Nutzer und Anfänger) positive Auswirkungen beim Information Retrieval haben kann.

4 Bradfordizing

Das Verfahren Bradfordizing wurde erstmals in White (1981) zur Umorganisation von Suchergebnissen in Onlinedatenbanken vorgeschlagen. Bradfordizing sortiert bzw. rankt eine beliebige Dokumentmenge nach Kernzeitschriften. Die Zeitschriften zu einer Suchanfrage werden gemäß ihrer Häufigkeit (Anzahl der Zeitschriftenaufsätze) gelistet. Bradfordizing kann daher als ein nicht-textorientiertes Rankingverfahren bezeichnet werden. Aufsätze der Zeitschriften aus Zone 1 werden beim Bradfordizing vor die Zeitschriftenartikel geordnet, deren Zeitschriften nur durchschnittlich (Zone 2) oder selten (Zone 3) gelistet werden.

Bradfordizing ist ein Sortierverfahren, das auf unterschiedliche Dokumenttypen in wissenschaftlichen Fachdatenbanken angewendet werden kann. White beschreibt das Verfahren folgendermaßen: "… That is sorting hits (1) by the journal in which they appear, and then sorting these journals not alphabetically by title but (2) numerically, high to low, by number of hits each journal contains. In effect, this two-step sorting ranks the search output in the classic Bradford manner, so that the most productive, in terms of its yield of hits, is placed first; the second-most productive journal is second; and so on, down through the last rank of journals yielding only one hit apiece." (S. 47).

Bradfordizing ist insofern ein innovatives Re-Ranking-Verfahren, als es für eine beliebige Suchanfrage über eine nachgeordnete Umorganisation des Suchergebnisses zuverlässig (hohe Robustheit dieser Häufigkeitsverteilung) die zentralen Publikationen für eine übergreifende Suche in der Ergebnisliste nach vorne sortiert, ohne die beschreibenden Textbestandteile der Dokumente zu analysieren. Bradfordizing „… erleichtert nicht nur das Auffinden von Aufsätzen und hilft Anfängern rasch zu erkennen, welche Zeitschriften für sie besonders wichtig sind, es erhöht auch die Wahrscheinlichkeit, gerade in diesen Quellen noch weitere wichtige Publikationen zu finden …" (Umstätter, 2005)

Der Ablauf einer Bradfordizing-Analyse lässt sich folgendermaßen zusammenfassen:

1. Identifikation von allen Dokumenten (Artikeln), die in einem Forschungsgebiet/Subject publiziert wurden, durch retrospektive Suche in einem ausreichend großen Zeitraum.
2. Auflistung der Quellen (Zeitschriften), in denen die Artikel publiziert wurden, in der Reihenfolge der Häufigkeit der Quelle. Das Ergebnis ist eine gerankte Liste der Zeitschriften (sog. „bradfordized list").
3. Einteilung der Zeitschriftenliste in Gruppen/Zonen (bei Beibehaltung der oberen Reihefolge), so dass die Anzahl der Artikel pro Gruppe/Zone in etwa gleich ist.

Das folgende idealisierte Beispiel in Abbildung 1 visualisiert die drei Zonen nach einer Bradfordizing-Analyse. Die Dokumentmenge umfasst in diesem Beispiel 450 Zeitschriftenartikel zu einem Forschungsthema, die in dem Fall auf insgesamt 39 unterschiedliche Zeitschriften verteilt sind (150 Artikel in jeder der drei Zonen). Das Kumulieren der Artikelzahlen nach dem Bradfordizing hat in diesem Fall ergeben, dass die ersten drei Zeitschriften zusammen 150 Artikel ergeben, also das erste Drittel der Gesamtdokumentenzahl von 450 Dokumenten. Diese ersten 150 Dokumente in den drei Zeitschriften definieren damit den Nukleus oder Core für dieses Topic. Die drei Zeitschriften werden „Core Journals" oder Kernzeitschriften genannt. Für das zweite Drittel bzw. Zone 2 werden 9 (3*3 Zeitschriften) weitere Zeitschriften und für das dritte Drittel bzw. Zone 3 werden 27 (3*3*3 Zeitschriften) Zeitschriften benötigt, um die Menge von jeweils 150 Zeitschriftenartikeln zu erreichen.

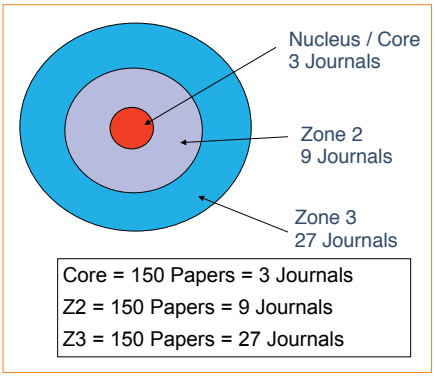


Abbildung 1: Ergebnis eines idealisierten Dokumentenpools für eine Fragestellung nach Bradfordizing. Einteilung in typische drei Bradford-Zonen (Core, Zone 2 und Zone 3). Die 450 Zeitschriftenartikel sind über 39 Zeitschriften verteilt.

Die Hauptidee der Anwendung von Bradfordizing in diesem Artikel ist es, ein beliebiges thematisch spezifiziertes Suchergebnis durch das einfach anwendbare Verfahren umzuorganisieren (Re-Ranking, vgl. Lin, 2008) und den Effekt dieses Re-Rankings zu evaluieren. Ziel ist es, nach dem Re-Ranking mehr relevante Dokumente für ein Topic im ersten Drittel der Ergebnismenge zu listen als in den

nachfolgenden Zonen. Dieser Zugewinn an relevanten Dokumenten im Kern soll als Bradfordizing-Effekt bezeichnet werden (siehe dazu Mayr, 2008).
Die Fokussierung auf Bradfordizing erscheint vielversprechend, weil dieses Verfahren per se universell und disziplinübergreifend angelegt ist und zuverlässig sowohl innerhalb einer Datenbank, einer Domäne (mehreren Datenbanken zu einem Fachgebiet) als auch zwischen Domänen beobachtet werden kann.
Bradfordizing liefert folgende unmittelbare Mehrwerte für den Nutzer:

- eine alternative Sicht auf Suchergebnisse, die nach Kernzeitschriften umorganisiert sind,
- eine alternative Sicht auf Publikationsquellen innerhalb eines Suchraums, die intuitiv näher am Forschungsprozess liegt als textstatistische Verfahren (z. B. best match) oder traditionelle boolesche Verfahren (exact match),
- eine vermutlich höhere fachliche Relevanz (topicality) der Dokumente nach dem Re-Ranking.

Aufgrund der Robustheit und Allgemeingültigkeit des BLS für die Verteilung von Forschungsliteratur ist davon auszugehen, dass das Bradfordizing gerade in föderierten und interdisziplinären Suchumgebungen mit unterschiedlichen Informationsbeständen praktisch operabel und gewinnbringend ist.

5 Forschungsfragen

Ziel der nachfolgenden Untersuchung ist es, eine umfangreiche Anzahl von standardisierten Fragestellungen in unterschiedlichen Datenbanken und Fachgebieten durch Bradfordizing zu analysieren. Dabei stellen sich folgende Fragen: Kann das nicht-textorientierte Rankingverfahren Bradfordizing ganz generell als ein Mehrwertdienst in Informationssystemen (z. B. Fachdatenbanken, Fachportalen mit mehreren Datenbanken) eingesetzt werden? Verbessert das Bradfordizing den Anteil der relevanten Dokumente im ersten Drittel (Core) einer Dokumentmenge gegenüber dem zweiten und dritten Drittel (Zone 2 bzw. Zone 3)? Zu diesem Zweck sollen der Re-Ranking-Ansatz Bradfordizing für Zeitschriftenartikel angewendet und in einem realen Suchszenario evaluiert werden. Des Weiteren wird geprüft, ob sich das Bradfordizing, das ursprünglich nur für Zeitschriftenliteratur beschrieben wurde, auch auf andere Dokumenttypen anwenden lässt.
Folgende Forschungsfragen lassen sich formulieren:

1. Welche Auswirkungen/Effekte hat die Anwendung von Bradfordizing als Re-Ranking-Ansatz für das Information Retrieval (IR)? Lassen sich die Effekte innerhalb eines klassischen IR-Tests evaluieren?
2. Lässt sich das Verfahren Bradfordizing als Re-Ranking-Ansatz auch auf andere Dokumenttypen und Szenarien übertragen? Konkret - kann Bradfordizing auch auf Monographien angewendet werden?

Im folgenden Abschnitt wird die Vorgehensweise der Evaluation erläutert. Eine detaillierte Beschreibung der angewendeten Methoden findet sich in Mayr (2010).

6 Design und Ablauf der Studie

In dieser Studie wurde Bradfordizing auf Dokumentenpools aus unterschiedlichen Fachdatenbanken aus den Bereichen Sozial- und Politikwissenschaften, Wirtschaftswissenschaften, Psychologie und Medizin zu verschiedenen Fragestellungen angewendet (siehe Tabelle 1). Für die Retrieval Tests wurden Daten aus zwei umfangreichen Evaluationsprojekten verwendet: zum einen Fragestellungen und Bewertungsdaten der jährlich stattfindenden internationalen CLEF-Initiative (siehe http://clef.iei.pi.cnr.it/), zum anderen Daten, die im Rahmen des GESIS-Projekts KoMoHe im Jahr 2007 entstanden sind (siehe Mayr & Petras, 2008). Die beiden Evaluationsprojekte eignen sich besonders für diese Studie, weil es sich bei den getesteten Dokumenten größtenteils um Zeitschriftenartikel und Monographien handelt, die bezüglich konkreter Fragestellungen intellektuell bewertet wurden. Für jede Fragestellung (im Kontext Information Retrieval Topic genannt) stehen jeweils ein Datensatz (Dokumentenpool) mit gefundenen Dokumenten und die entsprechenden binären Bewertungsdaten (relevant oder nichtrelevant) der Juroren zur weiteren Auswertung zur Verfügung (zum Verfahren siehe Harman & Voorhees, 2006). Die Topics bzw. die Dokumentenpools zu diesen Topics dienen jeweils als Grundlage für die Bradfordizing-Analyse (siehe dazu Tabelle 1).

Insgesamt wurden 164 intellektuell bewertete Dokumentenpools aus beiden Projekten analysiert. In diesen Dokumentenpools befanden sich mehr als 96.000 Dokumente, von denen mehr als 51.000 Dokumente durch das Verfahren Bradfordizing behandelt werden konnten.
In dieser Studie wird Bradfordizing auf die beiden Dokumenttypen Zeitschriftenartikel und Monographien angewendet. Da die Zeitschriftenartikel i. d. R. eine identifizierende Nummer der Zeitschrift (ISSN-Nummer) tragen, kann die anschließende Häufigkeitsanalyse (das eigentliche Bradfordizing) auf Basis der ISSN (International Standard Serial Number) erfolgen. Das gleiche Verfahren kann auf die ISBN-Nummer bei den Monographien angewendet werden. Die ISBN (International Standard Book Number) ist ein Identifier für Monographien und andere selbstständige Veröffentlichungen, in dem der Verlag, der die Publikation herausgibt, kodiert ist. Die Verlagsnummer ist eine ein- oder mehrstellige Zahl, die eindeutig einem Verlag zugeordnet ist. Dieser Verlags-Code wurde für die Häufigkeitsanalyse (Bradfordizing) verwendet.
Im Experiment wurden zunächst die beiden Dokumenttypen Zeitschriftenartikel und Monographien pro Topic aus den Dokumentenpools extrahiert und jeweils getrennt voneinander betrachtet. Abbildung 2 verdeutlicht den Prozess der Auswertung der evaluierten Topics aus den beiden Evaluationsprojekten KoMoHe und CLEF nach dem Verfahren Bradfordizing.

Die Analyse der Datensätze teilt sich in drei Schritte:

	CLEF	**KoMoHe**
Zeitraum	2003-2007	2007
Anzahl der Topics	125	39
Domäne/Disziplin	Sozial- und Politikwissenschaften	Sozial- und Politikwissenschaften, Wirtschaftswissenschaften, Psychologie und Medizin
Bewertete Dokumente total	65.297	31.155
Zeitschriftenartikel bradfordized	18.112	17.432
Monographien bradfordized	11.045	4.900
Datenbanken involviert	1 (SOLIS)	6 (SOLIS, SoLit, USB Köln OPAC, Psyndex, World Affairs Online, Medline)

Tabelle 1: Übersicht über die analysierten Topics, Dokumente und Datenbanken in den IR-Experimenten (CLEF und KoMoHe).

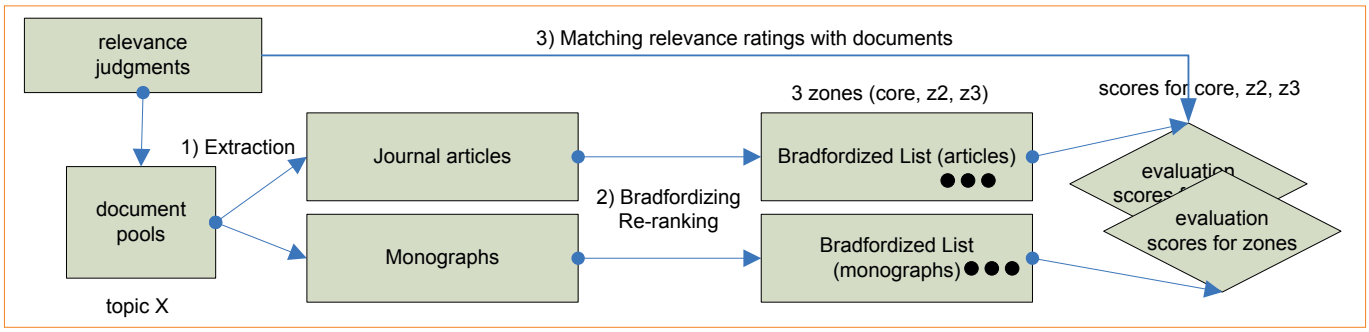


Abbildung 2: Auswertung der Relevanzbewertung im Re-Ranking-Experiment (KoMoHe und CLEF-Topics).

1. Die Dokumenttypen Zeitschriftenartikel (journal articles) und Monographien (monographs) werden aus dem ursprünglichen Datensatz extrahiert. Jedes Topic wird separat analysiert.
2. Anschließend wird die Dokumentmenge pro Topic nach dem Verfahren Bradfordizing (Häufigkeit einer Zeitschrift bzw. eines Verlages) umorganisiert und in drei etwa gleich große Zonen eingeteilt (core, z2 und z3).
3. Im nächsten Schritt wird die Verteilung der Relevanzbewertungen der Dokumente in diesen Zonen analysiert. Dazu werden die Dokumenten-ID mit den Bewertungen zu den jeweiligen Topics abgeglichen und ausgewertet.

Im Anschluss an die Untersuchung liegen für jedes Topic und jeden Dokumenttyp die Anteile der relevanten bzw. nichtrelevanten Dokumente für die jeweilige Zone des Dokumentpools vor. Daraus lassen sich Precision-Werte berechnen. Die Precision misst das Verhältnis von gefundenen relevanten Dokumenten zu einer Suchanfrage an allen gefundenen Dokumenten aus einer Kollektion. Die Precision misst damit die Präzision bzw. Exaktheit eines Retrieval-Ergebnisses. Wenn beispielsweise nur relevante Dokumente aus der Kollektion zu einer Fragestellung gefunden werden, nimmt die Precision den Wert 1 an.
Es wurde pro Topic und Zone (core, z2 und z3) die durchschnittliche Precision berechnet. Zudem wurde die durchschnittliche Precision für das gesamte Dokumentset (baseline) berechnet (siehe Tabelle 2 als Beispiel).

Tabelle 2: Beispiel für die Berechnung der Precision-Werte für ein Topic

	Retrieved	Relevant	Precision
Core	73	41	0.56 (P core)
Zone 2	65	25	0.38 (P z2)
Zone 3	70	14	0.20 (P z3)
Total	208	80	0.38 (P baseline)

Folgende Hypothese lässt sich für die Auswertung der Precisionwerte in unserem Test formulieren:

Wenn der Anteil der relevanten Dokumente, gemessen in Precision (P), in allen drei Zonen gleich groß ist, dann hat das Verfahren Bradfordizing keinen Effekt auf die Verteilung der Relevanz im gesamten Dokumentpool. Wenn der Anteil der relevanten Dokumente in Zone Core niedriger ist als in Zone 2 oder Zone 3, dann verschlechtert sich das Suchergebnis. Wenn der Anteil der relevanten Dokumente im Core höher liegt als in den übrigen Zonen, was erwartet wird, dann verbessert das Verfahren das Suchergebnis und hat somit einen positiven Effekt auf die Suche.

7 Ergebnisse

Die Hauptergebnisse der Analyse der Dokumentverteilung nach Bradfordizing werden nachfolgend knapp am Beispiel der Zeitschriften zusammengefasst. Die Auswertungen zur Monographieliteratur finden sich in Mayr (2010).

- Die Analyse der Topics aus den Projekten CLEF und KoMoHe zeigt deutlich, dass Bradford-ähnliche Verteilungen für beide Datensätze (Datenbanken SOLIS, SoLit, USB Köln OPAC, Psyndex, World Affairs Online, Medline) und beide Dokumenttypen (Zeitschriftenartikel und Monographien) nachgewiesen werden können. Prinzipiell kann das Verfahren Bradfordizing aufgrund des Auftretens der Bradfordschen Gesetzmäßigkeit für beide Dokumenttypen auf Basis der rein quantitativen Analyse angewendet werden.

Der Einsatz von Bradfordizing soll das Suchergebnis komplett neu zusammensetzen und es möglichst signifikant verbessern. Ziel ist es, mehr relevante Dokumente für das Topic im ersten Drittel der Ergebnismenge zu listen. Im Mittelpunkt der Analyse steht der Vergleich der durchschnittlichen Precision-Werte der drei typischen Dokumentzonen (Bradford-Zonen) für bewertete Dokumentenpools. Die Auswertung der Relevanzurteile ergibt folgendes Ergebnis:

- Die Auswertung der Relevanzurteile für den Dokumenttyp Zeitschriftenartikel zu den 125 Topics der CLEF-Reihe zeigt ein deutliches Ergebnis. Die Dokumente, die nach Bradfordizing in die Kernzone (core) gerankt werden, sind im Mittel häufiger relevant für die Fragestellungen als Dokumente, die sich in Zone 2 oder Zone 3 befinden. Tabelle 3 zeigt die gemittelten Precision-Werte für 125 CLEF-Topics. Im Durchschnitt beträgt die Precision in der Kernzone 0,28. Die Precision sinkt in den folgenden Zonen deutlich ab (Zone 2 = 0,23 und Zone 3 = 0,19). Die baseline-Precision fasst die durchschnittliche Precision über alle Zeitschriftenartikel zusammen. Tabelle 4 zeigt die prozentuale Verbesserung der Precision der drei Zonen untereinander. Als Vergleichswert ist zusätzlich die baseline-Precision mitgeführt. Signifikante Verbesserungen sind jeweils durch das Sternsymbol gekennzeichnet (*). Es wird nicht zwischen unterschiedlichen Signifikanzniveaus unterschieden. Die Kernzone (core) enthält demnach über alle CLEF-Jahrgänge ca. 56 Prozent mehr relevante Dokumente als Zone 3. Die Kernzone beinhaltet ebenfalls ca. 20 Prozent mehr relevante Dokumente als die nachfolgende Zone 2. Zone 2 beinhaltet wiederum ca. 28 Prozent mehr relevante Dokumente als Zone 3. Der Vergleich von Core und baseline-Precision fällt ebenfalls positiv für die Dokumente im Core aus. Die Zeitschriftenartikel im Core erzielen eine um knapp 22 Prozent bessere Precision als der Durchschnitt über alle Zeitschriftenartikel.

Die Ergebnisse der CLEF-Evaluation können im KoMoHe-Datensatz bestätigt werden. Tabelle 5 zeigt die gemittelten Precision-Werte für 39 KoMoHe-Topics. Die 39 Topics zeigen die gleiche Tendenz: Die Dokumente der Kernzone sind signifikant häufiger relevant als Dokumente der Zone 2 und Zone 3. Im Durchschnitt beträgt die Precision in der Kernzone 0,40. Die Precision sinkt in den folgenden Zonen ab (Zone 2 = 0,37 und Zone 3 = 0,34). Die Kernzone (core) enthält demnach über alle KoMoHe-Tests ca. 17 Prozent mehr relevante Dokumente als Zone 3 (siehe Tabelle 6). Die Kernzone beinhaltet ebenfalls ca. 9 Prozent mehr relevante

Tabelle 3: Durchschnittliche Precision für Zeitschriftenartikel nach Re-Ranking. Fünf CLEF-Jahrgänge (N=125 topics). Core, Zone 2 (Z2), Zone 3 (Z3) und baseline.

CLEF articles	Topics	P core	P Z2	P Z3	P baseline
2003	25	0.294	0.218	0.157	0.221
2004	25	0.226	0.185	0.134	0.179
2005	25	0.310	0.240	0.174	0.239
2006	25	0.288	0.267	0.244	0.265
2007	25	0.278	0.256	0.217	0.248

Tabelle 4: Durchschnittliche Precision-Verbesserung in % für Zeitschriftenartikel. Fünf CLEF-Jahrgänge (N=125 topics). Core, Zone 2 (Z2), Zone 3 (Z3) und baseline.

CLEF articles	Core against Z3 in %	Core against Z2 in %	Z2 against Z3 in %	core against baseline in %
2003	86.56 (*)	34.57 (*)	38.63 (*)	32.65 (*)
2004	69.23 (*)	22.45	38.20	26.25 (*)
2005	78.03 (*)	29.05 (*)	37.95 (*)	29.52 (*)
2006	17.63	7.66	9.27	8.46
2007	28.18 (*)	8.31	18.35	11.77
Average 2003-2007	**55.93 (*)**	**20.41 (*)**	**28.48 (*)**	**21.73 (*)**

Tabelle 5: Durchschnittliche Precision für Zeitschriftenartikel nach Re-Ranking. Drei KoMoHe-Tests (N=39 topics). Core, Zone 2 (Z2), Zone 3 (Z3) und baseline.

KoMoHe Articles	Topics	P core	P Z2	P Z3	P baseline
Test1	15	0.292	0.261	0.245	0.265
Test2	12	0.215	0.202	0.192	0.202
Test3	12	0.700	0.644	0.587	0.642

Tabelle 6: Durchschnittliche Precision-Verbesserung in Prozent für Zeitschriftenartikel. Drei KoMoHe-Tests (N=39 topics). Core, Zone 2 (Z2), Zone 3 (Z3) und baseline.

KoMoHe articles	Core against Z3 in %	Core against Z2 in %	Z2 against Z3 in %	Core against baseline in %
Test1	18.82	11.75	6.32	9.84
Test2	11.58	6.16	5.11	6.12
Test3	19.32 (*)	8.67 (*)	9.80 (*)	9.00 (*)
Average Test1-3	16.57 (*)	8.86	7.08 (*)	8.32 (*)

Dokumente als die nachfolgende Zone 2. Zone 2 beinhaltet wiederum ca. 7 Prozent mehr relevante Dokumente als Zone 3. Der Vergleich von Core und baseline-Precision fällt ebenfalls positiv für die Dokumente im Core aus. Die Zeitschriftenartikel im Core erzielen eine um knapp 8 Prozent bessere Precision als der Durchschnitt über alle Zeitschriftenartikel.

Ein Vergleich dieser Ergebnisse zeigt sich, dass die Verbesserung der Precision zwischen den Zonen bei den KoMoHe-Tests nicht so deutlich ausfällt wie bei den CLEF-Topics (siehe Tabelle 4 und 6). Die Dokumente der Kernzone zeigen für die meisten Testreihen signifikant höhere Precision als Dokumente der Zone 2 und Zone 3. Sowohl für die Zeitschriften als auch für die Monographien kann diese Aussage auf einer sehr breiten Basis von Themen und Fragestellungen an zwei unabhängigen Dokumentkorpora empirisch nachgewiesen werden. Das Verfahren Bradfordizing bzw. das Re-Ranking nach der Zeitschriftenhäufigkeit ist demnach für die zu einem Thema erfasste Literatur in Fachdatenbanken (insbesondere Zeitschriftenliteratur) aufgrund der signifikanten Precision-Verbesserung sehr vielversprechend.

8 Zusammenfassung

Die Besonderheit dieser Untersuchung liegt darin, dass für eine große Anzahl von sehr unterschiedlichen Fragestellungen die Konzentration der Literatur in vergleichsweise kleinen (unvollständigen) Dokumentkorpora untersucht wurde.

Die Precision für 125 getestete Topics in der Datenbank SOLIS erhöht sich nach Bradfordizing um durchschnittlich 56 Prozent (Vergleich der Precision der Kernzeitschriften und der Zone 3-Zeitschriften). Diese signifikante Verbesserung durch Bradfordizing tritt auch für andere Datenbanken und Disziplinen ein. Die Precision für 39 getestete Topics in den Datenbanken SOLIS, SoLit, USB Köln OPAC, Psyndex, Econis, World Affairs Online, Medline erhöht sich nach Bradfordizing um durchschnittlich 17 Prozent (Vergleich der Precision der Kernzeitschriften und der Zone 3-Zeitschriften).

Die Dokumente der Kernzone zeigen für die meisten Testreihen signifikant höhere Precision als Dokumente der Zone 2 und Zone 3. Sowohl für die Zeitschriften als auch für die Monographien kann diese Aussage auf einer sehr breiten Basis von Themen und Fragestellungen an zwei unabhängigen Dokumentkorpora empirisch nachgewiesen werden. Das Verfahren Bradfordizing bzw. das Re-Ranking nach der Zeitschriftenhäufigkeit ist demnach für die erfasste Literatur in Fachdatenbanken (insbesondere Zeitschriftenliteratur) zu einem Thema aufgrund der signifikanten Precision-Verbesserung sehr vielversprechend. Gestützt werden die empirischen Ergebnisse durch eine qualitative Fallstudie die ebenfalls im Rahmen der Dissertation durchgeführt wurden.

Zusätzlich zu den Zeitschriften wurde ein Teil der CLEF- und KoMoHe-Topics auch für die Monographienliteratur untersucht (siehe detaillierter in Mayr, 2010). Die bei Worthen (1975) beschriebene Übertragbarkeit der Bradfordschen Gesetzmäßigkeit auf die Monographienliteratur kann durch diese Studie für die analysierten Topics auf einer sehr breiten Basis bestätigt werden. Es zeigt sich, dass sich das in dieser Arbeit entwickelte Vorgehen der Analyse der Verlags-Codes aus der ISBN der Monographie-Datensätze als tragbares Verfahren erweist.

Sehr abstrakt ausgedrückt soll Bradfordizing durch die Identifikation von Kernbereichen als ein Kompensationsverfahren für erweiterte Suchräume (mehrere Datenbanken in einer Suchanfrage) eingesetzt werden. Der Heterogenität und der Vielzahl der Datenbanken wird als Ausgleich ein quasi vereinheitlichendes Prinzip entgegengesetzt, das Kernbereiche identifiziert, die über Datenbanken und Domänen hinweg existieren.

Folgende beiden Thesen können als Ergebnisthesen formuliert werden:

1. Die Anwendung des Bradfordizings bzw. das Re-Ranking nach Kernzeitschriften für thematische Do-kumentmengen führt zu signifikanten Verbesserungen der Precision zwischen den

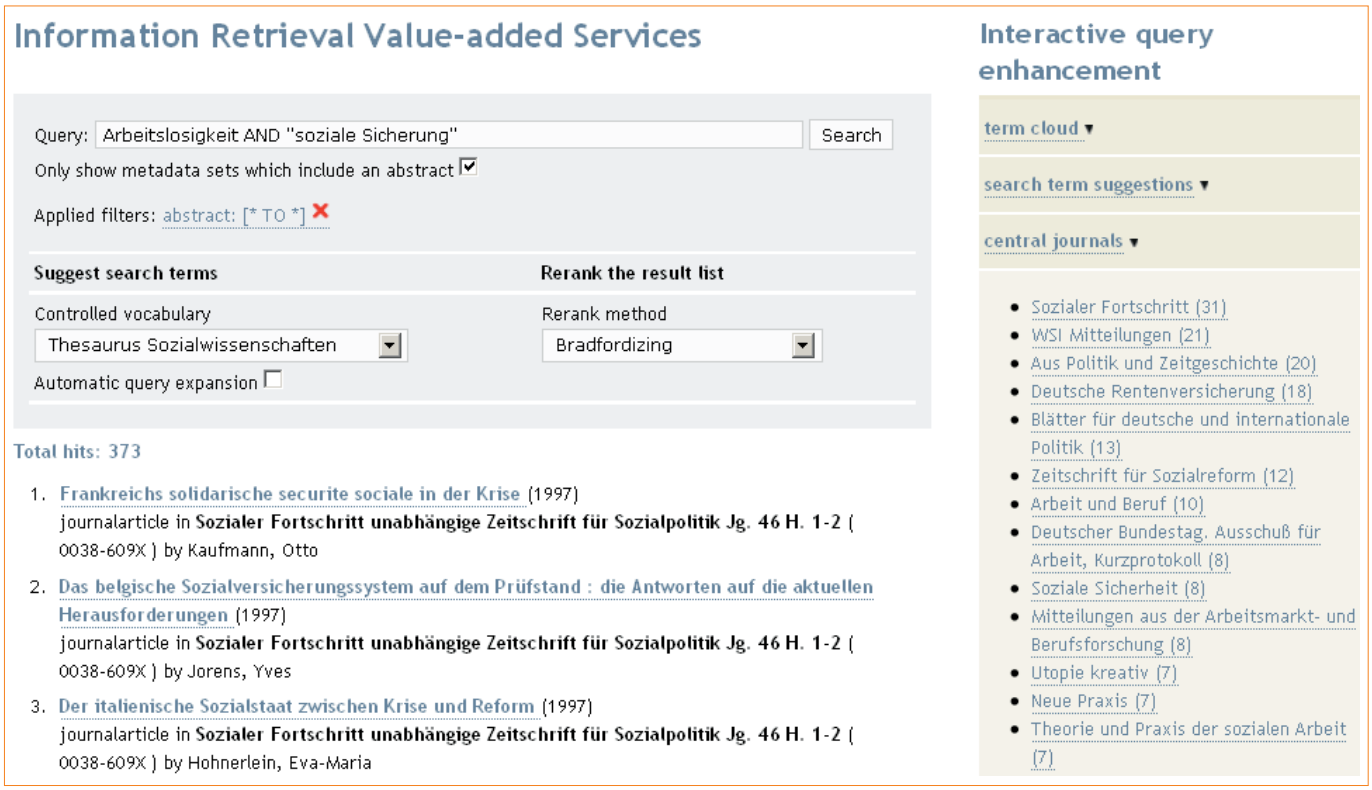


Abbildung 3: Beispiel eines neugerankten Ergebnisses (nach Bradfordizing) im IRM-Prototyp.

drei Zonen (Core, Zone 2 und Zone 3). Die Kernzeitschriften (Core) beinhalten signifikant mehr relevante Dokumente als Zeitschriften der Zone 2 oder der Zone 3. Der größte Precision-Gewinn ergibt sich zwischen Core und Zone 3-Zeitschriften.

2. Bradfordizing für thematisch konzentrierte Dokumentmengen lässt sich erfolgreich auf Monographien (Verlag als Selektionskriterium) übertragen. Die Anwendung des Bradfordizings bzw. das Re-Ranking nach Verlagen für Monographien führt zu geringeren Verbesserungen der Precision zwischen den drei Zonen (Core, Zone 2 und Zone 3). Der größte Precision-Gewinn ergibt sich auch hier wie bei den Zeitschriften zwischen Core und Zone 3-Verlagen.

9 Ausblick

Die Ergebnisse dieser Arbeit gehen in das GESIS-Projekt „Retrieval-Mehrwertdienste zur Weiterentwicklung wissenschaftlicher Fachportale wie vascoda und sowiport. Suchexpandierung und Re-Ranking“ (kurz IR-Mehrwertdienste) ein, das im Januar 2009 gestartet wurde. Das Projekt IR-Mehrwertdienste[1] (DFG-Förderung INST 658/6-1) konzentriert sich auf die Implementation und Evaluation von drei Mehrwertdiensten (u. a. Bradfordizing), die die Informationssuche in Digitalen Bibliotheken verbessern sollen (siehe dazu Mayr et al., 2011). Abbildung 3 zeigt ein Beispiel eines nach Bradfordizing gerankten Suchergebnisses im Prototypensystem des Projekts IRM[2].

Literatur

Bates, Marcia J. (1989): The Design of Browsing and Berrypicking Techniques for the Online Search Interface. In: Online Review 13, Nr. 5, S. 407-424. URL: http://www.gseis.ucla.edu/faculty/bates/berrypicking.html.

Bates, Marcia J. (2002): Speculations on Browsing, Directed Searching, and Linking in Relation to the Bradford Distribution. S. 137-150. In: Bruce, Harry; Fidel, Raya; Ingwersen, Peter; Vakkari, Pertti (Hrsg.): Fourth International Conference on Conceptions of Library and Information Science (CoLIS 4) URL: http://www.gseis.ucla.edu/faculty/bates/articles/Searching_Bradford-m020430.html.

Bradford, Samuel C. (1934): Sources of information on specific subjects. In: Engineering 137, Nr. 3550, S. 85-86.

Bradford, Samuel C. (1948): Documentation. London: Lockwood. 156 S.

Egghe, Leo; Rousseau, Ronald (1990): Introduction to Informetrics: quantitative methods in library, documentation and information science: Elsevier Science Publishers. URL: http://eprints.rclis.org/archive/00003297/.

Garfield, Eugene (1990): How ISI Selects Journals for Coverage: Quantitative and Qualitative Considerations. In: Current Contents 13, Nr. 22, S. 5-13. URL: http://www.garfield.library.upenn.edu/essays/v13p185y1990.pdf.

Garfield, Eugene (1996): The Significant Scientific Literature Appears In A Small Core Of Journals. In: The Scientist 10, Nr. 17, S. 13. URL: http://www.garfield.library.upenn.edu/commentaries/tsv10(17)p13y090296.html.

Harman, Donna K.; Voorhees, Ellen M. (2006): TREC: An Overview. In: Annual Review of Information Science and Technology 40, S. 113-156.

Hood, William W.; Wilson, Concepción S. (2001): The scatter of documents over databases in different subject domains: how many databases are needed? In: Journal of the American Society for Information Science and Technology 52, Nr. 14, S. 1242-1254.

Krause, Jürgen (2004): Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. In: ZfBB: Zeitschrift für Bibliothekswesen und Bibliographie 51, Nr. 2, S. 76-89.

Krause, Jürgen (2008): Total Package Design für Digitale Bibliotheken und Fachinformation. S. 185-206. In: Hutzler, Evelinde; Schröder, Albert; Schweikl, Gabriele (Hrsg.): Bibliotheken gestalten Zukunft. Kooperative Wege zur Digitalen Bibliothek. Dr. Friedrich Geißelmann zum 65. Geburtstag. Göttingen: Universitätsverlag Göttingen. URL: http://epub.uni-regensburg.de/4564/.

Krause, Jürgen; Niggemann, Elisabeth; Schwänzl, Roland (2003): Normierung und Standardisierung in sich verändernden Kontexten: Beispiel Virtuelle Fachbibliotheken. In: ZfBB: Zeitschrift für Bibliothekswesen und Bibliographie 50, Nr. 1, S. 19-28.

Lin, Jimmy (2008): PageRank without hyperlinks: Reranking with PubMed related article networks for biomedical text retrieval. In: BMC Bioinformatics 9. URL: http://www.biomedcentral.com/1471-2105/9/270.

Mayr, Philipp (2008): An evaluation of Bradfordizing effects. In: Kretschmer, H.; Havemann, F. (Hrsg.): Proceedings of WIS 2008, Berlin, Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting. Humboldt-Universität zu Berlin URL: http://www.collnet.de/Berlin-2008/MayrWIS2008ebe.pdf.

Mayr, Philipp (2010): Information Retrieval-Mehrwertdienste für Digitale Bibliotheken: Crosskonkordanzen und Bradfordizing. GESIS-Schriftenreihe 5.

Mayr, Philipp; Mutschke, Peter; Schaer, Philipp; Sure, York (2011 to appear): Science Models as Value-Added Services for Scholarly Information Systems. In: Scientometrics.

1 www.gesis.org/irm

2 www.gesis.org/beta/prototypen/irm/

Mayr, Philipp; Mutschke, Peter; Petras, Vivien (2008): Reducing semantic complexity in distributed digital libraries: Treatment of term vagueness and document re-ranking. In: Library Review 57, Nr. 3, S. 213-224. URL: http://www.ib.hu-berlin.de/~mayr/arbeiten/mayr-etal_LR08.pdf.

Mayr, Philipp; Petras, Vivien (2008): Cross-concordances: terminology mapping and its effectiveness for information retrieval. In: 74th IFLA World Library and Information Congress. Québec, Canada URL: http://www.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf.

Mutschke, Peter (2004): Autorennetzwerke: Netzwerkanalyse als Mehrwertdienst für Informationssysteme. S. 141-162. In: Bekavac, Bernard; Herget, Josef; Rittberger, Marc (Hrsg.): Information zwischen Kultur und Marktwirtschaft. Proceedings des 9. Internationalen Symposiums für Informationswissenschaft (ISI 2004). Chur: UVK Verlagsgesellschaft mbH. (Schriften zur Informationswissenschaft, Band 42) URL: http://www.inf-wiss.uni-konstanz.de/infwiss/download/cc-isi04-art9.pdf.

Nacke, Otto (1979): Informetrie: ein neuer Name für eine neue Disziplin. In: Nachrichten für Dokumentation 30, Nr. 6, S. 219-226.

Newman, M. E. J. (2005): Power laws, Pareto distributions and Zipf's law. In: Contemporary Physics 46, Nr. 5, S. 323-351. URL: http://arxiv.org/abs/cond-mat/0412004.

Umstätter, Walther (2005): Anmerkungen zu Birger Hjørland und Jeppe Nicolaisen: Bradford's Law of Scattering: Ambiguities in the Concept of „Subject". In: LIBREAS, Nr. 3. URL: http://www.ib.hu-berlin.de/~libreas/libreas_neu/ausgabe3/008ums.htm.

Vickery, Brian C. (1948): Bradford's law of scattering. In: Journal of Documentation 4, Nr. 3, S. 198-203.

Wagner-Döbler, Roland (1997): Time dependencies of Bradford distributions: structures of journal output in Twentieth century logic and Nineteenth century mathematics. In: Scientometrics 39, Nr. 3, S. 231-252.

White, Howard D. (1981): ‚Bradfordizing' search output: how it would help online users. In: Online Review 5, Nr. 1, S. 47-54.

Wilkinson, E. A. (1972): The ambiguity of Bradford's law. In: Journal of Documentation 28, Nr. 2, S. 122-130.

Worthen, D. B. (1975): The application of Bradford's law to monographs. In: Journal of Documentation 31, Nr. 1, S. 19-25.

Empirische Untersuchung, Informetrie, Zeitschrift, Literaturrecherche, Qualität

Der Autor

Dr. Philipp Mayr

ist derzeit Vertretungsprofessor für Informationswissenschaft mit Schwerpunkt Wissenrepräsentation an der Hochschule Darmstadt. Er studierte Bibliotheks- und Informationswissenschaft, Informatik und Soziologie an der Humboldt-Universität zu Berlin. Seit Ende 2004 arbeitet er am GESIS – Leibniz-Institut für Sozialwissenschaften in Bonn. Seit seiner Promotion im Jahr 2009 arbeitet er als Post-Doc und Teamleiter u.a. im DFG-Projekt „Value-Added Services for Information Retrieval". Zu seinen Forschungsschwerpunkten gehören alle Aspekte des Information Retrieval, Modelle digitaler Bibliotheken und Webometrie.

GESIS – Leibniz-Institut für Sozialwissenschaften
Lennéstraße 30, 53113 Bonn, philipp.mayr@gesis.org, www.ib.hu-berlin.de/~mayr/

Neue Umschriftnormen aus dem DIN

Der zuständige Arbeitsausschuss „Transliteration und Transkription" im Normenausschuss Bibliotheks- und Dokumentationswesen im DIN hat zwei weitere Projekte abgeschlossen und legt als Ergebnis seiner Arbeit Umschriftnormen für das hebräische und das arabische Alphabet in einer vollständigen Überarbeitung vor.

Im Januar 2011 ist **DIN 31636, Information und Dokumentation – Umschrift des hebräischen Alphabets** erschienen. Hier hatte insbesondere der teilweise fehlerhafte Druck der hebräischen Buchstaben in der Vorgängerausgabe eine Überarbeitung notwendig gemacht, die für weitere Verbesserungen genutzt wurde. So wurden die Erläuterungen und Beispiele gründlich überarbeitet und den hebräischen Buchstaben in den Umschrifttabellen die Unicode-Werte beigegeben.

Die Norm brücksichtigt die aus der jahrzehntelangen Anwendung der DIN 31636 gewonnenen Einsichten und Desiderata und ist deshalb besonders anwenderfreundlich. Sie ist in der Wiedergabe der Grundbuchstaben des Alphabets identisch mit der Transliterationstabelle der American Library Association/Library of Congress (ALA-LC) und übernimmt auch die dort angewandten Transliterationsregeln des Hebräischen und Jiddischen. Damit trägt die revidierte Norm vor allem den Entwicklungen des modernen Hebräisch Rechnung und strebt mit der Übernahme der anglo-amerikanischen Regeln die Anbindung an eine internationale Standardisierung an, welche den zukünftigen Datenaustausch wesentlich vereinfachen wird.

Ebenfalls neu erscheinen wird in wenigen Wochen eine Neuausgabe der **DIN 31635, Information und Dokumentation — Umschrift des arabischen Alphabets für die Sprachen Arabisch, Osmanisch-Türkisch, Persisch, Kurdisch, Urdu und Paschtu.** Die Erstellung dieser Norm war besonders aufwändig, da sie aktuelle, teilweise auch wechselnde sprachliche Entwicklungen in verschiedenen Ländern berücksichtigen musste.

Die bisherige Norm wurde für Arabisch und Persisch geringfügig, für Osmanisch-Türkisch, Kurdisch und Paschtu jedoch grundsätzlich revidiert. Die Norm wurde für Urdu nur hinsichtlich der Umschrifttabelle geändert.

Zur Verbesserung der Anwenderfreundlichkeit wurde die Norm neu strukturiert, indem die Umschriften der verschiedenen Sprachen so getrennt wurden, dass jede Sprache in einem eigenen Abschnitt behandelt wird. Die Hinweise zur Transliterationsarbeit bzw. Umschriftvorgaben wurden für jede Sprache speziell angepasst und teilweise erweitert. Wie bei allen Normen des Arbeitsausschusses wurdn den Umschrifttabellen die Unicode-Werte der arabischen und zusätzlich der mit diakritischen Zeichen versehenen lateinischen Buchstaben hinzugefügt.

Die Norm bezogen auf Arabisch und Persisch beruht weitgehend auf der 1935 vom Internationalen Orientalistenkongress angenommenen Empfehlung. Die für diese Sprachen vorgesehenen Umschrifttabellen sind fast identisch mit der Transliterationstabelle der ISO 233:1984. Die Norm unterscheidet sich von dieser aber in einzelnen Zusatzbestimmungen, die ausführlich erläutert werden.

Die Umschrifttabelle im Abschnitt für das Arabische ist auf die klassische und auf die moderne arabische Schriftsprache anwendbar. Für die Umschrift des Osmanisch-Türkischen, das in der Türkei bis 1928 Verwendung fand, wird auf eine Umschrifttabelle verzichtet und stattdessen eine Empfehlung für bestimmte Referenzwerke gegeben. Der Abschnitt Persisch gilt allgemein für die Staatssprache Irans wie für das Dari, das neben dem Paschtu offizielle Sprache in Afghanistan ist. Alle Dokumente in der pakistanischen Staatssprache wie der in Indien in der Regel Hindustani genannten Sprache sind nach der Vorgabe in Abschnitt 9 „Urdu" zu transliterieren.

Beide Normen sind als Druckausgabe und/oder E-Book beim Beuth-Verlag (www.beuth.de) erhältlich.

Rita Albrecht
Obfrau des Arbeitsausschusses Transliteration und Transkription

BibEval – Ein webbasierter Kriterienkatalog zur Usability-Evaluation von Bibliothekswebsites

Thomas Weinhold, Sonja Öttl, Bernard Bekavac, Chur (Schweiz)

Heuristiken sind ein häufig genutztes Mittel zur Usability-Evaluation von Bibliothekswebsites, wobei ihr Einsatz in der Regel ein hohes Maß an fachspezifischem Knowhow erfordert. Ohne fundierte Usability-Kenntnisse ist es daher kaum möglich, auf der Basis von Heuristiken oder auch anderer Normen und Guidelines detaillierte Hinweise zur Optimierung bibliotheksspezifischer Online-Angebote abzuleiten. Das Schweizerische Institut für Informationswissenschaft (SII) hat daher im Rahmen des Innovationsprojektes „E-lib.ch: Elektronische Bibliothek Schweiz" einen speziell auf die Bedürfnisse von Bibliotheken zugeschnittenen Kriterienkatalog („BibEval") entwickelt, mit dessen Hilfe Bibliotheken selbstständig eine Usability-Evaluation ihres Webauftrittes vornehmen können und der auch von Usability-Experten als Leitfaden genutzt werden kann, um Bibliothekswebseiten effizient auf Schwachstellen überprüfen zu können. Der Leitfaden BibEval ist modular aufgebaut und ermöglicht dadurch sowohl umfassende Analysen aller Online-Angebote einer Bibliothek als auch die gezielte Evaluation ausgewählter Komponenten bzw. Bereiche der Website. Umgesetzt wurde er in Form einer interaktiven Webanwendung, welche auf den Webseiten der Churer Evaluationslabors zur freien Verwendung zur Verfügung steht. BibEval erlaubt es interessierten Institutionen, sich einen auf ihre Bedürfnisse angepassten Fragenkatalog für Evaluationen zusammenstellen und die Resultate einer Evaluation wahlweise in Form einer CSV-Datei oder eines PDF-Dokumentes exportierten zu können.

BibEval – a web-based criteria catalogue for the evaluation of the usability of library websites
Heuristics are a widely used instrument to assess the usability of library websites. However, such heuristics are usually rather abstract, so that it is difficult for evaluators without the appropriate background knowledge to derive detailed information on how to optimize library specific online services on that basis. Therefore, in the context of the project "E-lib.ch: Swiss Electronic Library" the Swiss Institute for Information Research (SII) has developed a criteria catalogue ("BibEval") which is specifically tailored to the needs of libraries and thus can be used for self-evaluations by the library staff. This criteria catalogue uses a modular structure, so that it can be used for comprehensive evaluations of a library's entire online services as well as for the evaluation of only selected areas of a website. BibEval is available in form of an interactive web application, which can be used by interested institutions to create a customized list of questions for their usability evaluations, at which the results of an evaluation can either be exported as a CSV file or in form of a PDF document.

1 Einleitung und Motivation

Wissen wird heute nicht mehr primär nur in gedruckter Form zur Verfügung gestellt, sondern immer häufiger zusätzlich oder sogar auch ausschließlich über elektronische Informationssysteme. Bedingt durch die immer stärker fortschreitende Digitalisierung von Wissen ändert sich nicht zuletzt auch die Rolle der Bibliotheken. Neue Technologien wie E-Book-Reader oder aber auch soziokulturelle Phänomene, wie die wachsende Bedeutung von Social-Networking-Anwendungen, verändern das Umfeld, in dem Bibliotheken agieren, grundlegend. Für Bibliotheken gilt es sich diesen neuen Gegebenheiten anzupassen, um langfristig erfolgreich bestehen zu können. So überrascht es nicht, dass der Branchenverband IFLA (International Federation of Library Associations and Institutions) in seiner strategischen Planung Bibliotheken und Informationsdienstleister dazu aufruft, gemeinsam mit Partnern und Anwendern daran zu arbeiten, die Potentiale digitaler Technologien noch besser auszuschöpfen, um Nutzern einen nahtlosen und offenen Zugang zu Kulturgütern und Informationsressourcen zu ermöglichen (IFLA 2010, 2).

Im Zuge dieser Entwicklungen gewinnen die Online-Angebote von Bibliotheken beständig an Relevanz. Sie ermöglichen es den jeweiligen Institutionen, ihren Benutzern eine Fülle von Informationen und zusätzlichen Dienstleistungen zur Verfügung zu stellen, unabhängig von den eigentlichen Öffnungszeiten. Für den Anwender findet hierbei eine zunehmende Verschmelzung zwischen der Bibliothek und deren Online-Auftritt statt: Ist der Webauftritt attraktiv gestaltet, so kann dies eine positive Wahrnehmung der gesamtem Bibliothek bewirken. Führt hingegen die Website den Nutzer nicht möglichst schnell und einfach zum gewünschten Ziel, so besteht die Gefahr, dass nicht nur der Webauftritt ein negatives Image erhält, sondern vielmehr die Institution als Ganzes (Puscher 2001, 29). Aus diesem Grund sehen sich Bibliotheken mit der Herausforderung konfrontiert, eine große Menge an Informationen in übersichtlicher und einfach handhabbarer Form bereit zu stellen, was neben etablierten bibliothekarischen Tätigkeitsbereichen ein neues, Technologie- und innovationsorientiertes Aufgabenportfolio mit sich bringt.

In den letzten 20 Jahren haben sich OPACs (Online Public Access Catalogues) zum wichtigsten Recherchein-strument in Bibliotheken entwickelt, weshalb diese eine Schlüsselrolle bei der Entwicklung von nutzerzentrierten Bibliothekswebseiten spielen. Doch während durch diese Systeme die Zugangsmöglichkeiten zu den Bibliotheksbeständen unmittelbar verbessert wurden, wurde die Anwenderfreundlichkeit der Online-Kataloge leider lange vernachlässigt. Der Aufbau bzw. die Umsetzung von OPACs ist als Kompromiss zwischen Ansprüchen der Nutzer und bibliothekarischen Anforderungen der Erschließung anzusehen. Die angebotenen

Suchmöglichkeiten, die stark auf die in Zusammenhang mit der Katalogisierung notwendigen Metadaten ausgerichtet sind, sind daher für viele Anwender nur schwer verständlich (Ronthaler/Zillmann 1998, 1203). Es lässt sich festhalten, dass OPACs eher im Hinblick auf erfahrene Nutzer entwickelt wurden, die mit dem Umgang solcher Systeme vertraut sind. In der Regel sind Recherchen in solchen Online-Katalogen daher nur erfolgreich, wenn der Suchende genau weiß, was er benötigt und auch in der Lage ist, dieses Informationsbedürfnis entsprechend in der Sprache des Anfragesystems auszudrücken (Borgmann 1996, 494). Auch wenn es 1996 praktisch noch kaum Webbasierte OPACs gab[1], so haben die Aussagen von Borgmann insofern teilweise auch heute noch Gültigkeit, als dass sich das Bedienkonzept dieser Kataloge mit der Migration hin zu web-basierten Oberflächen kaum verändert hat. Erst in den letzten Jahren sind bedingt durch die Verfügbarkeit neuer Technologien, wie z.B. AJAX (Asynchronous JavaScript and XML), dynamischere und reichhaltigere Benutzungsschnittstellen entstanden. Beispielsweise haben Bibliotheken Module für die Rechtschreibkorrektur oder die Erweiterung von Suchanfragen in ihre Anwendungen integriert (Gozali/Kann 2007, 329).

Zudem beschränken sich Bibliotheken in ihren Webauftritten immer weniger auf den Katalog und zusätzliche Informationsseiten, sondern sie integrieren zusätzlich externe Daten (Linked Data) in ihre Angebote und nutzen Social Communities (z.B. Facebook und Twitter), um damit Präsenz im Web 2.0 zu zeigen, was insbesondere mit Hinblick auf die Attraktivität für ein junges Zielpublikum von Bedeutung ist.

Momentan mangelt es allerdings noch an empirischen Daten, die belegen welche dieser Komponenten den Anwendern wirklich einen Mehrwert liefern. Über reine Nutzungsstatistiken lässt sich der Nutzen solcher Angebote kaum definieren, da beispielsweise Hilfefunktionen in der Regel erst dann genutzt werden, wenn Probleme auftreten, wobei sie aber in solchen Situationen für den Nutzer unerlässlich sind. Insgesamt mangelt es insbesondere noch an klaren und übertragbaren Konzepten und Strategien, wie einzelne Funktionalitäten und Dienste in die Bibliothekswebseiten zu integrieren sind. Daraus resultieren neue Herausforderungen für die Bibliotheken, und zwar sowohl im Hinblick auf die Implementierung entsprechender Angebote als auch für deren Evaluation.

1 der Zugang zu diesen Katalogen erfolgte damals primär noch über Telnet-Verbindungen.

2 Merkmale benutzerzentrierter Bibliothekswebsites

Aus Anwendersicht gibt es drei Aspekte, die für die Qualität einer Bibliothekswebsite zentral sind: die angebotenen Funktionalitäten, die Usability einer Webseite und deren Usefulness.

Zunächst ist das Portfolio an Interaktionsmöglichkeiten, welche dem Anwender zur Verfügung stehen, entscheidend dafür, ob der Anwender seine individuellen Ziele angemessen erreichen kann. Hierbei ist darauf zu achten, dass die Website nicht mit allen nur denkbaren Features überladen wird. Vielmehr muss die Auswahl der angebotenen Dienste entsprechend den Nutzern und deren Aufgaben vorgenommen werden. Hierfür sind profunde Kenntnisse über den Anwenderkreis der eigenen Angebote und deren Informationsbedürfnisse erforderlich. Dieses Wissen ist Grundvoraussetzung um Systeme implementieren zu können, die eine möglichst optimale Aufgabenunterstützung bieten (Battleson et al. 2001, 188).

Neben der Ausrichtung an den Aufgaben der Anwender, ist es von zentraler Bedeutung, dass eine Website einfach zu handhaben ist. Alle angebotenen Funktionalitäten müssen intuitiv, effizient und mit einer geringen Fehlerquote zu bedienen sein. Sofern dies nicht gewährleistet ist, besteht das Risiko, dass Anwender die Nutzung dieser Tools zu vermeiden versuchen (Yushiana/Rani 2007, 538). Aus diesem Grund fällt der Usability eine Schlüsselrolle bei der Entwicklung qualitativ hochwertiger Websites zu. Dieser Aspekt ist insbesondere dann von Bedeutung, wenn eine Differenzierung zu Wettbewerbern nicht über die auf der Website angebotenen Funktionalitäten möglich ist.

Der dritte Aspekt bezieht sich auf die Zweckmäßigkeit bzw. die Usefulness einer Website. Innerhalb der HCI-Community (Human-Computer-Interaction) werden Usability und Usefulness als integrale Bestandteile eines ganzheitlichen Ansatzes betrachtet. Im Kontext dieses Beitrags bzw. in Zusammenhang mit Bibliotheken ist der Begriff der Usefulness allerdings ein wenig anders zu interpretieren, wobei darunter die Fähigkeit der Bibliothek zu verstehen ist, die Anwender mit für ihre Informationsbedürfnisse relevanter Literatur zu versorgen. Um eine optimale Informationsversorgung sicherzustellen, ist es erforderlich, den Anwendern unterschiedliche Zugänge zu den Beständen anzubieten. Dies bedeutet, dass beispielsweise sowohl eine zielgerichtete Suche nach bestimmten Objekten möglich sein muss, als auch ein Browsing im Bestand, sofern nur ein vages Informationsbedürfnis vorliegt. In Bezug auf die Zweckmäßigkeit bzw. die Usefulness eines Angebotes besteht also ein enger Bezug zum Begriff der Relevanz bzw. zum Ranking der Rechercheergebnisse. Da die Relevanz von Dokumenten jedoch ein subjektives Kriterium ist, stellt die Usefulness einen Aspekt dar, welcher nur schwer zu bewerten ist. Aus diesem Grund wird dieses Kriterium in der vorliegenden Arbeit nicht näher beleuchtet. Der Fokus liegt vielmehr ausschließlich auf funktionalen Aspekten sowie der Usability der entsprechenden Dienste.

Im Zusammenhang mit der Usability von Websites, insbesondere bei öffentlichen Institutionen, spielt auch das Thema Barrierefreiheit eine wichtige Rolle. Dieser Aspekt sollte ohnehin im Rahmen einer Usability-Evaluation mit den dafür etablierten Methoden[2] geprüft werden und wird daher in dem hier beschriebenen Kriterienkatalog nicht explizit berücksichtigt.

3 Usability-Evaluationsmethoden

Mittlerweile existiert eine Vielzahl unterschiedlicher Methoden, die für Usability-Evaluationen herangezogen werden können. Eine mögliche Art der Unterscheidung der einzelnen Methoden basiert auf dem Zeitpunkt, zu dem die Evaluation durchgeführt wird. Dabei unterscheidet man zwischen formativen und summativen Verfahren. Die formative Evaluation findet bereits während der Entwicklung eines Produktes statt, wobei konkrete Verbesserungsmöglichkeiten zur Optimierung der Benutzeroberfläche im Vordergrund stehen. Hierbei werden insbesondere qualitative Daten erhoben, also beispielsweise verbale Aussagen oder Beschreibungen anhand von Beobachtungen, die der Identifikation von Nutzungsproblemen dienen. Formative Evaluationen sollten in der Regel in frühen Phasen der Produktentwicklung angewendet werden, da hierbei der Spielraum für Änderungen am Produkt noch am größten ist (Burmester 2003, 99). Summative Evaluationen erfolgen dagegen meist nach Abschluss der Produktentwicklung, da durch diese Verfahren hauptsächlich Urteile und Gesamteinschätzungen ermittelt werden (Nielsen 1993, 170). Dabei werden quantitative Daten gesammelt, wie beispielsweise Angaben zur Zeit, die Testpersonen für die Erledigung einer Aufgabe benötigen, die Anzahl der Fehler, die sie dabei machen, oder wie häufig sie gewisse Fehler wiederholen. Diese Maßzahlen haben den Vorteil, dass sie objektiv gemessen und ausgewertet werden können (Dumas/Redish 1999, 184).

2 Nähere Informationen s. in IWP 56 (8/2005) Schwerpunktheft „Barrierefreiheit im Internet“. Wiesbaden: Dinges & Frick. ISSN 1434-4653.

Ein weiteres Unterscheidungskriterium basiert darauf, welche Art von Gutachtern für die Evaluation herangezogen wird. Man unterscheidet zwischen den expertenorientierten oder analytischen Methoden und den benutzerorientierten bzw. empirischen Methoden. Bei den benutzerorientierten Methoden der Usability-Evaluation werden künftige, reale Benutzer mit dem Produkt konfrontiert und während der Interaktion mit diesem Produkt beobachtet (Schweibenz & Thissen 2003, 43). Die Beobachtung der Nutzer ermöglicht Rückschlüsse über die Stärken und Schwächen der untersuchten Produkte. Dagegen werden bei der Verwendung von expertenorientierten bzw. analytischen Methoden Experten als Ersatz-Benutzer für die Untersuchung herangezogen. Diese untersuchen Benutzungsschnittstellen mit Hilfe vorgegebener Richtlinien und anhand ihrer Erfahrung auf mögliche Usability-Probleme (Heuer 2003, 116). Expertenorientierte Verfahren lassen sich bereits mit einer geringen Anzahl von Gutachtern durchführen. Diese Methoden haben daher den Vorteil, dass sie relativ schnell und flexibel durchführbar und auswertbar sind. Da diese Methoden nicht so strengen formellen Anforderungen unterliegen wie die benutzerorientierten Methoden, werden sie häufig auch als informelle Methoden der Evaluation bezeichnet (Schweibenz & Thissen 2003, 76).

Die wohl bekannteste analytische Evaluationsmethode ist die so genannte heuristische Evaluation, bei der ein Abgleich der Elemente einer Benutzungsoberfläche mit allgemein anerkannten Gestaltungsprinzipien vorgenommen wird (Nielsen 1994, 26). Basierend auf diesen Richtlinien untersuchen ein oder mehrere Gutachter ein Produkt hinsichtlich vorhandener Usability-Probleme. Da ein Produkt nicht immer von allen Gutachtern gleich bewertet wird, werden bei dieser Methode die besten Ergebnisse dann erzielt, wenn mehrere Gutachter das entsprechende Produkt zunächst unabhängig voneinander untersuchen und anschließend diese Ergebnisse einer gemeinsamen Diskussion und Konsolidierung unterzogen werden (George 2008, 131).

Da Usability-Tests zeitaufwändig und daher kostenintensiv sind, stellt der Einsatz von Heuristiken eine effektive und ökonomische Alternative dar, um die Usability-Mängel einer Website zu identifizieren. Ein weiterer Vorteil von Heuristiken ist der Umstand, dass diese bereits in einem frühen Entwicklungsstadium gewinnbringend eingesetzt werden können (Lindgaard 1994, 125). Aus diesem Grund stellen Heuristiken ein häufig genutztes Instrument zur Evaluation von Websites im Allgemeinen und speziell auch für Online-Angebote aus dem Bibliotheksbereich dar. So präsentieren bspw. Manzari/Trinidad-Christensen (2006) eine Studie, in der eine Kombination von heuristischer Evaluation und formativem Usability-Test für ein interaktives Redesign der Website der Bibliothek der Long Island University verwendet wurde. Aitta et al. (2007) haben die klassischen Heuristiken von Nielsen (1994) für eine Evaluation der Websites von 15 öffentlichen Bibliotheken herangezogen. Dieselben Heuristiken wurden von Yushiana/Rani (2007) verwendet um den webbasierten OPAC einer akademischen Bibliothek zu evaluieren.

In diesem Zusammenhang ist jedoch anzumerken, dass die in gängigen Heuristiken enthaltenen Kriterien, wie z.B. die zehn Bewertungsaspekte, die Nielsen (1994, 30) in seinen Heuristiken präsentiert, recht abstrakt gehalten sind. Aus diesem Grund ist ein bestimmtes Maß an Hintergrundwissen aus den Bereichen Usability-Engineering und Web- bzw. Interaction Design erforderlich, um auf deren Basis eine detaillierte Analyse bestimmter Komponenten einer Website vornehmen zu können.

Um hingegen auch Personen ohne umfassende Erfahrung in der Durchführung von Usability-Evaluationen die Möglichkeit zu geben, eine Analyse der Schwachstellen bibliotheksspezifischer Online-Angebote vornehmen zu können, sind detailliertere Kriterienkataloge erforderlich, die möglichst optimal auf die Branchenspezifika im Bibliotheks- und Informationssektor zugeschnitten sein sollten. Ein solcher Kriterienkatalog, der von den Autoren im Rahmen des Kooperations- und Innovationsprojektes „E-lib.ch: Elektronische Bibliothek Schweiz“ entwickelt wurde, wird nachfolgend vorgestellt. Dabei wird zunächst auf die Konzeption des Kriterienkatalogs eingegangen und darauf aufbauend erfolgt anschließend die Präsentation, der auf Grundlage des Kriterienkatalogs implementierten Webanwendung „BibEval“, mittels derer interessierte Institutionen eine Evaluation ihrer Online-Angebote vornehmen können.

4 Konzeption des Kriterienkatalogs

Die Konzeption des in diesem Beitrag beschriebenen Kriterienkatalogs basiert auf drei unterschiedlichen Säulen. In einer frühen Phase des Projekts wurde eine umfassende Literaturanalyse hinsichtlich Usability-Evaluationen in Bibliotheken durchgeführt, deren Ziel es war, geeignete Kriterien für die Beurteilung solcher Websites zu identifizieren. Zusätzlich wurde eine Best-Practice-Analyse von Bibliothekswebsites vorgenommen. Der Zweck dieser Analyse war es, einen besseren Eindruck des gegenwärtigen State-of-the-Art sowie Informationen über die Verbreitung und die Ausgestaltung bestimmter Funktionalitäten bzw. Website-Komponenten zu gewinnen. Basierend auf diesen Erkenntnissen wurde eine erste Version des Kriterienkatalogs erstellt. Anschließend wurde dieser Entwurf in einer Fokusgruppe, an der Experten aus den Bereichen Bibliothek, Webdesign und Usability-Engineering teilnahmen, diskutiert und weiter verfeinert.

Das Resultat dieser Bemühungen ist eine modular verwendbare, hierarchisch strukturierte Liste von Evaluationskriterien. In diesem Kontext bedeutet modular, dass die Autoren auf Basis der Best-Practice-Analyse eine Unterscheidung vorgenommen haben, welche Bestandteile bzw. Funktionalitäten einer Website für die Nutzer unverzichtbar sind und welche eher einen optionalen Charakter aufweisen (kategorisiert als „obligatorisch“ und „optional“). Das Ziel dieser Kategorisierung ist es, die Anwendbarkeit des Kriterienkatalogs für Bibliotheken unterschiedlicher Größe und unterschiedlichen Typs zu maximieren. Während sich dadurch kleine Institutionen mit geringen Ressourcen auf die Evaluation der wichtigsten Features konzentrieren können, haben große Bibliotheken anhand des Kriterienkatalogs trotzdem die Möglichkeit eine umfassende und detaillierte Analyse ihrer sämtlichen Online-Angebote vorzunehmen.

5 Struktur und Aufbau des Kriterienkatalogs

Um die zuvor beschriebene Anpassbarkeit und Flexibilität des Kriterienkatalogs zu gewährleisten, wurde für diesen eine hierarchische Grundstruktur gewählt, unterteilt in die drei Ebenen Sektionen, Subsektionen und Komponenten (vgl. Abbildung 1).

Auf der obersten Hierarchieebene wird zwischen vier Sektionen unterschieden, welche an den potentiellen Zielen des Anwenders ausgerichtet sind: „Information & Kommunikation“ deckt alle Aspekte bezüglich der Verbreitung von Informationen und des Anwendersupports ab, also all jene Informationen, welche der Nutzer über die Bibliothek und deren Dienstleistungen einholen kann, wie beispielsweise Öffnungszeiten der Bibliothek, einen Newsletter oder Beratung über einen Live-Chat-Service. „Recherchieren im Bestand“ beinhaltet alle Funktionalitäten, mit denen die Anwender konfrontiert werden, wenn diese nach Bibliotheksobjekten (z.B. Bücher, Zeitschriften, CDs, Karten, etc.) recherchieren und die Ergebnisse auswerten möchten. Innerhalb der Sektion „Personalisierung“ sind alle Funktionalitäten zusammengefasst, die es den Nutzern erlauben, die Website bzw. deren Dienste/Einstellungen an ihre eigenen Präferenzen anzu-

passen. „Nutzerpartizipation“ steht in engem Zusammenhang zum Begriff Web 2.0 und umfasst alle Funktionen, die es dem Nutzer ermöglichen, Inhalte zu bewerten, selbst zu erfassen und mit anderen auszutauschen (Linh 2008, 632).
Um eine gezieltere Eingrenzung innerhalb dieser Sektionen treffen zu können, wurden diese auf der zweiten Ebene in Subsektionen untergegliedert. So wird innerhalb der Sektion „Information & Kommunikation“ beispielsweise zwischen den Subsektionen „Kontakt & Zugang“, „Seitenüberblick“, „News & Veranstaltungen“ und „Anleitungen“ unterschieden, während die Sektion „Recherche im Bestand“ in „Suchen & Erkunden“ sowie „Präsentation & Zugriff“ unterteilt ist.
Auf der dritten Ebene sind nun die einzelnen Komponenten einer Website, wie beispielsweise ein Kontaktformular oder die einfache Suche angesiedelt. Die Zuordnung einer Komponente zur entsprechenden Sektion bzw. Subsektion wurde dabei mit Hinblick auf ihren hauptsächlichen Verwendungszweck vorgenommen. In den meisten Fällen war diese Zuteilung relativ einfach zu bewerkstelligen, wie bspw. bei der Zuordnung der einfachen Suche zur Subsektion „Suchen & Erkunden“ bzw. zur Sektion „Recherche im Bestand“. Es gab jedoch auch einige Komponenten, für die keine eindeutige Klassifikation möglich war. So können ein Blog oder ein Wiki, in Abhängigkeit der jeweiligen Inhalte entweder als Instrumente, die ausschließlich Informations- und Kommunikationszwecken dienen interpretiert werden oder aber auch als Tools, die eine interaktive Beteiligung der Bibliotheksnutzer ermöglichen. In diesem Fall hängt die Einordnung von organisatorischen Kriterien und der Konfiguration dieser Systeme ab (können bspw. nur Bibliotheksangestellte neue Beiträge verfassen oder ist dies auch der Öffentlichkeit möglich).
Neben dieser funktionalen Zuordnung erfolgte zusätzlich eine Unterscheidung zwischen obligatorischen und optionalen Komponenten. Diese Unterteilung basierte auf den Einschätzungen der Autoren bezüglich der Relevanz einer Komponente aus Anwendersicht und dem Verbreitungsgrad der jeweiligen Features (entsprechend den Resultaten der Best-Practice-Analyse). So wurde beispielsweise im Rahmen der Sektion „Information & Kommunikation“ die Bereitstellung der Öffnungszeiten der Bibliothek als verpflichtend eingestuft, während die Verfügbarkeit von sozialen Netzwerken (z.B. Facebook) oder Microblogging-Diensten (z.B. Twitter) als optional bewertet wurden.

Zu jeder der drei Ebenen wurden Fragen entwickelt, deren Detaillierungsgrad von oben nach unten kontinuierlich zunimmt. Während die Fragen zu den Sektoren einen starken Bezug zu klassischen Heuristiken aus dem Bereich des Webdesign haben (dabei werden z.B. Fragen bezüglich der Konsistenz der Darstellungen und der Einhaltung von gängigen Standards untersucht), gehen die Fragen zu den Subsektoren bereits konkreter auf den Funktionsumfang ein. Auf der Ebene der Komponenten werden schließlich detaillierte Fragen zur Verfügung gestellt, die eine umfassende Bewertung der Funktionsweise bzw. das Aufspüren von konkreten Schwachstellen erlauben. Durch diese modulare Aufteilung steht es den Anwendern frei, entweder eine allgemeine Bewertung der Webseite vorzunehmen, indem lediglich die Fragen zu den Sektoren und Subsektoren berücksichtigt werden, oder gezielt einzelne Funktionalitäten im Detail zu untersuchen.
In Analogie zu den Komponenten wurde auch bei den Evaluationskriterien eine Unterteilung in optionale und obligatorische Fragen vorgenommen. So ist die Frage „Kann die Suche sowohl durch das Drücken eines dedizierten Buttons als auch per „Enter“ gestartet werden?“ beispielsweise als obligatorisch deklariert, wohingegen „Wird eine Suchhistorie angeboten, mittels der der Anwender frühere Suchanfragen wiederholt durchführen kann?“ als optionale Frage eingestuft wurde. Hierbei wurde die Einstufung der Komponente nicht berücksichtigt, da auch eine optionale Komponente, wie z.B. das Angebot eines Microblogging-Dienstes, sofern sie auf einer Website angeboten wird, obligatorische Kriterien aufweist, welche für die Nutzerfreundlichkeit der Komponente unumgänglich sind.
Um die Auswirkungen der ermittelten Usability-Probleme einschätzen zu können, bietet es sich an, ein so genanntes Severity Rating durchzuführen. Eine solche Kategorisierung kann genutzt werden, um eine optimale Mittelallokation für ein etwaiges Redesign vorzunehmen, indem zunächst die gravierendsten Probleme beseitigt werden. Für diesen Zweck wird in Anlehnung an Nielsen die in Tabelle 1 dargestellte Bewertungsskala herangezogen.

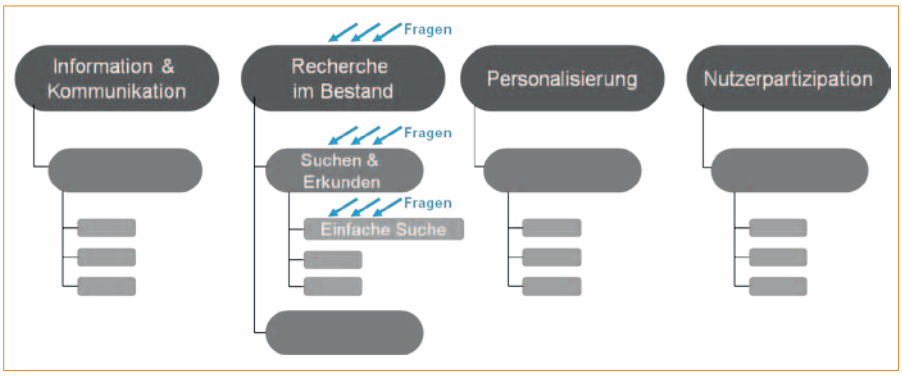


Abbildung 1: Struktureller Aufbau des Leitfadens.

Tabelle 1: Kategorien des Severity-Ratings (in Anlehnung an Nielsen, 1994, S. 49).

Rating	Beschreibung
0	Nicht zutreffend
1	Kein Usability-Problem
2	Geringfügiges Usability-Problem
3	Mittelschweres Usability-Problem
4	Gravierendes Usability-Problem
5	Feature nicht umgesetzt, obwohl notwendig.

6 Aufbau und Funktionsweise der Webanwendung

Um die modulare Struktur des Kriterienkatalogs in der Praxis effektiv nutzen zu können wurde vom SII unter der Bezeichnung „BibEval“ eine interaktive Webanwendung entwickelt, mit deren Hilfe interessierte Institutionen anhand des in Kapitel 5 beschriebenen Kriterienkatalogs eine Evaluation ihrer Web-Angebote vornehmen können. Realisiert wurde „BibEval“ mittels MySQL, PHP und JavaScript bzw. AJAX, wobei die Anwendung sowohl in Deutsch als auch in Englisch genutzt werden kann. Entsprechend dem vorgängig beschriebenen Konzept kann die Struktur der Webanwendung in drei Bereiche gegliedert werden, welche im Folgenden genauer beschrieben werden. Die einzelnen Dialogschritte werden hierbei – unabhängig von ihrer Reihenfolge - dynamisch angepasst, was es dem Anwender erlaubt, Einstellungen jederzeit zu modifizieren. Er erhält somit stets unmittelbares Feedback über die vom ihm getroffenen Auswahlentscheidungen.

Im ersten Dialogschritt bietet „BibEval“ den Anwendern die Möglichkeit eine Auswahl zu treffen, welche Bereiche (Sektionen und Subsektionen) für eine Evaluation berücksichtigt werden sollen (vgl. Abbildung 2). Hierbei wird neben der direkten Auswahl von Sektoren und Subsektoren per Checkboxen auch ein

Abbildung 2: Auswahl der Bereiche, die evaluiert werden sollen.

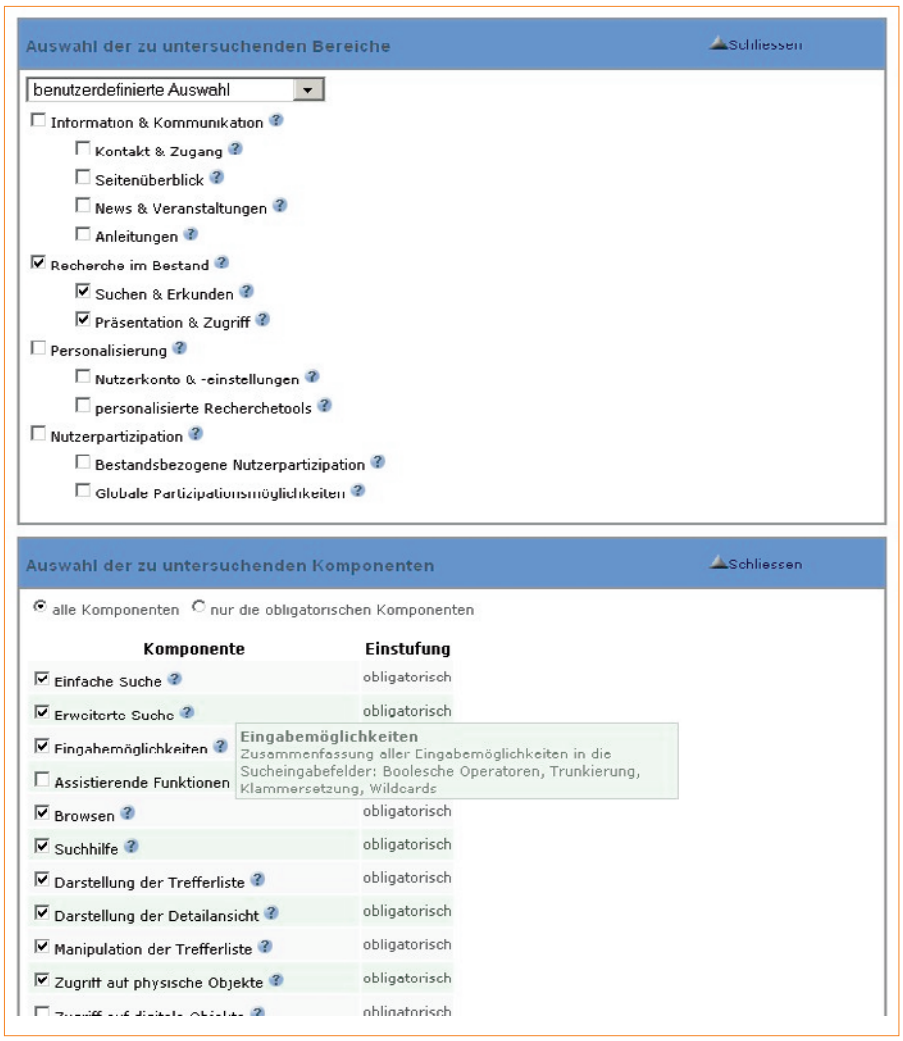


Abbildung 3: Auswahl der zu untersuchenden Komponenten.

Drop-Down-Menü angeboten, über welches unter verschiedenen Gesichtspunkten vorkonfigurierte Zusammenstellungen von Komponenten (und den jeweils übergeordneten Bereichen) ausgewählt werden können. Diese Vorauswahl ist nötig, da zwischen zwei Sichtweisen unterschieden werden muss. Während der Anwender die Webseite unter einer funktions- bzw. zielorientierten Sichtweise betrachtet, weicht eine bibliotheksinterne Perspektive in der Regel von der Nutzersichtweise ab und basiert auf der Unterscheidung zwischen den spezifischen Einzelprodukten wie Katalog, Webseite und gegebenenfalls Fachdatenbanken. Dies beruht auf dem Umstand, dass Bibliotheken teilweise Produkte von externen Anbietern nutzen und somit beispielsweise auf Fachdatenbanken nur verlinken, aber wenig Einfluss auf die Gestaltung des entsprechenden Webseitenbereichs nehmen können. Insgesamt stehen daher folgende Optionen zur Auswahl:

- „alle“ (wobei alle im System hinterlegten Fragen für die Evaluation herangezogen werden)
- „Vorauswahl: nur Katalog“ (es werden nur diejenigen Fragen berücksichtigt, die sich auf den Bibliothekskatalog beziehen)
- „Vorauswahl: nur Webseite“ (evaluiert wird lediglich der eigentliche Webauftritt der Bibliothek)
- „Vorauswahl: nur Fachdatenbanken“ (nur die Funktionalitäten werden evaluiert, die sich auf die Nutzung von über die Seite zugreifbaren Datenbanken beziehen)
- „benutzerdefinierte Auswahl“ (der Anwender kann selbst entscheiden, welche Bereiche und Subsektoren er in seiner Evaluation berücksichtigen möchte)

Entsprechend der Auswahl, die der Anwender in Bezug auf die für die Evaluation zu berücksichtigenden Bereiche getroffen hat, kann er in einem zweiten Bereich der Webanwendung diese Auswahl noch präzisieren (vgl. Abbildung 3). Hierbei bekommt er eine Liste der zugehörigen Komponenten dargestellt, welche gemäß seiner Vorauswahl als relevant erscheinen. In der Zusammenstellung der Komponenten sind dabei zunächst sowohl die obligatorischen als auch die optionalen Komponenten einer Bibliothekswebsite aufgelistet. Der Anwender hat die Möglichkeit, über Radiobuttons die zu untersuchenden Komponenten auf ausschließlich die obligatorischen Angebote einzuschränken. Darüber hinaus ist auch eine individuelle Zusammenstellung der zu untersuchenden Komponenten möglich, wobei die Auswahl der entsprechenden Komponenten über Checkboxen erfolgt. Sofern dem Anwender die Bezeichnungen der einzelnen Rubriken oder

Komponenten unklar sind, kann er jederzeit zusätzliche Informationen als Tooltip abrufen.
Im dritten Bereich von BibEval bekommt der Anwender dann entsprechend seiner Auswahlkriterien seinen individuellen Fragenkatalog inklusive der Anzahl der Fragen präsentiert (vgl. Abbildung 4). Es steht ihm an dieser Stelle frei, den Kriterienkatalog auf die allgemeinen Fragen zu den einzelnen Bereichen oder zu den jeweiligen Komponenten einzugrenzen. Zudem kann er den Kriterienkatalog auf obligatorische Fragen einschränken. Die erklärenden Beschreibungstexte zu den einzelnen Fragen werden hierbei direkt und nicht in Form eines Tooltips eingeblendet, um sicherzustellen, dass alle Anwender unter gleichen Bedingungen testen und keine Informationen übersehen werden könnnen. Der Anwender hat nun die Möglichkeit anhand dieser Kriterienliste einen spezifischen Webauftritt zu untersuchen, wobei die Webanwendung für jede Frage ein entsprechendes Auswahlfeld zur Verfügung stellt, über das eine Bewertung gemäß der in Tabelle 1 dargestellten Fehlerkategorien vorgenommen werden kann. Zusätzlich kann bei Bedarf über einen Button bei jeder Frage auch ein textueller Kommentar hinterlegt werden.
Am Ende der Evaluation hat der Nutzer die Möglichkeit den ausgefüllten Kriterienkatalog zu exportieren, wahlweise in Form eines PDF-Berichtes oder in Form einer CSV-Datei, um gegebenenfalls die Evaluationsergebnisse in einer Tabellenkalkulation weiter verarbeiten zu können. Hierbei werden alle Einstellungen und Kommentare übernommen.

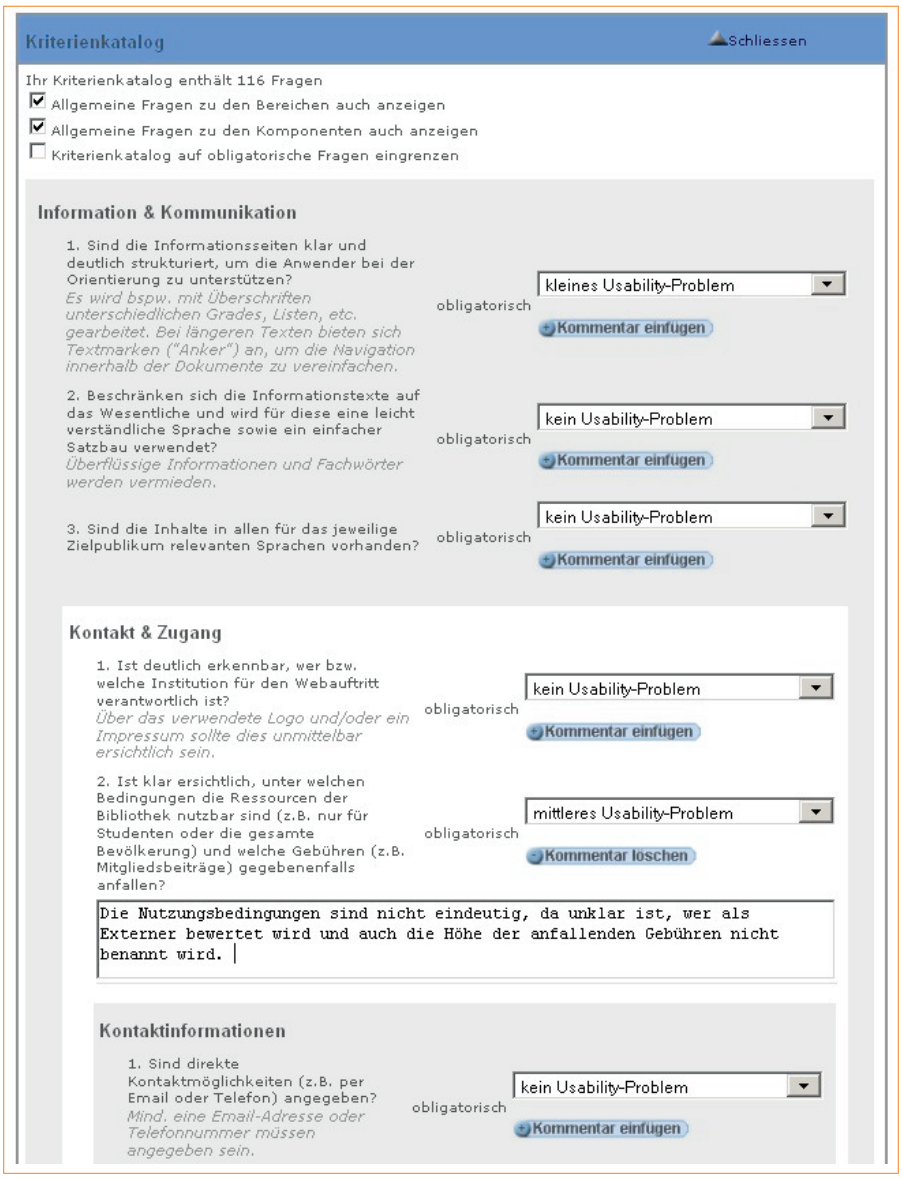


Abbildung 4: Auszug aus einer Kriterienliste.

6 Fazit und Ausblick

Nachdem der in diesem Beitrag beschriebene Kriterienkatalog zunächst einige Zeit nur intern im Rahmen der innerhalb von E-lib.ch durchgeführten Evaluationen eingesetzt und durch die dabei gewonnenen Erfahrungen kontinuierlich weiter verfeinert wurde, steht die entsprechende Webanwendung mittlerweile unter einer Creative Commons-Lizenz auf den Webseiten des Churer Evaluationslabors (CHEVAL) unter www.cheval-lab.ch/leitfaden-bibeval/ öffentlich zur Verfügung. Die Autoren erhoffen sich auf Basis des Feedbacks bzw. der Eindrücke und Erfahrungen, die Anwender mit dem Tool in der Praxis sammeln, den Kriterienkatalog nach und nach noch weiter optimieren zu können. Für eine zielgerichtete Weiterentwicklung und Optimierung werden aktuell zusätzlich auch im Rahmen einer studentischen Arbeit an der HTW Chur der praktische Mehrwert und die Alltagstauglichkeit des Leitfadens überprüft.

In einem nächsten Schritt ist die Implementierung einer Nutzerverwaltung angedacht, die es den für die Durchführung einer Evaluation verantwortlichen Personen erlaubt, alle entsprechenden Voreinstellungen zu treffen und die somit erstellte Kriterienliste gezielt an weitere Gutachter senden zu können. Dabei sollen anschließend die Einzelergebnisse der unterschiedlichen Evaluatoren von der Testleitung ebenfalls über die Webseite abgerufen und konsolidiert werden können.
Neben der in diesem Artikel beschriebenen Möglichkeit zur Durchführung einer expertenbasierten Usability-Evaluation ist darüber hinaus auch die Entwicklung eines speziellen Leitfadens geplant, welcher für benutzerorientierte Evaluationen von Bibliothekswebsites herangezogen werden kann. Denn generell bleibt festzuhalten, dass expertenorientierte Evaluationsverfahren zwar hilfreich sind, um die Schwachstellen einer Website zu identifizieren, sie jedoch die Einschätzungen und Anforderungen realer Benutzer nur bedingt wiedergeben können. Die beiden Leitfäden würden sich somit ergänzen und zusammen ein ideales Instrument zur Usability-Evaluation von Webauftritten im Bibliotheks- und Informationsbereich bilden. Im Rahmen des Benutzerleitfadens sollen dabei neben funktionalen Aspekten (pragmatische Qualität) auch verstärkt Fragen zu der Attraktivität des Webauftrittes und der Stimulation der Anwender durch bestimmte Elemente einer Website (hedonische Qualität) berücksichtigt werden.

Literatur

Aitta, M.R.; Kaleva, S.; Kortelainen, T. (2007): Heuristic evaluation applied to library web services. In: New Library World, 109(1/2), S. 25-45.

Battleson, B.; Booth, A.; Weintrop, J. (2001): Usability Testing of an Academic Library Web Site: A Case Study. In: The Journal of Academic Librarianship, 27(3), S. 188-198.

DABIS.eu
Gesellschaft für Datenbank-InformationsSysteme mbH

Ihr Partner für Archiv-, Bibliotheks- und DokumentationsSysteme

BIS-C 2000

Archiv- und Bibliotheks-InformationsSystem

DABIS.com · alle Aufgaben · ein Team

Synergien: Qualität und Kompetenz
Software: Innovation und Optimierung
Web · SSL · Warenkorb und Benutzeraccount
Lokalsystem zu Aleph-Verbünden

Software · State of the art · Open Source

Leistung	Sicherheit
Standards	Offenheit
Stabilität	Verläßlichkeit
Generierung	Adaptierung
Service	Erfahrenheit
Outsourcing	Support
Dienstleistungen	Zufriedenheit

GUI · Web · Wap · XML · Z 39.50

Archiv		Bibliothek
singleUser	System	multiUser
Lokalsystem		Verbund
multiDatenbank		multiServer
multiProcessing		multiThreading
skalierbar		stufenlos
Unicode		multiLingual
Normdaten		redundanzfrei
multiMedia		Integration

Portale

http://www.vthk.eu
http://www.landesbibliotheken.eu
http://www.bmlf.at
http://www.volksliedwerk.org
http://www.dabis.org:18093

DABIS.com
Heiligenstädter Straße 213
1190 · Wien, Austria
Tel.: +43-1-318 9 777-10
Fax: +43-1-318 9 777-15
eMail: office@dabis.com
http://www.dabis.com

DABIS.de
Herrgasse 24
79294 · Sölden/Freiburg, Germany
Tel.: +49-761-40983-21
Fax: +49-761-40983-29
eMail: office@dabis.de
http://www.dabis.de

Zweigstellen: 61350 · Bad Homburg vdH, Germany / 1147 · Budapest, Hungary / 39042 · Brixen, Italy

Borgman, C.L. (1996): Why are Online Catalogs Still Hard to Use? In: Journal of the American Society for Information Science, 47 (7), S. 493-503.

Burmester, M. (2003): Ist das wirklich gut? Bedeutung der Evaluation für die benutzerzentrierte Gestaltung. In Machate, J.; Burmester, M. (Hrsg.). User Interface Tuning – Benutzungsschnittstellen menschlich gestalten, S. 97-119. Frankfurt: Software und Support.

Dumas, J.; Redish, J. (1999): A Practical Guide to Usability Testing. Exeter: Intellect Books.

George, C.A. (2008): User-Centred Library Websites: Usability evaluation methods. Oxford: Chandos Publishing.

Gozali, J. P.; Kan, M. Y. (2007): A Rich OPAC User Interface with AJAX. In: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries (JCDL'07), S. 329-330.

International Federation of Library Associations and Institutions (2010): IFLA Strategic Plan 2010-2015, http://www.ifla.org/files/hq/gb/strategic-plan/2010-2015.pdf [10.11.2010].

Lindgaard, G. (1994): Usability testing and system evaluation: a guide for designing useful computing systems. London: Chapman & Hall.

Linh, N.C. (2008): A survey of the application of Web 2.0 in Australasian university libraries. In: Library High Tech, 26 (4), S. 630-653.

Manzari, L.; Trinidad-Christensen, L. (2006): User-centered design of a Web site for library and information science students: Heuristic evaluation and usability testing. In: Information technology and libraries, 25 (3), S. 163-169.

Nielsen, J. (1994): Heuristic Evaluation. In: Nielsen, J.; Mack, R.L. (Hrsg.). Usability Inspection Methods, S. 25-62. New York: Wiley.

Nielsen, J. (1993): Usability Engineering. San Diego: Academic Press.

Puscher, F. (2001): Das Usability-Prinzip – Wege zur benutzerfreundlichen Website. Heidelberg: dpunkt. verlag.

Ronthaler, M.; Zillmann, H. (1998): Literaturrecherche mit OSIRIS: Ein Test der OSIRIS-Retrievalkomponente. In: Bibliotheksdienst, 32 (7), S. 1202-1212.

Yushiana, M.; Rani, W.A. (2007): Heuristic evaluation of interface usability for a web-based OPAC. In: Library Hi Tech, 25(4), S. 538-549.

Bewertung, Benutzerfreundlichkeit, Website, Usability, Bibliothek, Checkliste

Die Autoren

Thomas Weinhold

MSc Business Information Systems, Dipl.-Informationswirt (FH)
Wissenschaftlicher Mitarbeiter am Schweizerischen Institut für Informationswissenschaft (SII). Forschungsschwerpunkte: Human-Computer-Interaction, Information Retrieval, Semantic Web.
thomas.weinhold@htwchur.ch

Sonja Öttl

M.Sc, M.A.
Wissenschaftliche Mitarbeiterin am Schweizerischen Institut für Informationswissenschaft (SII). Forschungsschwerpunkte: Informationsvisualisierung, Human-Computer-Interaction, Semantic Web.
sonja.oettl@htwchur.ch

Prof. Dr. Bernard Bekavac

Professor für Information Engineering im Studiengang Informationswissenschaft an der Hochschule für Technik und Wirtschaft (HTW) Chur. Forschungsschwerpunkte: Information/Web Retrieval, Architektur von Informationssystemen, Human-Computer-Interaction.
bernard.bekavac@htwchur.ch

Schweizerisches Institut für Informationswissenschaft (SII), Hochschule für Technik und Wirtschaft (HTW) Chur, Ringstrasse/ Pulvermühlestr. 57, 7004 Chur, Schweiz

Gesamtlösungen für Digitalisierungsprojekte

ImageWare

MyBib eDoc® Workflow-System
BCS-2® Scansoftware
C-3 Periodikaerschließung
MyBib eRoom Elektronischer Lesesaal
ScanRobot® Bookeye® Buchscanner

ImageWare Components GmbH | Am Hofgarten 20 | 53113 Bonn | Germany | www.imageware.de

Erfolgsfaktoren kirchlicher Community-Angebote im Social Web: Eine Analyse des Wikis „Evangelisch in Niedersachsen“

Natascha Pauls, Joachim Griesbaum und Thomas Mandl, Hildesheim

Vorliegender Artikel behandelt anhand einer Untersuchung des Wikis „Evangelisch in Niedersachsen“ die Frage von Erfolgsfaktoren kirchlicher Social Media Angebote. Hierzu werden Bewertungskriterien aus der Literatur erarbeitet und auf die Fallstudie „Evangelisch in Niedersachsen“ angewendet. Dabei werden statistische Nutzungsdaten ausgewertet sowie eine Expertenanalyse und eine Nutzerbefragung durchgeführt. In der konkreten Fallstudie werden die Ergebnisse zur Erarbeitung von Empfehlungen zur Optimierung des Kirchen-Wikis genutzt.

Success factors of ecclesiastical web communities in social media: A case study of the wiki „Evangelisch in Niedersachsen“

This article is based on the analysis of the wiki "Evangelisch in Niedersachsen" (Protestants in Lower Saxony) and deals with the research of key factors of ecclesiastic sites in social media. For this purpose, criteria for an evaluation have been gathered from relevant literature and applied to the case study "Evangelisch in Niedersachsen". Therefore, statistical usage data has been analysed, and an expert analysis as well as a user survey have been conducted. The results of the case study have been used in order to suggest potential improvements of the wiki.

1 Einleitung

Kirche und Social Web, zwei gegensätzliche Konzepte oder doch vereinbar? Folgt man Shirkys (2008) Argumentation, so kann (insbesondere die katholische) Kirche als stark hierarchisierte Organisation mit einem großen Beharrungsvermögen gegen strukturelle Veränderungen begriffen werden. Auf der anderen Seite lässt sich das Social Web als Kommunikationsinfrastruktur bzw. als Prozess verstehen, der in vielen Fällen hierarchische Strukturen auflöst und quasi jedermann eine Stimme verleiht (vgl. auch den Beitrag von Griesbaum in diesem Heft). In einer der konkreten Realität der evangelischen Kirche in Deutschland näher zugewandten Perspektive lässt sich erhebliches Potential der Nutzung von Social Media für kirchliche Kontexte argumentieren. Aktuell verliert die Institution immer mehr Mitglieder. Eine Folge davon ist das Zusammenlegen von Gemeinden bis hin zu Schließungen von Kirchen und Gemeindehäusern oder die Vermietung der Räumlichkeiten. Dadurch wird es wiederum zunehmend schwieriger, Interesse für Religion und Kirche sowie Begeisterung für Gemeindearbeit zu wekken. Hier können Soziale Medien als eine Chance zur Belebung der Gemeinden verstanden werden. Ihr Erfolg in unterschiedlichen Bereichen, im Privaten als auch in der Wirtschaft, hat inzwischen auch die Kirche erreicht. Mittlerweile sind die meisten religiösen Institutionen im Internet vertreten, vgl. (Malik et al. 2007: 203). Einer Studie von Trocholepczy et al. (2009: 36) zufolge, ist „[f]ast jeder zweite Studienteilnehmer […] Mitglied in einer religiösen Gruppe innerhalb bestehender Netzwerke.“ Doch inwieweit können Soziale Medien kirchliche Aktivitäten fördern? Was ist zu beachten, damit diese für ihre Mitglieder nützlich sind und angenommen werden, so dass von einer erfolgreichen Ergänzung und Unterstützung vorhandener Gemeindearbeit gesprochen werden kann?

Der vorliegende Text soll einen Beitrag zur Beantwortung dieser Fragen leisten. Hierzu wird zunächst kurz in den Kontext Kirche und Internet eingeführt, nicht zuletzt, um die besondere Stellung der Kirche, auch in diesem Medium, darzulegen. Vor diesem Hintergrund wird die Konzeption und der Umsetzungsstand eines konkreten Beispiels eines kirchlichen Sozialen Mediums vorgestellt: „Evangelisch in Niedersachsen“, ein Wiki, das im Herbst 2007 von der Evangelisch-lutherischen Landeskirche Hannovers initiiert worden ist. Darauf aufsetzend werden auf Basis einer Literaturanalyse wichtige Erfolgsfaktoren und Kriterien zur Ausgestaltung partizipativer kirchlicher Webangebote diskutiert. Damit soll ein grobes Gesamtbild zu Erfolgsfaktoren kirchlicher Social-Media-Angebote erschlossen werden. Der dabei kompilierte Kriterienkatalog dient wiederum als Orientierungsrahmen zur Untersuchung der oben genannten Fallstudie. Hierzu werden Nutzungsdaten analysiert sowie eine Expertenanalyse und eine Nutzerbefragung durchgeführt. Schließlich werden auf Basis der Ergebnisse Empfehlungen für „Evangelisch in Niedersachsen“ formuliert. Abschließend wird die Untersuchung eingeordnet und ein Ausblick auf mögliche weitere Untersuchungen gegeben.

2 Kirche im Web

Der Begriff Kirche lässt sich in mehrfacher Hinsicht konnotieren: zunächst als ein dem Gottesdienst gewidmeten Gebäude, des Weiteren bedeutet er auch Glaubensgemeinschaft beziehungsweise steht für die gesamte Institution. In letzterer Hinsicht kann Kirche als Non-Profit- und Nichtregierungsorganisation (NGO) eingeordnet werden. In Deutschland sind 63,4 Prozent der Bevölkerung Christen, Protestanten und Katholiken sind dabei nahezu gleichgewichtig vertreten (EKD Hannover 2009). Das Bundesland Niedersachsen weist ca. vier Millionen evangelische Kirchenmitglieder auf. Seit der ersten Internetpräsenz der Evangelischen Kirche Deutschlands (EKD) im Jahre 1996 hat sich das Angebot kirchlicher Präsenzen im Internet stetig erweitert. Allerdings existiert eine Schere zwischen den oftmals professionellen Präsenzen der Landeskirchen und den meist ehrenamtlich betriebenen Webangeboten der Gemeinden vor Ort. Bislang bestanden die vorhandenen Angebote dabei im Wesentlichen aus Informationsangeboten (Haese 2007). Folgt man Brok et al. (2007) so geht es bei solchen Angeboten insbesondere auch darum, Kirche und Glaube auch online leben zu können. D.h. die schlichte Bereitstellung von Informationen ist unangemessen, gefordert

sind Interaktions- und Partizipationsmöglichkeiten, so wie sie in bekannten Communities (Trocholepczy et al. 2009) bzw. in eigens dafür errichteten Web-Kirchen wie der Funcity-Kirche[1] teilweise bereits ausgeübt werden. Eine reine Internetkirche ohne jeglichen Bezug zu einer bestehenden Face-to-Face-Gemeinde wird jedoch als zweifelhaft angesehen. Deshalb wird von Brok et al. (2007) eine Rückbindung an die Gemeinden, d.h. die Kombination aus online- und realweltlicher Gemeinde vorgeschlagen. Dies gestatte es, das Gemeindeleben zu fördern, zumal das Internet heute vereinfachte Möglichkeiten des Zugangs zu spiritueller Praxis bietet. Zudem wird mit ergänzenden Internetangeboten die (Re)Integration in die Gemeinschaft der Kirche vor Ort ermöglicht und gefördert. Während Technologien, welche die Kommunikation und Partizipation unterstützen, kirchlicherseits auf Landesebene bereits nutzbar gemacht werden, sind Social Media auf Gemeindeebene, weitestgehend aus Mangel an personellen und finanziellen Ressourcen, kaum vorhanden. Dabei könnten alle beteiligten Ebenen von einer Vernetzung und gemeinsamem Austausch profitieren.

3 Das Portal „Evangelisch in Niedersachsen“

„Evangelisch in Niedersachsen“ ist ein Wiki, das im Herbst 2007 von der Evangelisch-lutherischen Landeskirche Hannovers initiiert worden ist. Auch als Kirchen-Wiki bezeichnet, dient es wie bei Szugat (2006: 56) vorgeschlagen zur Wissensspeicherung, Kommunikation und „gemeinsame[n] Koordination von Events“, also der Vorbesprechung und Dokumentation von Veranstaltungen. Nach einer etwa einjährigen Pilotphase wurde das Wiki auf der Landessynode am 08.05.2009 vorgestellt und ist seitdem offiziell online. Gemäß der Phaseneinteilung Iriberris und Leroys (2009) ist „Evangelisch in Niedersachsen“ der Creation-Phase eines Community-Lebenszyklus zuzuordnen. Die hauptsächliche technische Ausgestaltung wurde mit der Beendigung der Pilotphase bzw. offiziellen Bekanntgabe des Wikis abgeschlossen. Seitdem ist die Community in der Entstehung begriffen. Anders als eine offizielle kirchliche Website der evangelischen Kirche, ist das Wiki (im Rahmen der technischen Möglichkeiten) für spezielle Anforderungen von Gemeindemitgliedern, d.h. z. B. die Vorstellung von Gemeinden und Pastoren, Veranstaltungsübersichten und Gruppenangeboten ausgelegt. Der offizielle Charakter wurde bewusst abgelegt, um Raum für individuelle Entfaltung jedes interessierten Protestanten in Niedersachsen zu schaffen. Abbildung 1 zeigt das Portal in der Übersicht.

Abbildung 1: Startseite „Evangelisch in Niedersachsen“, URL http://ev-nds.evlka.de/wiki/index.php?title=Portal:Startseite.

Das Kirchen-Wiki besteht aus drei Bereichen, die durch Tabs visualisiert werden: „Evangelisch in Niedersachsen“, „Unser Netz“ und „Evangelische Orte“. Der erste Bereich ist das eigentliche Wiki, das aus einer Sammlung von Wiki-Artikeln besteht. Diese werden über Kategorien Themengebieten zugeordnet, wodurch eine Strukturierung des Wikis entsteht. Das zweite Tab „Unser Netz“ zeigt die Einträge, die von Wiki-Nutzern angelegt werden. Der dritte Reiter „Evangelische Orte“ visualisiert die von Nutzern eingetragenen postalischen Adressen auf einer Google Maps-Karte für Deutschland, vgl. Abb. 2.

Jeder Wiki-Nutzer hat eine eigene Benutzerseite. Die Kommunikation zwischen Wiki-Nutzern kann zum einen über öffentliche Einträge auf der jeweiligen Benutzerseite geschehen oder über E-Mail innerhalb des Systems. In die Wiki-Seiten können Bilder, Textdokumente (PDFs), Flickr[2]-Fotogalerien, sowie Audio- und Video-Dateien (MP3 und Flash) eingebunden werden. Mit diesen gegebenen Möglichkeiten kann jede Gemeinde ihre eigene Wiki-Seite gestalten und z.B. eine kurze Beschreibung zur Gemeinde geben, mit einem Bild der Kirche, einer Angabe der Öffnungszeiten des Pfarramtes und eines Kontaktes zu Pastoren sowie ausstehende Termine bekanntgeben und Berichte über vergangene Veranstaltungen einbringen.

1 http://www.kirche.funcity.de

2 http://www.flickr.com/

Abbildung 2: Tab Evangelische Orte, URL http://ev-nds.evlka.de/wiki/index.php?title=Orte:Portal.

Das Hauptziel, das aus Sicht der Verantwortlichen mit dem Wiki verfolgt wird, ist, eine Infrastruktur für die Internetpräsenz aller zur Landeskirche Hannover gehörenden Gemeinden aufzubauen. Des Weiteren sollen redaktionell eingestellte Veranstaltungshinweise bereitgestellt werden. Darüber hinaus sollen Diskussionen ermöglicht werden, wobei sowohl Raum für freie Diskussionsthemen zur Verfügung steht als auch redaktionelle Anregungen für kontroverse Themen angeboten werden. Als Minimalziel gilt, die Auffindbarkeit jeder niedersächsischen Gemeinde im Internet zu erleichtern. Dazu gehört, jeder Gemeinde Raum zur Präsentation anzubieten und dabei bestehende Gemeinde-Homepages im Wiki zu integrieren. Ein weiteres Ziel ist die Vernetzung aller zur Evangelisch-lutherischen Landeskirche Hannovers gehörenden Gemeinden, um Kooperationen zu erleichtern, gemeinsam an Projekten zu arbeiten und dadurch einen größeren Wirkungsradius abzudecken. Mit dem Kirchen-Wiki soll außerdem zum „Mitmachen“ angeregt werden, indem Menschen ihr eigenes Wissen, ihre Meinungen oder Fragen einbringen können und ein Austausch zwischen den Gemeinden auf diesem Weg auch das nicht virtuelle Gemeindeleben bereichert. Der Brückenschlag zwischen Gemeinde vor Ort und Internetgemeinschaft kann somit vielfach erreicht werden. Auch Kirchenfernen soll über dieses alternative Angebot (wieder) eine Annäherungsmöglichkeit geboten werden.

4 Erfolgsfaktoren partizipativer kirchlicher Webangebote

Erfolgsfaktoren zur Gestaltung partizipativer kirchlicher Webangebote lassen sich einerseits aus der Fachliteratur im Bereich der computervermittelten Kommunikation, der Mensch-Maschine-Interaktion sowie des Community Building und Social Media – bzw. direkt auf die Fallstudie bezogen – Wikis, erschließen. Zugleich sind kirchenspezifische Kontextfaktoren zu berücksichtigen. Darüber hinaus ist anzunehmen, dass eine erfolgreiche Community das Potenzial besitzt, die realweltliche (Face-to-Face) Kirchengemeinschaft zu fördern. Es wird also der Frage nachgegangen, welchen Beitrag Social-Media-Plattformen wie „Evangelisch in Niedersachsen“ für die Gemeinschaft in den Gemeinden und für die Kirche insgesamt leisten.

Eine grundlegende Orientierungshilfe für die benutzergerechte Gestaltung von Internetanwendungen stellen standardisierte Vorgaben zur Mensch-Maschine-Interaktion wie die ISO-Normreihe 9241 (ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten) dar, vgl. (Gimmler 2007: 326) . Des Weiteren nennt Huber (2008) eine Reihe von Motiven zur Nutzung von Web 2.0-Maßnahmen, die zum Erfolg bei der „Kommunikation 2.0“ beitragen können und hat hierzu eine Checkliste erstellt. Rotz (2007) stellt 16 Einflussfaktoren für erfolgreiche Communities auf, die im Wesentlichen mit denen von Huber übereinstimmen, zusätzlich liefert er community-spezifische Faktoren. Die meisten Autoren messen den Erfolg einer Community anhand der Anzahl der Beiträge bzw. des Grades der Beteiligung sowie der Art der Beziehungen unter ihren Mitgliedern, vgl. (Iriberri et al. 2009: 9). Preece (2000) und Preece et al. (2002) stellen eine Reihe von Einflussgrößen auf, die zum Erfolg einer Community beitragen können und ordnen diese den beiden Kategorien Sociability und Usability zu. Sociability wird als „planning and developing social policies“ verstanden, die das Ziel einer Community unterstützen und zum harmonischen Miteinander beitragen, sich also mit sozialer Interaktion beschäftigen. Usability ist hingegen auf die Mensch-Computer-Interaktion ausgerichtet und soll die technischen Aspekte von Kommunikationsprozessen innerhalb einer Community befördern. Ein zentraler Aspekt einer Community stellt ein gemeinsames Ziel aller Beteiligten dar. Dies befördert die notwendige Identifikation, die gemäß Moskaliuk (2008) wichtig für den Erfolg mit einer Community ist, damit sich ein Nutzer als Teil der Gemeinschaft fühlen kann. Moskaliuk (2008) spricht im Gegensatz zu Preece (2000) keine eindeutige Empfehlung aus, ob das Ziel einer Community bereits bei der Konzeption klar definiert sein sollte oder ob es im Laufe ihrer Entwicklung entstehen kann, obwohl vorstellbar ist, dass eine unklare Zielvorstellung die Partizipation an und Identifikation mit einer Community erschweren kann. Trotzdem ist denkbar, dass das Ziel einer Community sich im Verlauf ihres Daseins verschieben kann. Letztlich wird es von seinen Nutzern bestimmt. Im Rahmen des Kirchen-Wikis ist die Fokussierung auf eine bestimmte Interessensgemeinschaft von besonderer Bedeutung.

Um Erfolgsfaktoren aus kirchlicher Sicht näher zu bestimmen, stellt sich zunächst die Frage, was eine kirchliche Web 2.0-Anwendung an kontextspezifischen Inhalten grundsätzlich anbieten sollte, damit sie erfolgreich wird. Doch kann diese Frage überhaupt allgemeingültig beantwortet werden? Vorhandene Literatur, die sich bisher mit dem Internet und seinen Möglichkeiten für die Kirche auseinandergesetzt hat, liefert Anhaltspunkte, worauf zielgruppenspezifisch zu achten ist. Bereits 1998 hatte Bruns (1998) genaue Vorstellungen darüber, was in zukünftigen Internet-Angeboten nicht fehlen sollte, um sie für die Kirche optimal nutzbar zu machen. Technische Möglichkeiten ändern sich zwar laufend, doch der soziale Anspruch an eine Gemeinschaft hat auch nach mehr als einem Jahrzehnt nicht an Aktualität verloren. Dies wird sich vermutlich auch in Zukunft kaum ändern. Bruns Vorschläge für das Internet bleiben somit aktuell. Das Internet bietet weltumspannende Interaktionsmöglichkeiten, doch braucht es „gemeinsame Bezugspunkte“ (ebd.: 37), um die Kommunikation aufrecht zu erhalten. Wie auch Moskaliuk (2008) und Brok und Reinmann (2007) anführen, scheint ein Social-Media-Angebot nur dann attraktiv zu sein, wenn Menschen beteiligt sind, zu denen auch außerhalb des Internets eine Verbindung besteht. Hierzu gehören im kirchlichen Bereich z. B. das gemeinschaftliche Besuchen von Veranstaltungen oder regelmäßigen Programmen des Gemeindeangebots wie Chöre oder Kindergruppen. Die Vor- und Nachbereitung oder Terminübersicht ließe sich entsprechend online umsetzen. Die Kombination aus Online- und Offline-Kommunikation

fördert den Zusammenhalt einer Gemeinschaft und die Pflege von Kontakten. Selbst Internetkontakt und Internetbeteiligung, wie regelmäßiger sozialer Austausch in einer Community, stärken erwiesenermaßen die nicht virtuelle Beziehung.

Eine Aufgabe der Kirche ist es, Fragen zu Lebensorientierung und Weltanschauung zu beantworten. Hierbei könnte sie die neuen Medien nutzen, um Antworten und Denkanstöße zu platzieren und „Orientierung in der Fülle der Informationen" (Neukirch 1996: 25) zu geben. Die Kirche hat die Möglichkeit, ihre Aufgaben auch innerhalb des alternativen Mediums Internet, wahrzunehmen (Brok et al. 2007: 20). Damit kann Raum zum Gedankenaustausch und zur Diskussion geschaffen und grundlegende Fragen des christlichen Glaubens beantwortet werden (Bruns 1998: 37). So kann die Kirche als Ansprechpartner in Bereichen wie Seelsorge und Lebenshilfe dienen und dabei die Anonymität des Internets nutzen (ebd.). Wie glaubwürdig Kirche im Internet ist, lässt sich mit einem Konzept der PR-Ethik[3] für Glaubwürdigkeit für die Kirche darstellen, das sich auf den vorliegenden Kontext übertragen lässt (Uden 2007: 293). In diesem Konzept wird der Aspekt Glaubwürdigkeit u. a. anhand von Überzeugung durch Transparenz und „Generierung sozialen Nutzens" gemessen. Uden (2007) schlägt in diesem Zusammenhang vor, den Aspekt Transparenz beispielsweise anhand der Offenlegung kontroverser kirchlicher Themen umzusetzen. Der soziale Nutzen führt neben der Akzeptanz auch zu höherer Partizipationsbereitschaft (ebd.). Doch auch die Authentizität der Nutzer einer kirchlichen Social-Media-Anwendung ist ein zu beachtendes Kriterium, das Auskunft darüber gibt, auf welcher Kommunikations- und Selbstdarstellungsebene interagiert wird (Bruns 1998: 37). Auf beiden Seiten, sowohl aus Betreiber- als auch Benutzersicht, ist Glaubwürdigkeit wie oben erwähnt ein wichtiges Kriterium, das die Nutzung eines Internetangebots beeinflussen kann. Gräb und Weyel (2007) schlagen weiterhin vor, Möglichkeiten der spirituellen Praxis einzubringen, um kirchliche Rituale wie Online-Andachten oder Online-Exerzitien[4] auch im Internet ausüben zu können, wie dies einige Anbieter bereits umgesetzt haben[5]. Der Anspruch an das Internet besteht zum einen darin, wie in Face-to-Face-Situationen, Gleichzeitigkeit (Vogel 2007: 71) zu erleben, zum anderen gibt es den Vorteil des asynchronen Erlebens von Religiosität. Kirchliche Angebote im Internet sind somit als Chance zu sehen, Menschen zu erreichen und zu (re)integrieren, die über vorhandene Gemeindeaktivitäten nicht erreicht werden (Nethöfel 2000: 15). Doch ist dieser Vorschlag Gräb und Weyels (2007) im Rahmen der jeweiligen Zielvorstellungen über eine neue Internet-Anwendung und der aktuellen Angebotslandschaft zu betrachten. Auch insgesamt gilt bei diesen auf den Kontext Kirche bezogenen Punkten, dass nicht jeder einzelne bei der Erstellung eines partizipativen Onlineangebots zwingend zu erfüllen ist. Vielmehr liefern sie Möglichkeiten und Anregungen, die je nach Zielvorstellung und Zielgruppe bei der Konzeption und Umsetzung eingebracht werden können.

Tabelle 1: Katalog von Erfolgskriterien partizipativer kirchlicher Social Media.

Nr.	Kriterium
1	Zielbestimmtheit: Eindeutigkeit und Transparenz der Zielsetzung
2	Technologieangemessenheit: Zweckorientiertheit in Bezug zu Alternativen
3	Glaubwürdigkeit: Seriosität, Authentizität, Transparenz
4	Betreuung: Technische Hilfestellung, soziale Unterstützung, Optionen für Feedback und Anregungen
5	Bekanntheit: Marketingkanäle und Marketingmaßnahmen
6	Grad der Selbstdarstellung: Anonymität vs. Selbstdarstellung, Datenschutz und Privatsphäre
7	Zielgruppenadäquanz: · Identifikation, Erwartungshaltung (ähnliche Ziele/Erwartungen in der Zielgruppe) · Zielgruppenbedürfnisse (Vernetzung, Internetauftritt, Gemeindearbeit) · Motivation der Partizipation der Zielgruppe (Wikiseiten, Informationen, Termine)
8	Kommunikations- und Interaktionsmöglichkeiten Moderation, Anleitung durch Betreiber (Anreize zur Reflexion und Diskussion, Gedankenaustausch) · Rückbindung (Bezug zur Gemeinde vor Ort) · User-Profil (Identitätsmanagement) · Nachrichtenaustausch (Mitteilungen, Diskussionen)
9	Controlling: Nutzerverhalten/Aktivität · Nutzerstatistik (Besucherzahl, Benutzerkonten, Häufigkeit der Besuche, Seitenaufrufe) · Beteiligungsart, Nutzertypen (Leser, Autor bzw. passiv vs. aktiv) · Verhaltensregeln (Netiquette)
10	Informationsangebot: Aktualität, Archivierung
11	Technische-Ausgestaltung & Design · Anmeldung (Bedingungen technisch/sozial) · Bedienbarkeit (Hilfeseiten, Suchfunktion, Medien-Upload) · Wiki-Aufbau (Vorstrukturierung, Spezialseiten) · Server (Stabilität, Geschwindigkeit)

5 Kriterienkatalog

Nachdem relevante Einflussgrößen für den Erfolg einer Social-Media-Anwendung dargelegt wurden, soll nun versucht werden, daraus einen operationalisierbaren Kriterienkatalog zu Erfolgsfaktoren kirchlicher Social-Media-Angebote zu erschließen. Hierzu werden die dargestellten Aspekte zu Kriterien zusammengefasst und damit aggregiert und strukturiert. Anzumerken ist, dass die nachfolgende Aufstellung keinen Anspruch auf Vollständigkeit erhebt, sondern einen Ansatz darstellt, um das Wissen der Fachgebiete der computervermittelten Kommunikation, der Mensch-Maschine-Interaktion sowie des Community Building und Social Media und die Anforderungen der Anwendungsdomäne zusammenzuführen.

Der Katalog verdeutlicht die Spannweite unterschiedlicher Gestaltungsaspekte und Erfolgsmerkmale sozialer kirchlicher Medien. Die Social-Media-Anwendung sollte professionell durch einen Portalverantwortlichen betreut werden, der

3 PR steht für Public Relations (Öffentlichkeitsarbeit).

4 „Exerzitien ist abgeleitet aus dem Lateinischen und bedeutet […] „geistliche Übungen". [Es sind] Übungen für die Seele, um tiefer in Beziehung zu sich selbst, zum eigenen Leben und zu Gott zu kommen." (Arbeitsgemeinschaft der deutschen Diözesan-Exerzitien-Sekretariate 2009a) „Grundelemente sind [u. a.] schweigend meditierendes persönliches Beten, geistliche Betrachtung von Heiliger Schrift und geistlicher Tradition der Kirche […]" (Arbeitsgemeinschaft der deutschen Diözesan-Exerzitien-Sekretariate 2009)

5 Z.B. http://www.webandacht.de, http://www.kirche.funcity.de/

für die Administration der Teilnehmer und gegebenenfalls die Beseitigung von Spam oder gar Flames[6] verantwortlich ist. Hinzu kommt eine zuverlässige technische Betreuung. Voraussetzung für das technische Funktionieren einer Internetanwendung ist ein stabiler Server. Auch eine Mitgliederbetreuung ist empfehlenswert. Das Feedback der Teilnehmer wird nicht allein in der Konzeptionsphase benötigt, auch in einer späteren Phase einer Community-Plattform ist es förderlich, Raum für Anregungen für seine Nutzer zu schaffen, damit Rückmeldungen an die Betreiber gegeben werden können, die darauf reagieren sollten. Betreiber sollten die Bekanntheit der Social-Media-Anwendung in der Zielgruppe sicherstellen. Des Weiteren sollte sie der Zielgruppe angepasst sein und die Primärbedürfnisse, seien es nun Internetpräsenz, Vernetzung oder Gemeindearbeit fokussieren. Hinsichtlich Kommunikation und Interaktion ist es nicht ausreichend die technische Infrastruktur anzubieten, vielmehr ist es erforderlich, diese aktiv zu initiieren und zu pflegen. Hierzu sind z.B. regelmäßige Beiträge durch die Betreiber, etwa Denkanstöße in Text- oder Videoform eine probate Form der Stimulation. Effektives Controlling ist wiederum die Voraussetzung um in dieser Hinsicht als Betreiber effektiv agieren zu können. Die Social-Media-Anwendung sollte die Aktivität der Benutzer fördern. Da ebenso die Benutzerfreundlichkeit als grundlegender Aspekt eines solchen Angebots, gerade für ältere, vergleichsweise wenig webaffine Zielgruppen, von zentraler Bedeutung ist, sollten neben einer effektiven und effizienten Benutzerführung auch Hilfsangebote bereitgestellt werden.

6 Analyse von „Evangelisch in Niedersachsen“

Nachfolgend wird auf Grundlage des erarbeiten Kriterienkatalogs das Portal „Evangelisch in Niedersachsen“ aus drei verschiedenen Sichtweisen analysiert. Die Betrachtung der Nutzungsdaten des Portals ermöglicht eine grundlegende Einschätzung zur quantitativen Reichweite (z.B. Besucherzahlen) und des Umfangs sowie der Ausprägung der Partizipation. Die Expertenanalyse prüft, inwieweit die aus der Literatur erarbeiteten Kriterien berücksichtigt werden bzw. erfüllt sind. Die Benutzerbefragung gibt diesbezüglich eine auf Selbstauskünften basierte Einschätzung durch die Zielgruppe wider.

Tabelle 2: Nutzungsdaten von „Evangelisch in Niedersachsen“.

Anzahl	Nach Startphase: 24.10.2009	24.07.2010	Wachstum (%)
Wiki-Seiten	543	604	11,2
Hochgeladene Dateien	654	770	17,7
Registrierte Nutzer	242	308	27,2
Seitenbearbeitungen	6520	7670	17,6
Seitenabrufe insgesamt	427256 (in 5,5 Monaten ~ 77682/Monat)	868456 (in 15,5 Monaten ~ 56029/Monat)	103
Vertretene Gemeinden	90	106	17,8

6.1 Nutzungsdaten

Die Analyse der Nutzungsdaten stützt sich primär auf die durch die Wiki-Software des Portals bereitgestellten Statistiken[7]. Hierzu wurden zu zwei Erhebungszeitpunkten Analysen vorgenommen. Das erste Datum ist der 24. September 2009, ein Zeitpunkt, der – einige Monate nach dem Start des Portals auf der Landessynode am 8. Mai 2009 – den Stand nach dem offiziellen Start und der Durchführung erster Marketingmaßnahmen widergibt. Ergänzend werden zum Vergleich die Daten am 24. Juli 2010 angeführt. Diese liefern ein Bild darüber, welche weitere Entwicklung das Portal in den folgenden zehn Monaten genommen hat. Tabelle 2 fasst die Ergebnisse in einer Übersicht zusammen.

Die Zahlen geben interessante Einblicke zur Nutzung des Portals. Vor dem Hintergrund, dass die Evangelisch-lutherische Landeskirche Hannover 1346 Gemeinden umfasst, die sich insgesamt 57 Kirchenkreisen in Niedersachsen zuordnen lassen, zeigt sich, dass insgesamt nur ein kleiner Teil der Zielgruppe, 308 registrierte Nutzer und 106 Gemeinden, im Wiki aktiv bzw. vertreten sind. Die Zahl der Wiki-Seiten zeigt auf der anderen Seite, dass im Wiki Inhalte generiert werden und damit eine, wenn auch im Umfang relativ bescheidene, Wissensbasis aufgebaut wird. Der Vergleich von 2009 und 2010 verdeutlicht, dass der Großteil der Nutzer und Inhalte schon einige Monate nach dem Start des Portals registriert war respektive erarbeitet wurde. Dennoch befindet sich das Wiki diesbezüglich auf einem Wachstumspfad. Die Angabe der Seitenabrufe lässt nur rudimentäre Schlüsse in Bezug auf die rezeptive Nutzung von „Evangelisch in Niedersachsen“ zu. Dennoch ist hier eine Stagnation erkennbar bzw. lässt sich ein Rückgang der Zahl der Abrufe konstatieren. Einerseits ist das Wiki grundsätzlich lebendig, wie die Nutzerzahlen und die Wiki-Seiten verdeutlichen, andererseits noch weit davon entfernt die Mitglieder und Gemeinden der Evangelisch-lutherischen Landeskirche Hannover großflächig zu erreichen.

6.2 Expertenanalyse

Wie oben beschrieben, orientiert sich die Expertenanalyse an dem in Abschnitt 5 erarbeiteten Kriterienkatalog, für eine ausführliche Darstellung vgl. (Pauls 2009: 67-86).

1 Zielbestimmtheit: „Evangelisch in Niedersachsen“ ist ein evangelisches Social-Media-Angebot der Region Niedersachsen, das in einer fünfteiligen sogenannten Tour auf der Website selbst erkundet werden kann. In dem Portal werden Ziele und Partizipationsmöglichkeiten des Wikis explizit vorgestellt[8]. Einerseits bietet es durch seinen inhaltlich offenen Charakter Raum zur freien Gestaltung und kreativen Entfaltung für seine Nutzer. Andererseits birgt die geringe Konkretisierung des Ziels die Gefahr, zu wenig Anreize für eine Teilnahme zu geben. Dennoch kann das Wiki als Ergänzung zu nicht virtuellen Angeboten dazu beitragen, die Gemeinde und die Internetarbeit vor Ort zu fördern.

2 Technologieangemessenheit: Die Wiki-Technologie ermöglicht das kollaborative Erstellen einer Enzyklopädie mit regionalem inhaltlichem Schwerpunkt, primär aus Gründen der flächendeckenden Webpräsenz niedersächsischer Gemeinden. Des Weiteren ist ein Wiki aufgrund langfristiger Dokumentations- und technischer Anpassungsoptionen ein grundsätzlich geeignetes Werkzeug, um die angestrebten Ziele umzusetzen und bildet daher eine adäquate technologische Plattform.

3 Glaubwürdigkeit: Der Aspekt der Glaubwürdigkeit wird durch den recht-

6 Unangebrachte Äußerungen, z. B. Beleidigungen.

7 http://ev-nds.evlka.de/wiki/index.php/Spezial:Statistik [25.07.2010].

8 Vgl. http://ev-nds.evlka.de/wiki/index.php?title=Portal:UeberUns [25.07.2010]

lichen und organisatorischen Status des Trägers des Portals, der Evangelisch-lutherischen Landeskirche Hannovers abgesichert. Der Punkt Transparenz zeigt sich u. a. darin, dass auch kontroverse Themen im Wiki behandelt werden und zu einem späteren Zeitpunkt noch einsehbar sind[9].

4 Betreuung: Das Portal verfügt über fünf Ansprechpartner, die im Problemfall kontaktiert werden können. Neben Hilfestellungen zum Selbstnachlesen existiert eine Wikiseite „Fragen Sie uns“[10] für Fragen und Feedback, die allerdings selten genutzt wird. Insofern lässt sich eine hinreichende Betreuung argumentieren.

5 Bekanntheit: Um die Bekanntheit zu steigern wurde das Wiki verschiedenerseits beworben, z.B. durch Pressemeldungen auf EVLKA.de bzw. mit einer Standpräsenz auf dem Kirchentag 2009. Allerdings zeigt die Analyse der Nutzungsdaten hier erhebliches Verbesserungspotenzial. Insbesondere das Suchmaschinen- und Social Media Marketing scheint ausbaufähig zu sein. Die Website ist in Suchmaschinen wie Google gut indexiert (459 Seiten, Stand 25.07.2010), die Startseite weist hingegen einen Pagerank von 3 (ein eher niedriger Wert) auf.

6 Grad der Selbstdarstellung: Das Portal stellt Benutzerprofile bereit. Der Grad der Selbstdarstellung wird jedem Wiki-Nutzer freigestellt. Er selbst entscheidet über seine Positionierung und die Freigabe persönlicher Daten.

7 Zielgruppenadäquanz: Das Portal adressiert drei mögliche Zielgruppenbedürfnisse: Vernetzung, Internetauftritt und Gemeindearbeit. Der regionale Bezug schafft Möglichkeiten zur Identifikation und offen formulierte Ziele unterstützen eine breitgestreute Erwartungshaltung unter seinen Nutzern. Beispielseiten und leere, lediglich mit einem Titel versehene, Wiki-Seiten geben Anreize zur Teilnahme. Regelmäßige Informationen, die in Abständen von wenigen Tagen redaktionell eingestellt werden, erscheinen in Form von Artikeln und Hinweisen auf weitere regionale und überregionale kirchliche Projekte. Sie sollen das Interesse wecken und zu weiteren Besuchen oder eigenen Beiträgen motivieren. Mit aktuellen technischen Möglichkeiten wie Flickr-Fotogalerien und RSS-Feeds lassen sich die Bedürfnisse der Zielgruppe umsetzen und darüber hinaus ansprechende Seiten kreieren. Das Angebot ist also gut auf die postulierten Bedürfnisse abgestimmt. Der kreativen Beteiligung der Nutzer wird ausreichend Freiraum geboten.

8 Kommunikations- und Interaktionsmöglichkeiten: Für das Kriterium Kommunikations- und Interaktionsmöglichkeiten lässt sich festhalten, dass Anreize zur Teilnahme über redaktionelle Beiträge erreicht werden. Des Weiteren werden Denkanstöße und Einladungen zu Workshops und Tagungen angeboten. Im Wiki angelegte Artikel zur Gemeinde und ihren Terminen ergänzen die Gemeindearbeit, bilden Bezugspunkte zu Face-to-Face-Veranstaltungen und ermöglichen die Rückbindung an die Gemeinschaft vor Ort. Ein Nutzer-Profil, die Möglichkeit, dort Einträge zu hinterlassen, und eine hinterlegbare E-Mail-Adresse unterstützen die Kommunikation unter Wiki-Nutzern. Sie stärken über den virtuellen Kontakt gleichzeitig vorhandene soziale (Face-to-Face) Beziehungen und können ebenso über gemeindeübergreifende Projekte neue Kontakte entstehen lassen.

9 Controlling: Nutzerverhalten/Aktivität: Statistische Daten ermöglichen Portaladministratoren, Nutzer-Aktivität zu verfolgen und i.d.R. wertvolle Ableitungen zu unterschiedlichen Nutzerbedürfnissen, d. h. unterscheidbaren Nutzertypen anzustellen und mit entsprechenden Angeboten zu reagieren. Dabei kann das Verhalten der Benutzer beobachtet und im Falle eines Verstoßes sanktioniert werden. Mit den durch das Wiki selbst bereitgestellten Daten, sowie der installierten Piwik-Software steht für die Betreiber eine geeignete technische Infrastruktur zur Verfügung. Dieses Open Source Statistiktool dokumentiert das Besucherverhalten in Evangelisch in Niedersachsen. Es kann jedoch nicht zwischen angemeldeten und nicht angemeldeten Besuchern unterscheiden. Deshalb wird es zur Analyse des Benutzerverhaltens aller Besucher des Wikis genutzt.

10 Informationsangebot: Das Informationsangebot besteht aus Beiträgen der Wiki-Nutzer und redaktionell eingestellten Artikeln, sowohl von aktuellem als auch von zeitlosem Interesse. Um ein Spektrum an Informationen aus der Region anzubieten, wird von Seiten der Verantwortlichen des Wikis darauf geachtet, in Abständen von wenigen Tagen redaktionell Inhalte einzustellen. So werden derzeit Artikel zu Workshops des „Medientages der Landeskirche eingestellt“[11]. Durch seinen enzyklopädischen Charakter bietet das Wiki Möglichkeiten, Dokumente und Dateien zu archivieren. Bisherige Veranstaltungen werden darin festgehalten, z. B. das „Schülerforum 2008“[12].

11 Technische Ausgestaltung und Design: Die technische Ausgestaltung setzt auf Einfachheit, sowohl beim Zugang als auch in der Bedienung, wobei eine grundlegende Internetaffinität bei Portalnutzern vorausgesetzt wird. Diverse Hilfeseiten bieten Erklärungen zu den elementaren Funktionen. Die Suchfunktion ist mit einer Titel- und Volltextsuche ausgestattet, doch eine Fuzzy-Suche ist nicht integriert, die einem Suchenden ähnlich geschriebene Worte anbieten würde. Der Medien-Upload lässt sich in zwei Schritten ausführen und wird ebenfalls in den Hilfeseiten erläutert. Eine barrierearme Umsetzung ist mit Einschränkung von Flash-Animationen gegeben. Der Wiki-Aufbau bedient sich einer nachvollziehbaren Struktur. Ein stabiler Server erleichtert das reibungslose Bearbeiten von Wiki-Seiten im Portal. Hinsichtlich des Designs überzeugt „Evangelisch in Niedersachsen“ mit seiner innovativen, ansprechenden Optik.

Insgesamt ergibt die Expertenanalyse eine positive Einschätzung. Die einzelnen Kriterien werden von dem Portal meist in hinreichendem Maße berücksichtigt. Der offene Charakter des Wikis schafft einen Raum zur Ausgestaltung kirchlicher regionaler Themen. Unterstützungselemente und Vorstrukturierung sind in ausreichendem Maße vorhanden. Hinsichtlich des Marketings besteht offensichtlich Optimierungspotential. Letztlich entscheidend für den Erfolg sind die Faktoren Zielbestimmtheit bzw. Zielgruppenadäquanz. Die im Zeitablauf gering gestiegene Aktivität weist darauf hin, dass beide Faktoren einen verhältnismäßig geringen Teil der Zielgruppe ansprechen bzw. zu aktivieren vermögen. Dies deutet darauf hin, diese noch stärker auf die Zielgruppe abzustimmen bzw. Bekanntheit und Interesse über Marketing und Betreuung zu verbessern. Die engere sozio-technische Ausgestaltung des Systems vor allem in Bezug zu Technologieangemessenheit, Glaubwürdigkeit, Grad der Selbstdarstellung, Kommunikations- und Interaktionsmöglichkeiten, Controlling, Informationsangebot, technische Ausgestaltung und Design scheinen den Anforderungen an eine erfolgreiche Ausgestaltung zu entsprechen, ließen sich aber auf Basis von Nutzertests weiter optimieren.

6.3 Nutzerbefragung

Letztlich entscheidend für die Relevanz des Portals bzw. der durch die Kriterien

9 Vgl. http://ev-nds.evlka.de/wiki/index.php/Fragen_Sie_uns! [30.07.2010].

10 http://ev-nds.evlka.de/wiki/index.php/Fragen_Sie_uns! [25.07.2010].

11 Vgl. http://ev-nds.evlka.de/wiki/index.php/Medientag [25.07.2010].

12 Vgl. http://ev-nds.evlka.de/wiki/index.php/Schuelerforum [25.07.2010].

gefassten Aspekte sind die Bedürfnisse der Nutzer. Aus diesem Grund wurde mit Hilfe eines Online-Fragebogens eine Besucherbefragung durchgeführt. Der Fragebogen wurde ebenso wie die Expertenanalyse auf Basis der genannten Kriterien entwickelt und auf Grundlage einer Vorabbefragung auf dem Kirchentag 2009 optimiert. Für jedes der in Abschnitt 5 genannten Kriterien wurden Fragen in Alltagssprache formuliert. Nachfolgend wurden die Fragen der Online-Befragung den aufgestellten Kriterien zugeordnet und ihre Aufnahme in die Befragung begründet. Insgesamt wurden 18 Fragen konzipiert. Die Umfrage war vom 29. Juli bis zum 30. August 2009 aktiv und konnte über einen Link auf „Evangelisch in Niedersachsen" aufgerufen werden.

Unter den 76 Teilnehmern der Befragung befinden sich 58 Wiki-Nutzer, die das Wiki mindestens ein Mal vor der Befragung besucht haben und 18 Erstnutzer. Die Erfassung demographischer Angaben dient dazu, ein aktuelles Abbild der Zielgruppe zu erhalten und Abhängigkeiten zwischen diesen Angaben und ihren Aussagen herausarbeiten zu können. Der überwiegende Anteil der Probanden besteht aus männlichen Teilnehmern. Hauptsächlich befinden sich die Befragten im Alter von 31 bis 50 Jahren. Zu 80 Prozent besteht die Gruppe der Befragten aus Ehrenamtlichen und kirchlichen Mitarbeitern. Die meisten Befragten stammen aus den Sprengeln Hannover, Hildesheim-Göttingen und Stade. Sie weisen überwiegend sowohl eine gute Internet- und Social-Media-Kompetenz auf als auch weitreichende Kenntnisse über religiöse Medien, was für ihre Eignung als Teilnehmer dieser Untersuchung spricht. Anhand von vier Eingangsfragen konnte die Eignung der Probanden bestätigt werden. Im Kontrast dazu steht die niedrige Frequentierung des Wikis durch die Mehrheit der Nutzer (monatlich oder seltener), obwohl die meisten Befragten angeben, ein Benutzerkonto zu besitzen.

Es ist bereits deutlich geworden, dass Marketingmaßnahmen zur Förderung des Bekanntheitsgrades des Wikis in Zukunft weitaus häufiger zum Einsatz kommen könnten. Hierbei ist anzumerken, dass diejenigen Personen, die nach Zielvorstellungen der Betreiber zuerst erreicht werden sollten, tatsächlich über die Internetpräsenz EVLKA.de der Landeskirche Hannovers erreicht worden sind und dieser Kanal somit ein erfolgreiches Medium zur ersten Verbreitung eingestuft werden kann.

Tabelle 3 fasst die Bewertungen aller Aussagen zu den Aspekten Vernetzung, Aktivität, Kommunikation und Technische Ausgestaltung in einem Ranking zusammen.

Tabelle 3: Mittelwerte zu den Aspekten Vernetzung, Aktivität, Kommunikation und Technische Ausgestaltung, Vierstufige Skala von 1 „trifft nicht zu" bis 4 „trifft zu".

Rang	Vernetzung, Aktivität, Kommunikation und Technische Ausgestaltung	Mittelwert
1	Mir ist wichtig, an realen Veranstaltungen meiner Gemeinde teilzunehmen.	3,59
2	Es ist hilfreich, dass jede Gemeinde in Niedersachsen das Wiki als kostenlosen Internetauftritt nutzen kann.	3,56
3	Es ist sinnvoll, dass jede Gemeinde sich im Wiki präsentieren kann.	3,31
4	Mir ist wichtig, dass die Vernetzung aller niedersächsischen Gemeinden gefördert wird.	3,22
5	Die Möglichkeit „Evangelische Orte" einzutragen (Google Maps Eintrag) bereichert das Wiki.	3,13
6	Beispielseiten helfen Neulingen sich zu orientieren.	2,93
7	Die „Spezialseiten" (Besondere Wiki-Seiten wie Statistik oder Kategorien) bieten interessante Links.	2,79
8	Das Wiki ist leicht zu bedienen.	2,77
9	Das Wiki ist eine gute Möglichkeit, um die Gemeindearbeit vor Ort zu unterstützen.	2,76
10	Mit der Suchfunktion finde ich sofort die richtigen Einträge.	2,71
11	Ich sehe gern Termine und Veranstaltungen nach.	2,71
12	Die Hilfeseiten bieten nützliche Unterstützungen.	2,69
13	Ich informiere mich gern über andere Gemeinden in Niedersachsen.	2,67
14	Das Wiki ist inhaltlich gut vorstrukturiert.	2,64
15	Ich informiere mich gern über meine Gemeinde.	2,59
16	Es ist leicht, Fotos, Videos und andere Dateien in Wiki-Seiten einzubinden.	2,56
17	Ich kündige gern Termine und Veranstaltungen an.	2,39
18	„Unser Netz" (Artikelsammlung und ihre Verbindungen) ist eine sinnvolle Animation.	2,31
19	Ich lege gern Wiki-Artikel an.	2,00
20	Mir ist wichtig, eine eigene Profilseite zu haben.	1,89
21	Ich diskutiere gern mit anderen Wiki-Nutzern (virtuell).	1,60
22	Ich korrigiere gern Wiki-Artikel.	1,58

Es zeigt sich in Bezug auf Zielgruppenadäquanz, dass Vernetzungsaspekten für die Befragten bei der Bewertung des Kirchen-Wikis eine wesentliche Rolle zukommt. Eine kostenlose Infrastruktur zur Selbstdarstellung und Kooperation mit anderen Gemeinden ist auf „Evangelisch in Niedersachsen" gegeben. So werden diese Voraussetzungen für eine Vernetzung als besonders wichtig empfunden und erhalten gute bis sehr gute Bewertungen.

Hinsichtlich Kommunikations- und Interaktionsmöglichkeiten wird Face-to-Face-Kommunikation gegenüber virtueller Kommunikation deutlich bevorzugt. In dieser Hinsicht scheint die Mehrheit der Teilnehmer dieselbe Einstellung zu teilen und ähnliche Erwartungen an das Wiki zu haben. Sie sind sich darüber einig, was ihnen im Wiki weniger wichtig ist: eine Wiki-Profilseite und virtuell stattfindende Diskussionen.

In Bezug auf das Informationsangebot erfreut sich die Rezeption von Inhalten, d. h. passive, lesende Tätigkeiten im Wiki, bei den Befragten höherer Zustimmung als informationsbereitstellende Aktivitäten wie die Informationsweitergabe und Wissensaufbereitung in Form von Veranstaltungsankündigungen sowie das Schreiben und Korrigieren von Wiki-Artikeln.

Die Meinungen der Teilnehmer bezüglich der technische Ausgestaltung und des Designs des Wikis sind ähnlich, sie bewegen sich im mittleren bis positiven Bereich. Aus Antworten zu offenen Fragen (Ergänzungs- oder Veränderungsvorschläge) geht hervor, dass eine Optimierung der Bedienbarkeit des Wikis sowie der Hilfeseiten und der Übersichtlichkeit des Wikis erwünscht ist.

Des Weiteren werden eine stärkere redaktionelle Betreuung und sinnvolle Konventionen zur Benennung von Kirchengemeinden vorgeschlagen. Auch wird Unklarheit über das Ziel des Wikis angegeben, ein Vorschlag für eine Erweiterung von „Unser Netz“ gemacht und Kritik an der Namensgebung des Kirchen-Wikis geübt. Die Kommentare daraus verdeutlichen, dass ein gewisser Mehrbedarf an Bekanntheit und Zielkommunikation besteht, was wiederum mit den Ergebnissen der Expertenanalyse konform geht.

7 Ergebnis und Empfehlungen

Durch die Untersuchung ist deutlich geworden, welchen Kriterien bereits in hinreichendem Maße entsprochen wird und in welchen Bereichen noch Optimierungsbedarf bzw. Verbesserungspotential besteht. Letzteres fokussiert sich vor allem auf die von (Preece 2000) angeführten Sociability-Bereiche und Marketing: Zielbestimmtheit, Zielgruppenadäquanz, Bekanntheit. Daneben lassen sich, wenn auch weniger zentral, ergänzend auch Verbesserungen in Bezug auf Kommunikations- und Interaktionsmöglichkeiten und Informationsangebot vorschlagen.

Das Ziel des Wikis wird an mehreren Stellen im Portal aufgegriffen und mit Beispielen versehen. Die Philosophie, die hinter Wikis steckt und die Übertragung auf dieses spezielle Kirchen-Wiki könnte jedoch noch prominenter kommuniziert werden. Dieses erschließt sich bislang primär über Texte, die von der Startseite verlinkt sind. Es wäre sinnvoll, das Ziel gut sichtbar auf der Startseite zu kommunizieren. Dabei könnte zudem noch deutlicher auf die zur Verfügung gestellte Infrastruktur für die Webpräsenz jeder niedersächsischen Gemeinde eingegangen werden. Auch auf den Austausch und die Dokumentation von religiösen Inhalten oder aktuell einsehbaren Terminen der eigenen und benachbarten Gemeinden könnte in der Zielbeschreibung hingewiesen werden.

Hinsichtlich des Bekanntheitsgrades ist zu konstatieren, dass eine flächendeckende Verbreitung in ganz Niedersachsen noch nicht stattgefunden hat. So sind weitere Marketingmaßnahmen anzuraten. Neben Online-Marketing-Maßnahmen (vor allem Suchmaschinen und Social Media Marketing) könnten auch Offline-Aktionen z. B. Anzeigen im Gemeindebrief oder Empfehlungen in Gruppenstunden (Frauen-/Männertreff, Kinder- und Jugendgruppen) vermehrt eingesetzt werden, um über den Kreis der Internen hinaus Gemeindemitglieder und Kircheninteressierte anzusprechen. Der geringen Frequentierung könnte z. B. eine Benachrichtigung an Wiki-Nutzer entgegen wirken, die nach einer festgelegten Zeit greift und einen Wiki-Nutzer an die Nutzung seines Accounts erinnert. Diese Nachricht könnte z. B. eine freundliche motivierende Erinnerung enthalten, das Wiki wieder einmal zu besuchen und die derzeit aktuellen Artikel oder vorhandene Themengebiete (ausgewählt nach einem Zufallsprinzip) im Portal als Anreiz zum Besuch mitschicken.

Die Zielgruppenadäquanz ist hinsichtlich der Aspekte Vernetzung, Internetauftritt und Gemeindearbeit grundsätzlich erfüllt. Da Vernetzung ein wichtiges Thema für die Teilnehmer der Befragung darstellt, ist es empfehlenswert, dieses Ziel in Zukunft weiterhin zu verfolgen und umzusetzen. Z. B. kann durch die Betreiber vermehrt auf die Möglichkeit für Kollaborationen hingewiesen werden oder zentrale Projekte der Landeskirche Hannovers im Wiki bearbeitet und dokumentiert werden, um Wissen zu teilen und Erfahrungen auszutauschen.

Für das Erweitern von Kommunikations- und Interaktionsmöglichkeiten empfiehlt es sich, weiter zu beobachten, ob Nutzer sich in Zukunft häufiger beteiligen. Dies könnte weitere Kommunikationsmöglichkeiten erforderlich machen, wie z. B. eine synchrone Messaging-Funktion. Des Weiteren wären Awareness-steigernde Dienste, z.B. eine „Wer ist online?-Funktion“ für das Portal empfehlenswert. So können die Nutzer sehen, welche anderen Nutzer gleichzeitig online sind. Dies kann zur Teilnahme und zur Nutzung der Kommunikationsfunktionen motivieren. In Kombination mit einer Messaging-Funktion kann es die Kommunikation untereinander, z. B. für gemeindeübergreifende Projekte, erleichtern. Um eine höhere Partizipation zu erreichen, könnten z. B. Wettbewerbe angesetzt werden, die Nutzer dazu animieren, aktiv zu werden, etwa durch die Teilnahme an einem Foto- oder Videowettbewerb zu einem bestimmten Thema. Gleichzeitig könnten solche Aktionen auch als Marketinginstrument zur Verbreiterung der Nutzerbasis wirken.

Den wohl größten Anreiz zur Teilnahme an „Evangelisch in Niedersachsen“ bieten derzeit lokalbezogene Informationen. Die eher informationssuchende Ausrichtung der Nutzer deutet darauf hin, das Wiki eher als Nachschlagewerk und Terminkalender einzustufen und weniger als Portal zur virtuellen Community-Bildung. „Evangelisch in Niedersachsen“ ist somit eine besondere Community mit einem Fokus auf Informationsbeschaffung, welcher der Face-to-Face-Kontakt weitaus wichtiger ist als der virtuelle. Obwohl aus Sicht der Betreiber ein Wiki das angemessene Medium für das von ihnen angestrebte Ziel ist, bleibt abzuwarten, ob es sich tatsächlich als das richtige Medium erweist. Denn wer wird die Inhalte einstellen, wenn Nutzer in erster Linie informiert werden wollen? Zwar unterstützen sie die Idee eines lokalen Informations- und Austauschportals, jedoch wird von ihnen derzeit noch eine passive Haltung eingenommen. Es bleibt somit spannend, ob sich mehr Menschen dieser konkreten Zielgruppe dazu animieren lassen oder selbst das Potential erkennen, eigene Inhalte zu erstellen. Zunächst wird das Portal sicherlich als Informations- und Dokumentationsmedium, vorrangig aus redaktioneller Quelle, ergänzend zu bestehenden Angeboten genutzt werden.

Das Portal sollte jedoch nicht isoliert betrachtet werden, sondern ist globaler und distanzierter gesehen ein Angebot, das neben anderen existiert. Sieht man den Zweck des Portals primär in der Ergänzung und Erweiterung vorhandener Gemeindeaktivitäten, wäre aufgrund der Ergebnisse der Untersuchung eine Anpassung des Zielfokus und der Softwareinfrastruktur zu bedenken. Wikis als Social Software lassen sich primär zur Beförderung der kooperativ-kollaborativen Wissensgenerierung nutzen. Für ein Webangebot, dass primär dem Sozialen Publizieren von Inhalten bzw. der Selbstdarstellung (Informationspräsentation der Gemeinden und ihrer Mitglieder) und der Anbahnungs- und Anschlusskommunikation dient, wäre wohl ein Soziales Netzwerk geeigneter als eine Wiki-Software.

8 Einordnung und Ausblick

Das in dieser Arbeit gewählte methodische Vorgehen stellt einen möglichen Ansatz dar, um Erfolgsfaktoren Sozialer Medien in dem speziellen Kontext Kirche aufzustellen und sie auf ihre Erfüllung zu überprüfen. Hierzu wurde ein Katalog relevanter Einflussfaktoren kompiliert, der ein breites Spektrum an möglichen Faktoren abdeckt. Zwar bleibt kritisch zu hinterfragen inwieweit „Erfolg“ durch das additive Zusammentragen und analysieren einzelner Kriterien tatsächlich ursächlich geprüft werden kann. Dennoch bleibt festzuhalten, dass ein derartiges Vorgehen das Einnehmen einer holistischen Perspektive befördert oder anders ausgedrückt der Gefahr einer methodischen Verengung – z.B. auf isolierte Sichten und begrenzte Methoden z.B. der Mensch-Maschine-Interaktion oder des Community Building – vorbeugt. Idealerweise wird mit dem gewählten Vorgehen das relevante Wissen aus den zentralen Fachgebieten der computervermittelten Kommunikation, der Mensch-Maschine-Interaktion sowie des Community Buil-

ding und Social Media mit dem Hintergrund der jeweilige Anwendungsdomäne – hier der Kirche – kontextualisiert. Damit wird der inhaltliche Rahmen zur Analyse kirchlicher Social Media bereitgestellt. Dieser kann, jenseits des hier behandelten Fallbeispiels „Evangelisch in Niedersachsen" grundsätzlich auf eine Vielzahl weiterer kirchlicher Social Media Anwendungskontexte bezogen werden. Dabei sind die erarbeiteten Erfolgskriterien nicht absolut zu setzen. Vielmehr sollte sich aus methodischer Sicht idealerweise ein iterativer Prozess aus Analysen und Umsetzung der ermittelten Ergebnisse entfalten. Dies würde es schließlich gestatten, die aus der Literatur abgeleiteten Kriterien selbst zu überprüfen. Im Rahmen der begrenzten Ressourcen, der diesem Text zugrundliegenden Magisterarbeit von Pauls (2009) konnte dies zeitlich nicht geleistet werden, so dass sich die Untersuchung letztlich auf eine erste Iteration beschränkt. Eine kritische Analyse der hier verwendeten Erfolgsfaktoren sollte Gegenstand weiterer Untersuchungen sein.

Literatur

Arbeitsgemeinschaft der deutschen Diözesan-Exerzitien-Sekretariate (2009): Exerzitien. http://www.exerzitien.info/rund_um_exerzitien.html [26.07.2010].

Brok, T. O.; Reinmann, R. P. (2007): Gottesdienst und Gemeinde im Internet? In: Brok, Tom O.; Friedrichs, Lutz (Hg): Liturgie per Mausklick. Hannover: Gemeinsame Arbeitsstelle für Gottesdienstliche Fragen der Evangelischen Kirche in Deutschland (GAGF),18-2.

Bruns, C.h (1998): Wie werden Medien „menschenverträglich"? In: Nethöfel, W.; Schnell, M. (Hg): Cyberchurch? Kirche im Internet. Frankfurt: Gemeinschaftswerk der Evang. Publizistik Abt. Verl., 34-39.

Gimmler, R. (2007): Qualität von Medienangeboten: Analyse und Bewertung. In: Six, U.; Gleich, U.; Gimmler, R. (Hg) (2007): Kommunikationspsychologie – Medienpsychologie. Weinheim: Beltz, 317-334.

Gräb, W.; Weyel, B. (2007): Handbuch Praktische Theologie. Gütersloh.

Haese, B.-M. (2007): Im Westen nichts Neues! In: Hermelink, J.; Schmidt-Biesalski, A.a (Hg): Praktische Theologie. Zeitschrift für Praxis in Kirche, Gesellschaft und Kultur. Vierteljahresschrift. Vol 42. No. 4, 282-288.

Huber, M. (2008). Kommunikation im Web 2.0. Konstanz: UVK.

Iriberri, A.; Leroy, G. (2009): A life-cycle perspective on online community success. ACM Computing Surveys Vol. 41 No. 2 Art. 11.

EKD Hannover (2009): Evangelische Kirche in Deutschland – Zahlen und Fakten zum kirchlichen Leben 2009. http://www.ekd.de/download/broschuere_2009_internet.pdf [23.07.2010].

Malik, J.; Rüpke, J.; Wobbe, T. (2007): Religion und Medien. Vom Kultbild zum Internetritual. Münster: Aschendorff.

Moskaliuk, J. (2008): Konstruktion und Kommunikation von Wissen mit Wikis. Theorie und Praxis. Boizenburg: Hülsbusch.

Nethöfel, W. (2000): Menschenfischer im Netz. Sind Internetgemeinden die Zukunft der Kirche? In: Evangelische Kommentare, Vol. 33. No. 2, 15-17.

Neukirch, J. (1996): Religion im Internet. Elektronische Kommunikationsformen und die Aufgabe der Kirchen. In: Lutherische Monatshefte 2/96, 22-26.

Pauls, N. (2009): Kirche im Web 2.0 –Erfolgsfaktoren kirchlicher Social-Media-Angebote am Beispiel des Wikis „Evangelisch in Niedersachsen". Universität Hildesheim.

Preece, Jenny (2000): Online Communities: Designing usability, Supporting sociability. Chichester: Wiley.

Preece, Jenny; Maloney-Krichmar Diane (2002): Online Communities: Focusing on sociability and usability. In: The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications book contents. Hillsdale: Erlbaum, 596-620.

Rotz, B. (2007): Online-Communities – Best Practices II. http://www.contentmanager.de/magazin/artikel_1779_online_community_practice.html [23.07.2010].

Shirky, C. (2008): here comes everybody: The Power of Organizing Without Organizations, London: Penguin Books.

Szugat, M.; Gewehr, J. E.; Lochmann, C. (2007): Social software. [Blogs, Wikis & Co.]. Frankfurt: Entwickler. press.

Trocholepczy, B.; Pelzer, J.; Heeg, D. (2009): Studie der Akademie Bruderhilfe-Pax-Familienfürsorge. Kirchliche Sinnangebote im Web 2.0. http://www.kirche-im-web20.de/?m=200903 [22.07.2010].

Uden, R. (2007): PR-Kampagnen der Kirche – Verkündigung zwischen Selbstdarstellung und Gottesrede. In: Praktische Theologie. Zeitschrift für Praxis in Kirche, Gesellschaft und Kultur. Vol. 42. No. 4, 288-294.

Vogel, W. (2007): www.internetgottesdienste.at. Ein Rückblick. In: Brok, T. O.; Friedrichs, L. (Hg): Liturgie per Mausklick. Hannover: GAGF, 71-75.

Kirche, Niedersachsen, Elektronischer Dienst, Fallstudie, Methode, Akzeptanz, Benutzerpartizipation, Bewertung

Die Autoren

Natascha Pauls, M.A.

Studium des Internationalen Informationsmanagements an der Universität Hildesheim, Nebenfächer Betriebswirtschaftslehre und Psychologie und Auslandsstudium an der Universidad de Valladolid, Spanien an der Facultad Filosofía y Letras.

natascha.pauls@uni-hildesheim.de

Prof. Dr. Joachim Griesbaum

Seit 2001 wissenschaftlich im Bereich der Informationswissenschaft tätig. 2008 Ruf auf die Juniorprofessur für Informationswissenschaft mit dem Schwerpunkt Social Networks and Collaborative Media an der Universität Hildesheim. Forschungsschwerpunkte: Web Information Retrieval, Suchmaschinen- und Social-Media-Marketing, E-Learning und kollaboratives Wissensmanagement.
joachim.griesbaum@uni-hildesheim.de

apl. Prof. Dr. Thomas Mandl

war nach dem Studium an der Universität Regensburg und der University of Illinois at Urbana/Champaign wissenschaftlicher Mitarbeiter am Informationszentrum Sozialwissenschaften in Bonn und ab 1998 wissenschaftlicher Mitarbeiter an der Universität Hildesheim. Im Jahr 2000 wurde er an der Universität Hildesheim promoviert und habilitierte sich 2006 zu Qualitätsaspekten im Web Retrieval. Seine Forschungsschwerpunkte sind Information Retrieval und Mensch-Maschine-Interaktion. Er ist Sprecher der Fachgruppe Information Retrieval in der Gesellschaft für Informatik (GI).
thomas.mandl@uni-hildesheim.de

Universität Hildesheim
Institut für Informationswissenschaft und Sprachtechnologie
Lübecker Straße 3
31141 Hildesheim

DGI-Praxistage 2011

Trends, Updates, Change, Management

Information: gift or poison? Die Kompetenz entscheidet!

Quelle: Stadtmarketing Karlsruhe

Ansprechpartner und Veranstalter:

Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V. (DGI)
Windmühlstraße 3
60329 Frankfurt am Main
Tel.: +49 (0)69 430313
Fax: +40 (0)69 4909096
E-Mail: mail@dgi-info.de

Veranstaltungsort:
Best Western
Queens Hotel Karlsruhe
Ettlinger Straße 23
76137 Karlsruhe
Tel.: +49 (0)721 3727-0
Fax: +49 (0)721 3727-170

Informationen zur Veranstaltung und Anmelde-unterlagen finden Sie unter www.dgi-praxistage.de

Für Mitglieder der DGI und Partnerverbänden gelten attraktive Mitgliedertarife.

Sondertarife auch für Studenten und Studentengruppen.

Informationsgesellschaft und Wissenswirtschaft entwickeln sich mit rasanter Geschwindigkeit. In unserer globalisierten Arbeitsumgebung wird der kompetente Umgang mit Information, Wissen und Know-how zu einem wichtigen Wettbewerbsfaktor.

Informations- und Wissensmanagement sind Schlüsselfunktionen für Unternehmen und spielen für strategische Planung, Geschäftsentwicklung, Compliance und Risikomanagement eine wichtige Rolle. Unter dem Motto „Information: gift or poison? Die Kompetenz entscheidet!“ werden Trends und aktuelle Entwicklungen im Informationsmanagement ins Visier genommen

Donnerstag, 7. April 2011

10:30-12:00 Eröffnung
Information: gift or poison? Die Kompetenz entscheidet!
Wie findet die theoretische Diskussion um Informationskompetenz Eingang in die Arbeitswelt? Wenn sie keinen Eingang findet – woran liegt das?
PD Dr. Matthias Ballod [A&O Akademie, Pirmasens]
Michael Fanning [Online Consultants International, Karlsruhe]

13:30-15:00 Session 2
Berufsbild „Information Professional“ - Anspruch und Wirklichkeit in der betrieblichen Praxis
Prof. Dr. Bernd Jörs [Hochschule Darmstadt]
Dr. Sabine Graumann [TNS Infratest Business Intelligence, München]

15:30-17:00 Session 3
Tue Gutes und lass Dir dabei zuschauen - das Management als Kunde
Dr. Ulrich Wieneke [Roche Diagnostics GmbH, Penzberg]
Dr. Jörg Hamann [Köln]

17:30-19:00 Session 4
Ist Web 2.0 der „Blogbuster“ der Zukunft? - Chancen und Risiken beim Einsatz im Unternehmen
RA Dr. Carsten Ulbricht M.C.L. [DIEM & PARTNER, Stuttgart]
Dr. Antje Stobbe [Deutsche Bank Research, Frankfurt am Main]

ab 19:30 Get together
Freuen Sie sich auf einen Abend voller interessanter Gespräche in gemütlicher Atmosphäre.

Freitag, 8. April 2011

09:00-10:30 Session 5
Best Practices - Vernetzung von Kompetenz in Unternehmen
Dipl. Kfm. Georg von Nessler [IP-ShareMedia, Frankfurt am Main]
Dr. Anne Petry-Eberle [Daimler AG, Stuttgart]

11:30-13:00 Closing/Podiumsdiskussion
Information: gift or poison? Wessen Kompetenz entscheidet?

Im Anschluss lädt die DGI alle Mitglieder zur Mitgliederversammlung 2011 ein!
ab 14:30 DGI-Mitgliederversammlung

Wir bedanken uns für die Unterstützung bei:

makrolog CONTENT MANAGEMENT AG
swets
Information Wissenschaft & Praxis
fachbuch journal
B.I.T. online

Benutzerforschung anhand von Log-Dateien: Chancen, Grenzen und aktuelle Trends

Thomas Mandl , Julia Maria Schulz, Nadine Mahrholz, Katrin Werner, Hildesheim

Die Analyse des Verhaltens von Benutzern von Informationssystemen stellt einen Kern der Informationswissenschaft dar. Die Sammlung von umfangreichen Verhaltensdaten fällt mit den heutigen technischen Möglichkeiten leicht. Der Artikel fasst Möglichkeiten und Chancen der Analyse von Log-Dateien zusammen. Der Track LogCLEF wird vorgestellt, der Forschern erstmals die Möglichkeit eröffnet, mit den denselben Log-Dateien und somit vergleichend arbeiten zu können. Die Datengrundlage und einige Ergebnisse von LogCLEF werden vorgestellt.

User research on the basis of log data: opportunities, limitations and current trends

The analysis of user behavior is an important research field in information science. The collection and the tracking of user behaviour are facilitated by modern technology. This article summarizes the potential of the analysis of log files. The track LogCLEF is presented, which provides researchers with an opportunity, to work with the same log data and to compare their results. The data for LogCLEF and some results are shown.

1 Die Ziele des LogCLEF Tracks

Die Forschung mit Log-Dateien von Informationssystemen liefert interessante Einsichten in das Verhalten von Benutzern und die Qualität von Informationsdiensten und stellt eine wichtige Ergänzung zu Benutzertests dar. Während bei interaktiven Labor-Tests immer nur das Verhalten eines kleinen Teils von Testnutzern ausgewertet wird, erlauben Log-Dateien Längsschnittanalysen des Verhaltens aller Benutzer. Auf der anderen Seite ist die Analyse auf die Interpretation der durch Logfiles erfassten Einträge beschränkt, so dass das Nutzerverhalten oft nur sehr ungenau erfasst werden kann. Während beim Benutzertest die Video-Beobachtung jedes Wimpernzucken und die Software jede Interaktion kleinteilig abspeichert, erfolgt der Blick auf den Benutzer bei der Logfile-Auswertung gewissermaßen durch ein Fernrohr. Zwar erlaubt moderne Web-Technologie inzwischen sogar das Verfolgen der Mausbewegungen von realen Benutzern in Log-Dateien (Schulte-Mecklenbeck et al. 2009), gleichwohl bleibt die Auswertung von Log-Dateien ein eher vages Verfahren.

Meist gelten Logs als Betriebsgeheimnis der Betreiber von Diensten. Nicht zuletzt deshalb blieb derartige Forschung bisher meist eine Domäne großer Firmen. Im Rahmen der Evaluierungsinitiative CLEF (Cross Language Evaluation Forum) gelang es erstmals, Forschern Zugang zu umfangreichen Log-Dateien zu verschaffen. Somit können Algorithmen vergleichend oder mit unterschiedlichen Zielrichtungen auf die gleichen Daten angewandt werden.

Der Track LogCLEF zielt auf das bessere Verständnis von Benutzerverhalten und Benutzerabsichten ab. LogCLEF fördert die Analyse und Klassifikation von Anfragen, um so das Suchverhalten insbesondere in mehrsprachigen Umgebungen genauer zu untersuchen. Langfristig tragen die Ergebnisse zur Anpassung der Systeme an die Benutzerabsichten und damit zu ihrer Verbesserung bei.

Im Folgenden werden zunächst die Grundlagen und Anwendungen von Log-Dateien erläutert. Exemplarisch zeigt der vierte Abschnitt aktuelle Forschungsthemen. Im Anschluss steht LogCLEF im Zentrum. Die Datengrundlagen aus digitalen Bibliotheken liefern zahllose Möglichkeiten, das Verhalten der Benutzer zu beobachten. Beispielhaft werden einige Ergebnisse dargestellt. Abschließend wird ein System vorgestellt, das die Exploration von Benutzer-Pfaden unterstützt

2 Logfiles und ihre Struktur

Log-Dateien werden meist in Form sogenannter Webserver-Logs gesammelt. Jeder Klick eines Benutzers im Internet hinterlässt Spuren in Form eines Eintrags in einem Server des genutzten Web-Angebots, wenn dieser entsprechend konfiguriert ist, was meist der Fall ist. (Derartige) Webserver-Logs verfügen über eine einfache Struktur, welche in Tabelle 1 gezeigt wird.

Aus den gesammelten Daten gilt es nun nachträglich die einzelnen Benutzer, ihre Pfade und Interaktionen zu extrahieren (Jansen et al. 2009). Zunächst werden alle Aktionen, die vermutlich auf einen Benutzer zurückzuführen sind, in Sessions zusammengefasst. Allein an der meist vorgenommenen Gleichung „IP-Adresse ist gleich Benutzer“ erkennt man deutlich, dass die Log-Datei-Analyse von Anfang an von Unsicherheiten und Interpretationen durchsetzt ist. Dies liegt zum einen daran, dass Internetprovider IP-Adressen oft dynamisch vergeben. Zum anderen kann es zu IP-Überschneidungen kommen, wenn mehrere Benutzer sich einen Rechner teilen oder einen gemeinsamen Proxy als Zugriffsknoten zum Internet nutzen. Im Rahmen der Identifikation von Benutzer-Sessions sollten daher immer mehrere Parameter herangezogen werden. Zu nennen sind in diesem Zusammenhang insbesondere die IP-Adresse, die Browserversion und die Dauer der Interaktion.

3 Einsatzmöglichkeiten von Logfile-Analysen

Die Auswertung der Log-Dateien kann mehreren Zielen dienen:

- Erkennen des Benutzerverhaltens
- Verbesserung des Systems hinsichtlich der Gebrauchstauglichkeit
- Verbesserung des Systems hinsichtlich wirtschaftlicher Aspekte

Tabelle 1: Die Struktur von Webserver-Logfiles.

Beispiel für Eintrag in Log-Datei	Bezeichnung	Bedeutung
123.123.123.123	remotehost	IP-Adresse des Computers
[01/Jan/2000:15:02:10 -0800]	date	Datum, Zeit und Zeitverschiebung zur Greenwich Mean Time
"GET /info/index.html HTTP/1.0"	request	Transaktionsart - „GET" oder „POST" mit dem angefragten Dateinamen
200	status	Code der angibt ob die Transaktion erfolgreich war
6822	Bytes	Größe der transferierten Datei in Bytes
"http://www.ihrefirma.de/index.html"	referrer	URL von der aus per Link oder Suchmaschine auf die Webseite zugegriffen wurde
"Mozilla/4.01 (WinNT)"	useragent	Name und Version des Browsers und das Betriebssystem

Besonders der dritte Aspekt wird heute meist als Web-Analytics bezeichnet (Hassler 2009). Dabei kommen unter anderem die folgenden Grundtechniken zum Einsatz (Erinaki & Vazirgannis 2003):

- Kollaboratives Filtern: Das Verhalten eines Nutzers wird auf einen ähnlichen Nutzer übertragen (z.B. könnten Nutzer, die sich für ein gleiches Produkt interessieren auch noch andere gemeinsame Vorlieben haben).
- Assoziationsanalyse: Welche Daten werden oft im Zusammenhang aufgerufen? (Assoziationsregeln: Interessiert sich der Nutzer für das Produkt X, wird er sich mit einer bestimmten Wahrscheinlichkeit auch für Produkt Y interessieren.)
- Sequenzanalyse: Erweiterung der Assoziationsanalyse, berücksichtigt auch die Reihenfolge in der die Daten/Internetseiten aufgerufen wurden
- Segmentierung: Ähnliche Nutzer werden zu einer möglichst abgrenzbaren und homogenen Gruppe zusammengefasst (Clusteranalyse)
- Klassifikation: Die Nutzer werden in vorab bestimmte Klassen/Nutzerprofile eingeordnet (Entscheidungsbaumverfahren)

Aus methodischer Perspektive sollte die Log-Analyse stets mit anderen Untersuchungsverfahren kombiniert werden. Die Analyse einer Site für Ferienreisen (Bartel 2003) und die Analyse einer populären Site für Mobiltelefone (Jürgens et al. 2010) zeigen beispielhaft, wie verschiedene Methodenkombinationen umgesetzt werden können.

Die Benutzerforschung mit Log-Dateien steht oftmals in einem Konflikt mit Grundsätzen des Datenschutzes. Der AOL-Skandal warf ein Schlaglicht auf die Problematik: AOL stellte eine umfangreiche Log-Datei seiner Benutzer öffentlich zur Verfügung. Die Daten wurden anonymisiert, indem alle Benutzer-IP-Adressen durch einen Code ersetzt wurden. Somit konnten die Benutzer nicht mehr über die eindeutige IP-Adresse erkannt werden. Jedoch reichten teilweise die Anfragen der anonymisierten Nutzer aus, um die jeweiligen Personen zu identifizieren, was einem Reporter gelang, der damit den Skandal auslöste (Anderson 2006). Die Identifizierbarkeit einer einzelnen Person liegt an sogenannten Ego-Suchen, in denen Nutzer ihren eigenen Namen, ihre Sozialversicherungsnummer oder Ähnliches als Suchanfrage eingeben. Da Menschen auch ihre oder andere Adressen, ihre Kreditkartennummer oder Namen von Krankheiten in eine Suchmaschine eingeben, gilt die Veröffentlichung von Log-Dateien als ethisch und datenschutzrechtlich hochproblematisch. Der Skandal trug dazu bei, dass Suchmaschinen heute sehr zurückhaltend mit der Veröffentlichung solcher Daten sind und Logfile-Forschung in der öffentlichen Forschungs-Community nur beschränkt stattfindet.

4 Forschung zu Logfiles

Die Forschung zur Analyse von Logfiles verfolgt zum einen die Verbesserung der Werkzeuge für die Erfassung und Auswertung und zum anderen deren innovative Auswertung. Darüber hinaus gibt es dezidierte Forschung zur Anonymisierung, um Probleme wie beim oben geschilderten AOL-Skandal in den Griff zu bekommen und dennoch sinnvolle und aussagekräftige Daten bereitstellen zu können (Korolova et al. 2009).

Die Erfassung beschränkt sich heute nicht mehr auf einfache Server-Logs, vielmehr kommen innovative Elemente im HTML-Code dazu, welche die Erfassung der Nutzeraktionen sehr viel genauer aufzeichnen. Ein Beispiel dafür ist das System UsaProxy (Atterer et al. 2006). Es basiert auf einem Proxy, der zwischen Server und Client geschaltet ist und automatisch einen kurzen JavaScript-Code in die Webseiten einfügt, bevor er sie an den Client weiterleitet. Die Autoren sehen den Hauptverwendungszweck in Usability-Tests der Website, können sich aber auch einen Einsatz bei der Website-Entwicklung, während Beta Tests oder kontinuierlicher Evaluation der Site vorstellen.

Zahlreiche Erkenntnisse etwa zur Länge von Anfragen und anderen formalen Merkmalen wurden bisher publiziert (Jansen et al. 2009), wobei einige unten anhand der Analyse der MSN-Log-Datei aus GeoCLEF dargestellt werden.

Aus inhaltlicher Perspektive steht oftmals die Analyse der (Um)Formulierung von Anfragen im Fokus. Eine neuere Arbeit auf Basis der bereits erwähnten AOL-Logs schätzt die Anzahl von Anfragenmodifikationen auf 10% der Menge aller Suchanfragen (Huang & Efthimiadis 2009). Wie in anderen Studien zeigte sich auch hier, dass zahlreiche Umformulierungen aus Unkenntnis über die Funktionalität der Suchmaschine vorgenommen werden. So werden Stoppwörter, z.B. Artikel, eingefügt oder getilgt, welche sich nach Lehrbuchwissen nicht auf das Resultat auswirken. Natürlich bleibt für die Suchmaschinenbetreiber die Frage, ob es den Benutzer zufrieden stellt, wenn er nach einer Modifikation das gleiche Ergebnis erhält. Denn es ist davon auszugehen, dass Nutzer, die beispielsweise nach „Gold" suchen und im weiteren Verlauf ihre Anfrage in „das Gold" modifizieren, in den meisten Fällen mit dem ersten Ergebnis unzufrieden waren. In der aktuellen Untersuchung konnten die Autoren auch zeigen, dass je nach Typ der Umformulierung das Nutzerverhalten abweicht. Dies betrifft die Zeit zwischen den Suchanfragen sowie das Klick-Verhalten. Des Weiteren zeigte sich, dass nach den Anfragereformulierungen durchschnittlich auf höhere Ränge in den Ergebnislisten geklickt wurde, so dass die Benutzer offensichtlich häufig Erfolg hatten (Huang & Efthimiadis 2009).

Einen gewissen Prozentsatz von Anfragen stellen konkrete Wissensfragen dar, die an sich durch Frage-Antwort-Systeme am besten beantwortet werden sollten. Die Unterscheidung zwischen Informationsanfragen mit dem Ziel, Dokumente zu lesen, und Wissensfragen mit dem Ziel, lediglich einen kleinen Ausschnitt eines Dokuments mit der Antwort zu erhalten, stellt eine weitere Herausforderung für die aktuelle Suchmaschinentechnologie

dar. Hilfen kann hier wieder die Log-Datei-Analyse bieten, indem sie derartige Fragestellungen extrahiert (Sutcliffe et al. 2010).

5 Logfile-Ressourcen in CLEF

Log-Dateien stellte erstmals der GeoCLEF Track im Query Parsing Task zur Verfügung. Ziel des Track war es, optimale Algorithmen für die Identifikation von Anfragen mit geographischen Bezügen zu evaluieren. In den Jahren 2010 und 2011 gelang es dem neu initiierten Track LogCLEF, große Mengen von Log-Dateien von digitalen Bibliotheken zu gewinnen und der Community anzubieten. Zudem wurde eine Suchmaschinen-Log-Datei bereitgestellt.

GeoCLEF berücksichtigte als erster Track einer großen Evaluierungsinitiative räumliche Einschränkungen, insbesondere bei Anfragen in cross-lingualen Szenarien (z.B. *Presse-Artikel über Unruhen in der Umgebung von Belfast*, *Haiangriffe an der Küste Australiens*). GeoCLEF wurde nach einem Testlauf im Jahr 2005 von 2006 bis 2008 als vollwertiger Track im Rahmen von CLEF angeboten (Mandl et al. 2009). *Trás-os-MontesVila RealBragançaDonau-DanubioDanubeDunaWaldheim*Im Rahmen von GeoCLEF 2007 wurde ein Query Parsing Task organisiert, in dem es um die Identifikation von geographischen Bedingungen in einer Menge von 800 000 Anfragen, welche im August 2006 von MSN gesammelt wurden, ging. Die Log-Datei enthält keine Informationen außer den Anfragen selbst. Somit sind keine Schlüsse auf das Benutzerverhalten im weiteren Sinn möglich, typische Informationsbedürfnisse können aber erkannt werden. Aus der Hauptmenge wurden 500 Anfragen für Training und Test der Systeme manuell ausgewertet (Li et al. 2007). Die Ergebnisse zeigen, dass die gestellte Aufgabe keineswegs trivial ist.
In dem 2009 neu etablierten Track LogCLEF gelang es, ein weiteres Suchmaschinen-Logfile anzubieten, das von der portugiesischen Suchmaschine Tumba![1] bereitgestellt wurde. Die Anonymisierung erfolgte manuell. Die Aufgabe bestand, in Anlehnung an die Query Parsing Task von 2007, wiederum in der Erkennung von geographischen Elementen (Mandl et al. 2010).

*Chaves*Für die Jahre 2009 und 2010 konnten zudem umfangreiche Daten aus digitalen Bibliotheken für die Forschung zugänglich gemacht werden, namentlich Logs von The European Library (TEL)[2] und des Deutschen Bildungsservers (DBS)[3]. Die geographische Identifikation wurde 2009 auch für TEL angeboten. Der Kern von LogCLEF bestand jedoch in beiden Jahren in einer explorativen Auswertung der Log-Dateien der digitalen Bibliotheken, ein Task welcher bewusst sehr offen formuliert wurde. Für LogCLEF 2009 galt folgende Aufgabenbeschreibung:
„The Log Analysis for Digital Society (LADS) task deals with logs from The European Library (TEL) and intends to analyze user behavior with a focus on multilingual search. Potential targets are query reformulation, multilingual search behavior and community identification. This task is open to different approaches, in particular data mining approaches in order to extract knowledge from the data and find interesting user patterns."

TEL bietet einen integrierten Zugang zu den Beständen zahlreicher Nationalbibliotheken Europas und ist damit ein Nachweissystem. Zu erwarten sind hier Anfragen nach Büchern in vielen Sprachen. Tatsächlich stammen die Nutzer aus vielen Ländern und allein die Identifikation der Sprache der Anfrage stellt sich als schwierig heraus, was typisch für derartige, kurze Texte ist (Gottron & Lipka 2010). Die TEL-Daten bieten mehrere Anhaltspunkte für die Sprache: die eingestellte Sprache bei der Benutzungsoberfläche, das mit der IP-Adresse assoziierte Land, die primär durchsuchte Nationalbibliothek und den Text der Anfrage. Die Log-Dateien wurden von einem spezifisch entwickelten Logdienst aufgezeichnet und zeigen neben den Suchaktionen auch weitere Tätigkeiten des Benutzers, wie etwa das Navigieren in der Ergebnisliste.

Der Deutsche Bildungsserver des Deutschen Instituts für Internationale Pädagogische Forschung[4] bietet Zugang zu qualitätskontrollierten Bildungsinformationen. Der DBS stellt ein Clearinghouse für Ressourcen der Erziehungswissenschaft dar, das dessen Benutzer stark frequentieren. Den Kern bilden kommentierte Internet-Links, welche in einer Verzeich-

Tabelle 2: In CLEF verfügbare Logfiles Ressourcen.

Jahr	Ursprung	Größe	Typ
2007	MSN	800 000 Anfragen	Anfragen-Sammlung
2009	Tumba!	350 000 Anfragen	Anfragen-Log
2009	TEL	1 870 000 Einträge	Anfragen- und Aktivitäts- Log
2010	TEL	2 600 000 Einträge	Anfragen- und Aktivitäts-Log
2010	DIPF.de	5 GB	Anfragen- und Aktivitäts-Log

```
[Dec/2003] GET /pesquisa?terms=helena+ranaldi HTTP/1.1
"http://www.tumba.pt/pesquisa?docs=40&terms=besouro;0
[Dec/2003] GET /pesquisa?terms=vorten HTTP/1.1;0
[Dec/2003] GET /pesquisa?terms=sp+filmes HTTP/1.1 "http://www.tumba.pt/";0
[Dec/2003] GET /pesquisa?terms=renault HTTP/1.1 "http://www.tumba.pt/";0
[Dec/2003] GET /pesquisa?terms=isep&0 HTTP/1.1;0
[Dec/2003] GET /pesquisa?terms=final+fantasy+7-+knigths+of+the+round HTTP/1.1
"http://www.tumba.pt/pesquisa?&terms=final+fantasy+7";0
[Dec/2003] GET /pesquisa?terms=fabulas+de+la+fontaine HTTP/1.1
"http://www.tumba.pt/pesquisa?&terms=fabula+de+la+fontaine";0
[Jan/2004] GET /pesquisa?terms=Society+of+Automotive+Engineers HTTP/1.1;0
[Feb/2004] GET /pesquisa?terms=praias+naturistas+fotos HTTP/1.1
"http://www.tumba.pt/pesquisa?&terms=praias+naturistas";0
[Feb/2004] GET /pesquisa?terms=pesquisa+emprega HTTP/1.1 "http://www.tumba.pt/";0
[Feb/2004] GET /pesquisa?terms=integralista HTTP/1.1 "http://www.tumba.pt/pesquisa?&terms=teologicamente";0
[Feb/2004] GET /pesquisa?site=ZZZ&terms=horarios+biologia HTTP/1.1 "ZZZ/";0
[Feb/2004] GET /pesquisa?terms=consulta+de+notas+na+net HTTP/1.1
"http://www.tumba.pt/pesquisa?docs=10&terms=site:ZZZ+notas&";0

[Feb/2004] GET /pesquisa?terms=servidores+nacionais HTTP/1.1;0
[Jan/2004] GET /pesquisa?terms=neurociencia HTTP/1.1 "http://www.tumba.pt/";0
```

Abbildung 1: Beispiele für Einträge im Tumba! Logfile.

```
900903;guest;136.159.xxx.xxx;null;en;medieval literature;search_url
900951;guest;161.53.xxx.xxx;8ad;en;("kohnen bernardo");search_sim
900952;guest;161.53.xxx.xxx;8ad;en;("kohnen bernardo");view_full
900996;guest;85.75.xxx.xxx;mlg;en;("homer");search_res
901119;guest;209.11.xxx.xxx;ljj;en;(title all "commentaria in psalmos davidicos");view_full
901367;guest;89.44.xxx.xxx;b60;en;("romania");search_sim
902846;guest;143.169.xxx.xxx;null;en;("bossaert leo");search_sim
903779;guest;83.33.xxx.xxx;83e;en;("pomegranate fertilization");search_sim
```

Abbildung 2: Beispiele für Einträge im TEL Logfile (ohne Tag und Uhrzeit).

1 http://www.tumba.pt.
2 http://www.theeuropeanlibrary.org.
3 http://www.bildungsserver.de.
4 http://www.dipf.de.

```
f64.alicedsl.de - - [09/Nov/2009:00:23:09 +0100] "GET /zeigen.html?seite=5892 HTTP/1.1" 200 22436
http://www.bildungsserver.de/zeigen.html?seite=2521 "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.1.5)
Gecko/20091102 Firefox/3.5.5"
80d.superkabel.de - - [09/Nov/2009:00:26:28 +0100] "GET /db/fwulesen.html?Id=200006289 HTTP/1.1" 200 16301
http://www.google.de/search?hl=de&source=hp&q=+landes+filmstelle&btnG=Google-Suche&meta=&aq=f&oq
=&fp=6013614429992176 "Mozilla/5.0 (Windows; U; Windows NT 6.0; de; rv:1.9.1.4) Gecko/20091016 Firefox/3.5.4
(.NET CLR 3.5.30729)"
937.googlebot.com - - [09/Nov/2009:00:27:09 +0100] "GET /db/ffach2.html?fach=2&Rnum=12&Snum=3 HTTP/1.1" 200 16019
- "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 5bd.ono.com - -
[09/Nov/2009:00:30:46 +0100] "GET /db/mlesen.html?Id=42021 HTTP/1.1" 200 180746  - "Java/1.6.0_13"
8f4.primacom.net - - [09/Nov/2009:00:30:45 +0100] "GET /zeigen.html?seite=771 HTTP/1.1" 200 45871
http://www.bildungsserver.de/metasuche/qsuche.html?feldinhalt1=aktive+medienarbeit&bool1=AND&finden=finden&search
all= ja&datenbanken%5B%5D=dbs_seiten&DBS=1&art=einfach "Mozilla/5.0 (Windows; U; Windows NT 6.0; de; rv:1.9.1.4)
Gecko/20091016 Firefox/3.5.4 (.NET CLR 3.5.30729)"
```

Abbildung 3: Beispiele für Einträge im DBS Logfile (IP Code verändert).

nisstruktur organisiert sind. Ein interner Suchdienst erschließt die Inhalte.
Die DBS-Logs bestehen aus Server-Log-Dateien und erlauben die Analyse der zuletzt besuchten Website und damit der eventuell bei einer Suchmaschine eingegebenen Begriffe ebenso wie die Auswertung der internen Suche. Der Schwerpunkt bei der Analyse soll auf der Beziehung zwischen Anfrage und dem betrachteten Ergebnis liegen. Die Benutzer können sowohl den internen Kommentar betrachten als auch zur kommentierten Web-Site wechseln. Für die weiterführende Log-Datei-Analyse können sowohl der Inhalt der Seiten des Bildungsservers als auch der Inhalt der darin verlinkten Angebote einbezogen werden.

6 Ergebnisse des Query Parsing Tasks

Die Analyse der MSN-Log-Daten für GeoCLEF 2007 offenbart einige typische Eigenheiten von Anfragen-Log-Dateien. Anfragen bestehen meist aus wenigen Wörtern und längere Anfragen kommen kaum vor, vgl. Abb. 4. Die meisten Anfragen bestehen in diesem Log aus drei Wörtern.

Ein weiteres im Internet allgemein sehr häufig auftretendes Phänomen sind Verteilungen nach dem Potenzgesetz. Solche Verteilungen bedeuten, dass Frequenzen sehr unterschiedlich auf Objekte verteilt sind. Dies gilt etwa für die Verteilung von eingehenden Links auf Web-Seiten. In derartigen Verteilungen überwiegen Elemente mit sehr niedrigen Werten während einige wenige Elemente sehr hohe Werte erzielen (Mandl 2006).
Dieses globale Muster findet sich auch beim MSN-Log. Nur sehr wenige Anfragen treten mehrfach und noch weniger sehr häufig auf. Zeichnet man eine derartige Verteilung in eine Grafik, welche die X- und Y-Achse logarithmisch skaliert, dann ergibt sich eine fast gerade Linie. Dies zeigt Abbildung 5 für das MSN-Log. Es ist erstaunlich, dass auch bei einer eher kleinen Menge analysierter Suchanfragen dieses Phänomen so deutlich zu Tage tritt.

Unter den sehr häufigen Anfragen dominieren die sogenannten Navigationssuchen, wie Tabelle 3 illustriert. Mit solchen Anfragen will der Nutzer die Homepage eines bestimmten Angebots finden, sei es eine Suchmaschine, ein soziales Netzwerk oder ein E-Commerce-Angebot. Nachdem die Site gefunden wurde, ist die Aufgabe der Suchmaschine erfolgreich erledigt, das Informationsbedürfnis des Benutzers dagegen eher nicht. In der Regel wird er auf dem gefundenen Angebot weiterarbeiten, beispielsweise könnte er bei der gefundenen Website versuchen sein eigentliches Informationsbedürfnis durch Anfrage in der lokalen Suchmaske zu formulieren.

Das Ziel bei der Bearbeitung des MSN-Logs im Rahmen von CLEF bestand in der geographischen Klassifikation der Suchanfragen. Die Analyse von Li et al. (2007) schätzt deren Anteil auf insgesamt 61 Prozent. Im Detail sollten dann auch einzelne Bestandteile von Anfragen erkannt werden:

- Geographische Einheit
- Geographischer Bezug (z.B. im Süden von, neben)
- Informationskomponente
- Typ der Anfrage (Informations-Anfrage, Gelbe Seiten-Anfrage, Landkarte)

Tabelle 3: Die zehn häufigsten Anfragen im MSN Logfile.

goOGLE
pirates of the Caribbean
yAHOO
Jesse Mccartney
MYSPACE
ebAY
td Canada trust
Yahoo JAPAN
naruto
BANK OF AMERICA

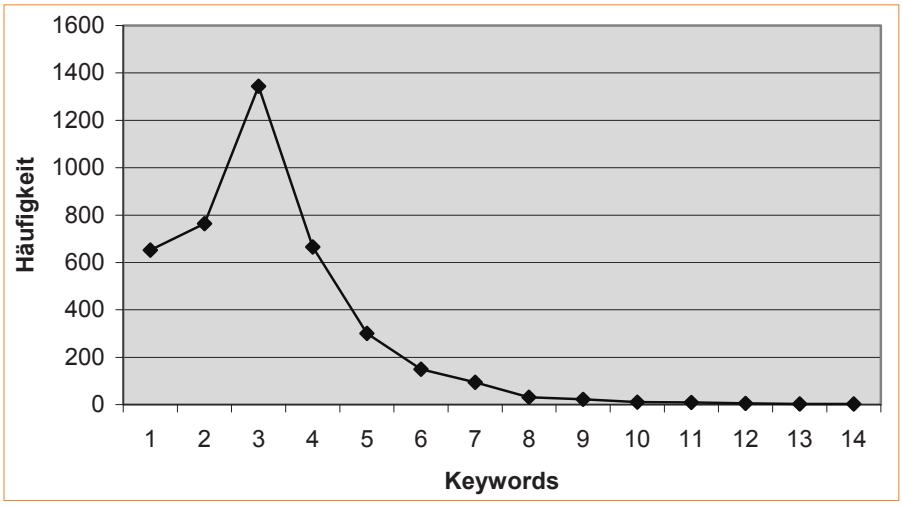


Abbildung 4: Verteilung der Länge der Anfragen im MSN Log.

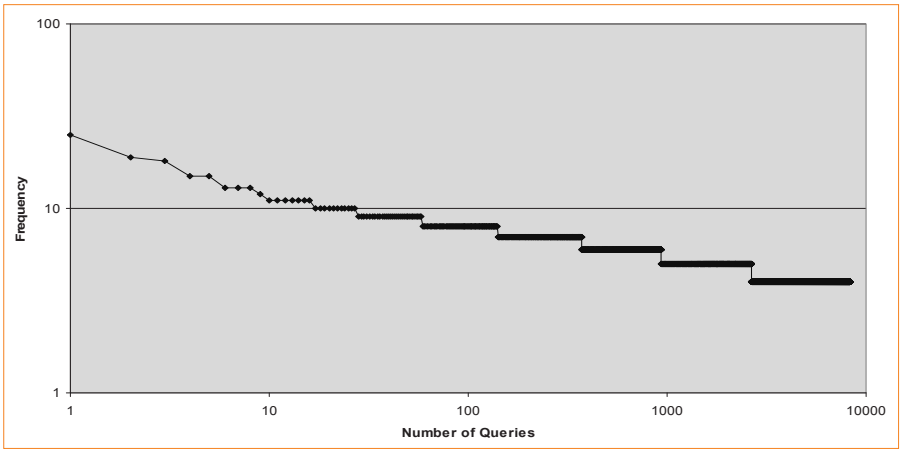


Abbildung 5: Häufigkeitsverteilung der Anfragen.

Die Aufgabe war nicht trivial. Das beste System erzielte ein F-Maß, welches Recall und Precsion kombiniert, von 0,48. Allerdings wurden nur Antworten als korrekt gewertet, welche bezüglich aller Elemente richtig zugeordnet wurden. Dazu wurden 500 Anfragen von je drei Juroren bewertet. Bei Unstimmigkeiten wurde ein Konsens gesucht. Als Ansätze wurden vor allem Listen mit geographischen Namen sowie Regel-Systeme genutzt, welche Anfragen auf typische Muster hin analysieren (Li et al. 2007).

7 Ergebnisse des TEL Tasks

Viele Aspekte des Benutzerverhaltens, die sonst nur am Rande betrachtet werden, rücken im Rahmen des TEL Tasks in den Blickpunkt des wissenschaftlichen Beobachters. Angesichts der Offenheit der Aufgabenstellung reichen die untersuchten Fragestellungen von der Analyse der verwendeten Sprache für die Suchanfrage sowie dem Wechsel der Sprache innerhalb einer Session (Oakes & Xu 2010) über die Identifikation übersetzter Suchanfragen (Bosca & Dini 2010) bis hin zur Visualisierung des Benutzerverhaltens (Lamm et al. 2010) (siehe auch Abschnitt 8).

Ungeachtet der methodischen Unterschiede kommen alle beteiligten Studien zu dem Schluss, dass die Analyse von Such-Logfiles trotz der eingangs genannten Unsicherheiten eine wertvolle Methode darstellt, um die Nutzung von IR-Systemen zu untersuchen. So sehen Ghorab et al. (2010) beispielsweise viele Ansatzpunkte für die Personalisierung und Verbesserung mehrsprachiger Suchen. Wünschenswert wäre auch, dass die Logdateien Informationen über die von den Benutzern angesehenen Dokumente enthalten. Solche Zusatzinformationen könnten wesentlich dazu beitragen, besser zu verstehen, wie natürliches Nutzer- und Nutzungsverhalten aussieht. So ist es durchaus denkbar, dass diese Informationen dabei helfen könnten, nachzuvollziehen, wann eine Suche erfolgreich ist und wann nicht.

Beispielhaft sollen hier die Ergebnisse von Ghorab et al. (2010) kurz dargestellt werden. Ghorab et al. (2010) analysierten drei Aspekte des Benutzerverhaltens: die Reformulierung von Suchanfragen, deren Länge sowie die Abfolge verschiedener Suchaktivitäten. In Bezug auf die Reformulierung von Suchanfragen unterscheiden sie das Hinzufügen und Löschen von Termen, sowie die Veränderung der Terme oder deren Modifizierung. Zu letzterem zählen sie die Termänderungen bei Suchanfragen, die nur aus einem einzelnen Suchterm bestehen. Die verwendete Sprache berücksichtigen sie nicht, mit der Begründung, dass die meisten Suchanfragen in englischer Sprache gestellt werden. Ein wichtiges Ergebnis von Ghorab et al. (2010) ist, dass den Nutzern des TEL Portals dessen Suchfunktionalität nicht bewusst ist. So besteht ¼ der Reformulierungen der Suchanfragen aus dem Hinzufügen, Entfernen oder Ändern von Stoppworten und ein Großteil der Termmodifikationen entfällt auf morphologische Variationen von Substantiven, wie der Verwendung von Singular- bzw. Pluralformen.

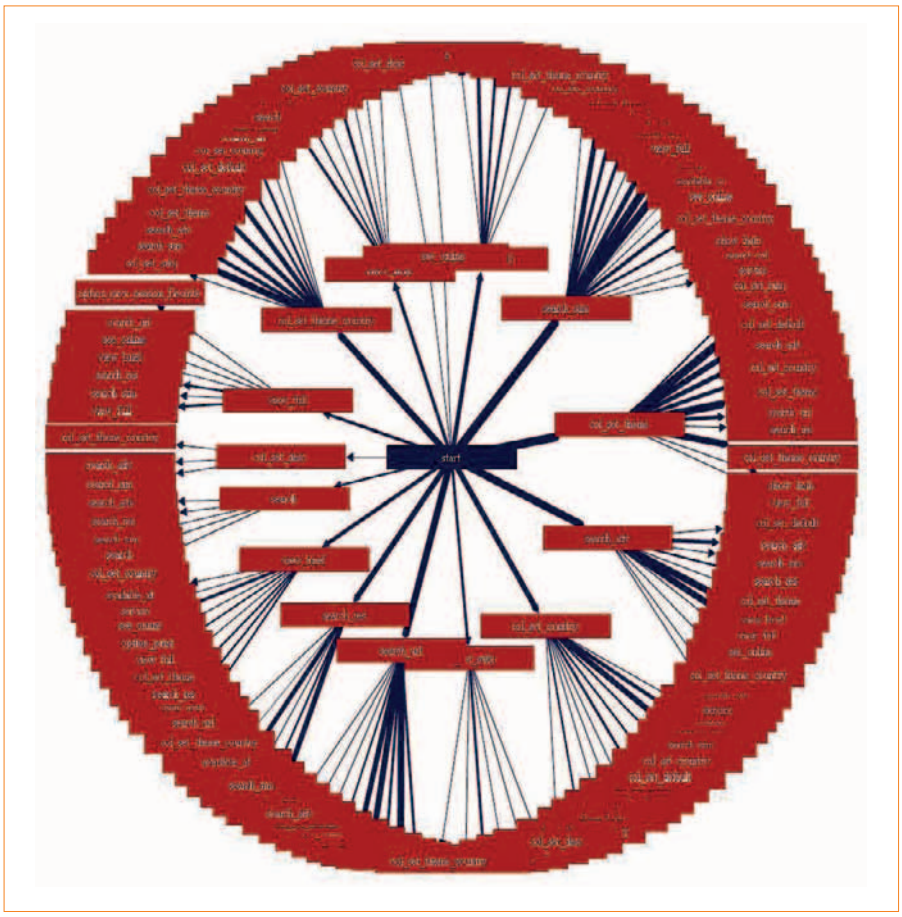


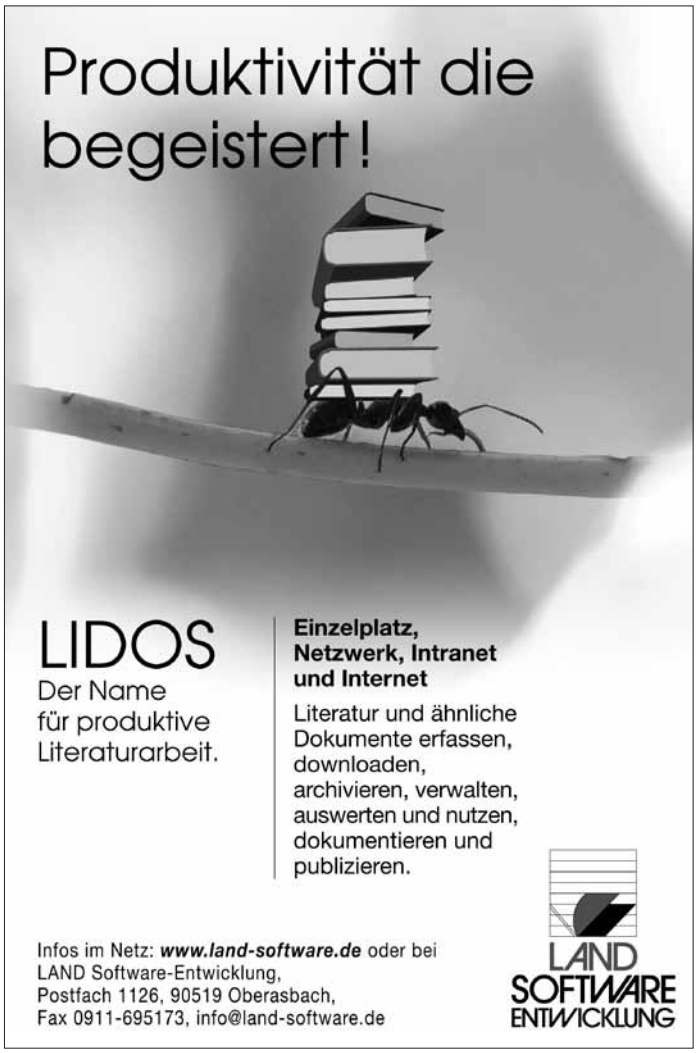
Abbildung 6: Pfad-Darstellung erfolgreicher Benutzer-Sessions.

Weiter untersuchten Ghorab et al. (2010) den Zusammenhang der vom Nutzer gewählten Interfacesprache und der Länge der Suchanfragen sowie der Anzahl und Art von durchgeführten Aktionen, wobei sie sich vorwiegend auf die am häufigsten eingestellten Sprachen konzentrieren: Englisch, Französisch, Polnisch, Deutsch und Italienisch. Anzumerken ist hier, dass Englisch mit mehr als 86% deutlich am häufigsten verwendet wird, was sich zum Teil darauf zurückführen lässt, dass die Sprache standardmäßig voreingestellt ist. Sowohl die Anzahl der Aktionen, als auch die Anzahl der eingegeben Suchanfragen sind für das Englische geringer als in den anderen Sprachen, was die Autoren die Vermutung anstellen lässt, dass Nutzer anderer Sprachen bei TEL mehr Suchanfragen benötigen, um ihr Informationsbedürfnis zu befriedigen.

Die Hypothese, dass sich Nutzer mit verschiedenem sprachlichem oder kulturellen Hintergrund in ihrem Suchverhalten unterscheiden, wird von den Ergebnissen bezüglich der Art der durchgeführten Aktionen gestützt: So wird von Nutzern der italienischen Sprache häufiger die erweiterte Suche verwendet als von den übrigen untersuchten Nutzern und polnische Nutzer wählen bspw. häufiger eine einzige für die Suche zu verwendende Kollektion im Vorfeld der Suche aus, als dies Nutzer mit einem anderen kulturellen Hintergrund tun. Auch die Art der am häufigsten verwendeten Suchterme stützt diese Hypothese. Im Englischen lassen sich 40 Prozent der häufigsten Suchterme der Kategorie Thema (subject) zuordnen und zehn Prozent dem Urheber (Autor, Komponist, etc.), während im Deutschen nur zehn Prozent dieser Suchterme dem Thema und 45 Prozent dem Urheber zugeordnet werden.

8 Werkzeug zur Pfadauswertung

Ebenfalls im Rahmen des TEL Tasks verwendeten Lamm et al. (2010) ein Werkzeug zur visuellen Pfadauswertung, das die Abfolge einzelner Benutzerinteraktionen in einem hyperbolischen Baum zusammenfasst. Diese Darstellungsform ermöglicht es dem Betrachter, einen einfachen Überblick über das Benutzerver-



halten in TEL zu gewinnen, ohne sich in den Details zu verlieren. Die Abbildung 6 zeigt beispielhaft zwei Screenshots des im Folgenden beschriebenen Werkzeugs. Der Startknoten ist Ausgangspunkt aller im Baum dargestellten Interaktionspfade. Jede Hierarchieebene im Baum repräsentiert eine Benutzeraktion in TEL. Klickt der Betrachter auf eine Aktion, wird dieser Knoten in den Mittelpunkt gerückt. Dabei visualisiert die Stärke der Kanten die Anzahl der Sessions mit gleicher Interaktionsabfolge.
Die Ausgangshypothese der von Lamm et al. (2010) durchgeführten Untersuchung bildet die Annahme, dass die Analyse von Logfiles wertvolle Hinweise auf den Erfolg oder Misserfolg einer Benutzer-Session geben kann. In diesem Zusammenhang wird davon ausgegangen, dass bestimmte Benutzeraktionen wie zum Beispiel das Drucken, Speichern oder Versenden eines Suchergebnisses vorzugsweise in erfolgreichen Sessions ausgeführt werden. Die in Abbildung 6 dargestellten Interaktionspfade enthalten jeweils mindestens einen solchen Erfolgshinweis. Anhand der Stärke der Kanten lässt sich sehr schnell sehen, dass der Einstieg über die einfache Suche (*search_sim*) mit am häufigsten verwendet wird. In einem frühen Stadium der Logfile-Analyse könnten derartige Visualisierungen zum Beispiel genutzt werden, um wiederkehrende Muster im Verhalten der Benutzer sichtbar zu machen und darauf aufbauend neue Forschungsfragen zu entwickeln.

Die Darstellung zeigt alle Knoten gleich groß, so dass die Häufigkeiten nicht visuell ablesbar sind. Verberne et al. (2010) stellen die Übergangshäufigkeiten für die TEL2010-Daten in einem nicht interaktiven Netzwerk dar. Dabei zeigt sich, dass View-Aktionen häufiger sind als Such-Aktionen und dass View-Full häufiger auftritt als View-Brief, dass also häufiger die Vollansicht als eine Kurzansicht gewählt wurde.

9 Ausblick

Die Auswertung von Log-Dateien bietet zahlreiche Chancen für die Analyse von Benutzerverhalten. Durch die Bereitstellung von Logs in der Evaluierungsinitiative CLEF können Algorithmen verglichen und weiter entwickelt werden. Der wissenschaftliche Fortschritt wird in der Praxis zu neuen Modellen und innovativen Verfahren der Web-Analyse führen.
Gleichwohl lässt die Auswertung der Wissensquelle Logfiles in vielen Firmen noch zu wünschen übrig. Laut einer Umfrage unter 192 Entscheidungsträgern in Schweizer Firmen werten lediglich sieben Prozent der Unternehmen die Logfiles ihrer Enterprise-Suchen aus, wobei sie auch nur einfache Verhaltensmuster wie den Gebrauch von Funktionen analysieren (Braschler et al. 2009). Auch die Untersuchung von Weischedel & Huizingh (2006) kommt zu dem Ergebnis, dass Webmetriken in Unternehmen noch nicht optimal genutzt werden, obwohl sie Unternehmen bei der Optimierung ihrer Website helfen können. Dazu befragten die Autoren acht Manager eines IT-Unternehmens und es zeigte sich, dass vor allem sehr einfache statistische Auswertungsverfahren bekannt sind und genutzt werden (Weischedel & Huizingh 2006).

Literatur

Anderson, N. (2006) The esthics of using AOL search data. In: Ars Technica. http://arstechnica.com/news.ars/post/20060823-7578.html.

Atterer, R.; Wnuk, M.; Schmidt, A. (2006): Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In: Proceedings of the 15th International Conference on World Wide Web (WWW ,06) Edinburgh, Scotland, May 23 – 26, pp. 203-212.

Bartel, T. (2003): Die Verbesserung der Usability von WebSites auf der Basis von Web Styleguides, Usability Testing und Logfile-Analysen. Berlin: WiKu-Verlag Dr. Stein.

Bischoff, K.; Mandl, T.; Kölle, R.; Womser-Hacker, C. (2007): Geographische Bedingungen im Information Retrieval: Neue Ansätze in Systementwicklung und Evaluierung. In: Open Innovation – neue Perspektiven im Kontext von Information und Wissen? Proc 10. Internationales Symposium für Informationswissenschaft (ISI 2007) Köln 30. Mai - 1. Juni 2007. Konstanz: Universitätsverlag, S. 15-26. Bosca, A.; Dini, L (2010): CACAO project at the LogCLEF Track. In: Multilingual Information Access Evaluation I: Text Retrieval Experiments: Proceedings 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece. Springer [LNCS 6241].

Braschler, M.; Briner, N.; Fischer, H.; Häni, M.; Mandl, T.; Schäuble, P.; Steinbach, M.; Stuker, J. (2009): Enterprise-Search-Systeme im internen Wissensmanagement: Ergebnisse einer Studie zu Perspektiven der Unternehmen. In: Datenbank-Spektrum: Zeitschrift für Datenbanktechnologie und Information Retrieval Heft 30, S. 5-12.

Erinaki, M.; Vazirgannis, M. (2003): Web mining for web personalization. In: ACM Transactions on Internet Technology (TOIT) 3 (1) ACM Press, pp. 1-27.

Ghorab, M.R.; Leveling, J.; Zhou, D.; Jones, G.; Wade, V. (2010): TCD-DCU at LogCLEF 2009: An Analysis of Queries, Actions, and Interface Languages. In: Multilingual Information Access Evaluation I: Text Retrieval Experiments: Proceedings 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece. Springer [LNCS 6241].

Gottron, T.; Lipka, N. (2010): A Comparison of Language Identification Approaches on Short, Query-Style Texts. In: Advances in Information Retrieval, 32nd European Conference on IR Research (ECIR 2010) Milton Keynes, UK, March 28-31, Springer [LNCS 5993], pp. 611-614.

Huang, J.; Efthimiadis, E. N. (2009): Analyzing and evaluating query reformulation strategies in web search logs. In: Proceeding of the 18th ACM Conference on Information and Knowledge Management (CIKM) Hong Kong, China, Nov. 02 - 06, 2009. ACM Press, pp.77-86.

Jansen, B.; Spink, A.; Taksa, I. (eds.) (2009): Handbook of Research on Web Log Analysis. Idea Group Reference: Hershey et al.

Jürgens, J.; Mandl, T.; Womser-Hacker, C. (2010): Das Potenzial von Web Analytics für Usability-Evaluierungen. In: Ziegler, Jürgen; Ziegler; Schmidt, Albrecht (Hrsg.): Mensch & Computer 2010. 10. Fachübergreifende Konfe-

renz für interaktive und kooperative Medien (M&C 2010) 12.-15. Sept. Duisburg. S. 261-270.

Korolova, A.; Kenthapadi, K.; Mishra, N.; Ntoulas, A. (2009): Releasing Search Queries and Clicks Privately. In: International World Wide Web Conference Proceedings of the 18th international conference on World Wide Web (WWW 2009). Madrid, Spain: ACM Press. 171-180. http://www2009.eprints.org/18/.

Lamm, K.; Mandl, T.; Koelle, R. (2010): Search Path Visualization and Session Success Evaluation with Log Files from The European Library (TEL). In: Multilingual Information Access Evaluation I: Text Retrieval Experiments: Proceedings 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece. Springer [LNCS 6241] Preprint in Working Notes. http://www.clef-campaign.org/2009/working_notes/lamm-paperCLEF2009.pdf.

Li, Z.; Wang, C.; Xing, X.; Ma, W.-Y. (2007): Query Parsing Task for GeoCLEF 2007 Report. In Cross Language Evaluation Forum (CLEF) 2007 Working Notes. http://www.clef-campaign.org/2007/working_notes/LI_OverviewCLEF2007.pdf.

Mandl, T. (2006): Die automatische Bewertung der Qualität von Internet-Seiten im Information Retrieval (Habilitationsschrift). http://www.uni-hildesheim.de/~mandl/Publikationen/HabilMandl2006QualityInfoprod.pdf.

Mandl, T. (2010): Evaluierung im Information Retrieval. In: Information. Wissenschaft & Praxis 61 (6-7), S. 341.

Mandl, T.; Carvalho, P.; Gey, F.; Di Nunzio, G.; Larson, R.; Santos, D., Womser-Hacker, Christa (2009): GeoCLEF 2008: the CLEF 2008 Cross-Language Geographic Information Retrieval Track Overview. In: Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, Revised Selected Papers. Berlin et al.: Springer [LNCS 5706] 808-821. Preprint in Working Notes. http://www.clef-campaign.org.

Mandl, T.; Agosti, M.; Di Nunzio, G.; Yeh, A., Mani, I.; Doran, C.; Schulz, J. (2010): LogCLEF 2009: the CLEF 2009 Cross-Language Logfile Analysis Track Overview. In: Peters, Carol; Di Nunzio, Giorgio; Kurimo, Mikko; Mandl, Thomas; Mostefa, Djamel; Peñas, Anselmo; Roda, Giovanna (Eds.): Multilingual Information Access Evaluation I: Text Retrieval Experiments: Proceedings 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece. Revised Selected Papers. Berlin et al.: Springer [Lecture Notes in Computer Science 6241] S. 508-517. Preprint in Working Notes: http://www.clef-campaign.org/2009/working_notes/LogCLEF-2009-Overview-Working-Notes-2009-09-14.pdf

Schulte-Mecklenbeck, M., Murphy, R. O.; Hutzler, F. (2009): Flashlight: Recording Information Acquisition Online (July 13). http://ssrn.com/abstract=1433225.

Sutcliffe, R.; Kruschwitz, U.; Mandl, T. (2010): Web Logs and Question Answering. In: Proceedings Web Logs and Question Answering (WLQA2010) Workshop at the Seventh International Conference on Language Resources and Evaluation (LREC) Malta, 22nd May. 1-7. http://www.csis.ul.ie/wlqa2010/proceedings.htm.

Verberne, S.; Hinne, M; van der Heijden, M; Hoenkamp, E.; Kraaij, W.; van der Weide, T (2010): How does the Library Searcher behave? In: Working Notes 11th Workshop of the Cross-Language Evaluation Forum, CLEF 2010. http://clef2010.org/resources/proceedings/clef2010abs_submission_42.pdf

Weischedel, B.; Huizingh, E. K. R. E. (2006): Website optimization with web metrics: a case study. In: Proceedings of the 8th International Conference on Electronic commerce (ICEC '06) ACM Press, 463-470.

Information Retrieval, Suchstrategie, Informationsverhalten, Forschung, Übersichtsbericht

Die Autoren

apl. Prof. Dr. Thomas Mandl

Thomas Mandl war nach dem Studium an der Universität Regensburg und der University of Illinois at Urbana/Champaign wissenschaftlicher Mitarbeiter am Informationszentrum Sozialwissenschaften in Bonn und ab 1998 wissenschaftlicher Mitarbeiter an der Universität Hildesheim. 2000 wurde er an der Universität Hildesheim über neuronale Netze im Information Retrieval promoviert und habilitierte sich 2006 zu Qualitätsaspekten im Web Retrieval. Seine Forschungsschwerpunkte sind Information Retrieval und Mensch-Maschine Interaktion. Er ist Sprecher der Fachgruppe Information Retrieval in der Gesellschaft für Informatik (GI).
Universität Hildesheim, Institut für Informationswissenschaft und Sprachtechnologie · Lübecker Straße 3 · 31141 Hildesheim · Telefon 05121 883-837 · thomas.mandl@uni-hildesheim.de · www.uni-hildesheim.de/~mandl

Julia Maria Schulz, M.A.

hat nach ihrem Abschluss des Studiengangs Internationales Informationsmanagement an der Universität Hildesheim zunächst als Online Marketing Managerin bei einem Erlebnisgeschenkeanbieter gearbeitet. Seit April 2008 ist sie als Lehrkraft für besondere Aufgaben an der Universität Hildesheim tätig und fertigt ihre Promotion im Bereich des Opinion Minings an.
julia-maria.schulz@uni-hildesheim.de

Nadine Mahrholz, M.A.

studierte Internationales Informationsmanagement an der Universität Hildesheim. Dort ist sie seit 2010 als wissenschaftliche Mitarbeiterin und Doktorandin am Institut für Informationswissenschaft und Sprachtechnologie tätig. Ihre Dissertation schreibt sie im Rahmen des Forschungsprojekts „Bildungsbezogene Informationssuche im Internet und beim Deutschen Bildungsserver".
nadine.mahrholz@uni-hildesheim.de

Katrin Werner, M.A.

ist seit Februar 2009 Doktorandin an der Universität Hildesheim und Absolventin des Studiengangs Internationales Informationsmanagement. Die hier vorgestellte Untersuchung basiert auf ihrer Magisterarbeit, für die sie 2009 mit dem VFI-Förderungspreis ausgezeichnet wurde. In ihrem Promotionsprojekt im Bereich des interaktiven Information Retrieval knüpft sie inhaltlich an diese Arbeit an. Katrin Werner ist daneben federführend im Projekt EFISU tätig, in dem es um den Aufbau eines Weiterbildungsangebots zum Thema Effektive Internetkommunikation und Suchmaschinenoptimierung für KMU geht.
katrin.werner@uni-hildesheim.de

www.dgi-praxistage.de

Informacionnye sistemy v èkonomike : učebnik

Informationssysteme in der Wirtschaft. Lehrbuch
Georgij Nikolaevič Isaev. - 2. izd. - Moskva : Omega, 2009. - 462 S. : Tab., graph. Darst. In kyrill. Schr., russ. ISBN 978-5-370-00855-9.

Aus der langjährigen Lehr- und Forschungserfahrung des Autors, Professor an der „Russischen Staatlichen Universität für Touristik und Dienstleistung" in Moskau, erwuchsen zahlreiche Publikationen. Bereits 2004 konnte die erste in der IWP rezensiert werden.
Interessante Informationen zum russischen Hochschulwesen lassen sich schon aus den erweiterten bibliographischen Angaben entnehmen. Nach wie vor muss das hierfür zuständige Ministerium für Ausbildung und Wissenschaft das Buch als offizielles Lehrbuch für die Fachrichtungen „Finanzen und Kredit" sowie „Buchhaltung, Analyse und Audit" zu-

lassen. Als zweiter Aspekt sind die ebenfalls traditionell tiefgehenden Studienspezialisierungen erkennbar.
Von den 14 Kapiteln behandeln allein zwölf die vielfältig unterschiedlichen Aspekte Automatisierter Informationssysteme (AIS), u.a.

1. Informationsprozesse in der Ökonomie
2. Die Struktur von AIS
4. Technologie der automatisierten Bearbeitung ökonomischer Prozesse
5. Informationssicherheit in AIS
12. AIS für das Versicherungswesen
13. Methodologie der AIS.

(Der Rezensenten fragt sich, ob die in diesem Kapitel besonders zahlreich vertretenen mathematischen Formeln in den entsprechenden deutschen Studiengängen von den Studenten rezipiert werden könnten.)
Dem Charakter als Lehrbuch entsprechend hat jedes Kapitel einen Anhang mit Fragen und Aufgabenstellungen sowie ein Schlagwortregister. Das Literaturverzeichnis verzeichnet von den 71 Nachweisen 65 russischsprachige. Der Rezensent möchte eine Unkenntnis dieser Literatur außerhalb der russischsprachigen Community wie weiland vor dem Sputnikschock annehmen. Gerade durch die tiefgreifende Spezialisierung, sei es in der Grundlagenforschung oder wie hier in der Lehre, erscheint eine vertiefte Kenntnisnahme in der „westlichen" Welt geboten.

Modelirovanie informacionnych resursov: teorija i rešenie zadač : učebnoe posobie

Modellierung von Informationsquellen. Handreichung
Georgij Nikolaevič. Isaev. Moskva : Al'fa-M . Infra-M, 2010. - 223 S. : Tab., graph. Darst. In kyrill. Schr., russ. ISBN 978-5-98281-211-7 und 978-5-16-004232-9

Dieses Lehrmaterial steht Studenten in der Hochschulausbildung für „Dienstleistungen und Touristik" in der Spezialisierung „Angewandte Informatik" zur Verfügung.
In sieben Kapiteln wird die Modellierung von Informationsressourcen (IR) dargestellt.

1. Grundlagen und Bedeutung der Modellierung von IR
2. Modellstrukturen von IR
3. Organisation und Technik der Modellierung von IR
4. Beschreibende Modellierung von IR
5. Formalisierte Modellierung von IR
6. Computergestützte Modellierung von IR
7. Lösungsvarianten von Modellierungen

Auch in diesem Buch fällt die ausgeprägte Verwendung mathematischer Formeln auf. Das Literaturverzeichnis berücksichtigt vorwiegend russischsprachige Literatur, wobei sich Isaev mit zahlreichen Beiträgen als einer der führenden russischen Vertreter der mathematischen Durchdringung auf dem Gebiet der Information und Dokumentation erweist.

Aufruf zur Kandidatur für den DGI-Vorstand

Die **Mitgliederversammlung 2011 der DGI** findet am **8. April 2011 um 14:30 Uhr** im Anschluss an die DGI-Praxistage in Karlsruhe statt. Bei dieser Mitgliederversammlung wird der Vorstand neu gewählt. Der gesamte Vorstand setzt sich laut neuer Satzung aus Präsident oder Präsidentin, Schatzmeister oder Schatzmeisterin, bis zu sechs Vorstandsmitgliedern und dem oder der Vorsitzenden des Fach- und Regionalgruppenbeirats zusammen.
Bis zum **11. Februar 2011** hat jedes Mitglied die Möglichkeit für ein Vorstandsamt zu kandidieren. Dazu muss das entsprechende Formular (unter www.dgi-info.de/aktuelles.aspx zu finden) zur Kandidatur ausgefüllt und elektronisch an die DGI-Geschäftsstelle geschickt werden.

Information, Zeichen, Kompetenz

Fragen an Rafael Capurro zu aktuellen und grundsätzlichen Fragen der Informationswissenschaft

von Linda Treude

Information ist zu einem omnipräsenten Begriff geworden. Wir begegnen ihm in der Wissenschaft ebenso wie in unserem Alltag. Man spricht von der Informationsgesellschaft, von Informationsprozessen, von Informatik und wahlweise von Informationsflut oder -mangel. Information wird als Ding, als Eigenschaft, als Relation oder Ware betrachtet. Aber was meinen wir (jeweils), wenn wir von Information sprechen?

Der Philosoph und Informationswissenschaftler Rafael Capurro beschäftigte sich bereits in seiner 1978 erschienenen Dissertation mit dem Titel „Information. Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs“[1] mit den verschiedenen Dimensionen des Informationsbegriffs. Die Motivation zu einer solchen Arbeit gab Capurro vor allem die fehlende philosophische Auseinandersetzung mit dem Informationsbegriff in Deutschland. Der konkrete Anlass jedoch war die Lektüre der Schriften Carl Friedrich von Weizsäckers, bei der er auf Sätze stieß, die ihm die Bedeutung des Problems klar vor Augen stellten. Weizsäcker formulierte es wie folgt:

> „Man beginnt sich [...] heute daran zu gewöhnen, daß Information als eine dritte, von Materie und Bewußtsein verschiedene Sache aufgefaßt werden muß. Was man aber damit entdeckt hat, ist an neuem Ort eine alte Wahrheit. Es ist das platonische Eidos, die aristotelische Form, so eingekleidet, daß auch ein Mensch des 20. Jahrhunderts etwas von ihnen ahnen lernt.“[2]

Dieser ideengeschichtliche Rückgriff gab Capurros Werk seine besondere Ausrichtung. Rafael Capurro setzt sich seitdem besonders mit Problemen der Informationsethik auseinander. Angesichts der technischen Entwicklung und ihrer Folgen ist heute zu fragen, wie Begriffsklärung, Anwendungsmöglichkeit und gesellschaftliche Bedeutung in Bezug gesetzt werden können. Im Umgang, vor allem mit digitalen Medien ist scheinbar eine neue Kompetenz oder sogar eine neue Kulturtechnik erforderlich. Als Reaktion darauf entstehen im bibliothekarischen Bereich immer neue Projekte, die den Begriff der „Informationskompetenz“ im Titel tragen. Derartige Vermittlungskonzepte, welche eine steigende Konjunktur erfahren, sind mit der angesprochenen terminologischen Unklarheit konfrontiert und zeigen so eine Lücke zwischen Theorie und Praxis im informationswissenschaftlichen Diskurs auf. Welche Forderungen an die Informationswissenschaft ergeben sich daraus? Die Diskussion dieser Problematik und deren Lösungsansätze, wie sie beispielsweise die Semiotik bieten könnte, bilden den Hintergrund für das folgende Interview.

Herr Capurro, heutzutage wird „immer mehr Gewicht auf das Erzeugen von reinen Informationen und immer weniger Gewicht auf das Erzeugen‘ informierter Gegenstände‘ gelegt.“ So formuliert es Vilém Flusser in seinem Vortrag „Die Informationsgesellschaft“. Sehen Sie diese Diskrepanz ebenso und was macht die Frag-würdigkeit der Information in Zeiten des World Wide Web aus?

Wenn Flusser mit „informierten Gegenständen“ die Herstellung intelligenter, das heißt digitalisierter und vernetzter Artefakte meint, dann hätten wir heute eher das umgekehrte Problem, das mit dem Schlagwort „Internet der Dinge“ gekennzeichnet wird. Das Cyberspace ist keine von der alltäglichen Welt abgetrennte Sphäre, wie sie von John Perry Barlow 1996 in Davos verkündet wurde, sondern die digitale Weltvernetzung dringt in alle Lebensbereiche ein. In diesem Sinne leben wir in digital ‚in-formierten‘ Gesellschaften, wobei der Anteil und die Fähigkeit der Menschen damit zu recht zu kommen, unterschiedlich ist. Deshalb führen wir seit Jahren auch in der Informationsethik eine Diskussion um die ‚digitale Spaltung‘, die mehr eine Frage des technischen Zugangs zum Internet ist da diese Art von Spaltung mit anderen gesellschaftlichen Spaltungen, Differenzen und Spannungen verknüpft ist.[3]

Den Begriff Informationskompetenz (bzw. information literacy) als Schlüsselkompetenz im Umgang mit Informationen tragen zur Zeit geradezu inflationär Projekte des Bildungs- und Bibliothekswesens im Titel. Dies suggeriert, dass man heute spezielle Fähigkeiten haben müsse, um den Anforderungen einer „Informationsgesellschaft“ gewachsen zu sein. Für wie brauchbar halten Sie den Begriff und auf welche Kompetenz(en)[4] verweist er Ihrer Meinung nach?

Wenn man diesen Begriff nicht auf den Umgang mit verschriftlichten Informationen einschränkt, sondern auf ein theoretisches und praktisches Sichauskennen mit den vielen Formen der Digitalisierung und der Handhabung digitaler *codes* und *devices* erweitert, dann finde ich die Frage nach der Entwicklung dieser Fähigkeiten bereits in der schulischen Ausbildung sehr wichtig. Es müsste dabei kritisch reflektiert werden wie digitale Technologien in verschiedene Bereiche des Alltags eindringen und welche positiven und negativen Auswirkungen sie haben oder haben können. Ich meine damit konkret die Erziehung zu einem verantwortlichen Umgang zum Beispiel mit Blogs, E-Mail oder *social software* bis hin zur kreativen Nutzung des Internet für Lernprozesse und sozial-politisches Engagement.

Die Demokratie wird zukünftig vermehrt durch interaktive digitale Prozesse bestimmt sein. In diesem Sinne bedeutet

1 Capurro, Rafael: Information. Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs. München: Saur, 1978.

2 Weizsäcker, Carl F. v.: Sprache als Information. In: ders.: Die Einheit der Natur, München: DTV, 1974, S.51. Dieser Vortrag wurde in der Vortragsreihe Die Sprache der Bayerischen Akademie der schönen Künste in München und Berlin 1959 gehalten.

3 Siehe dazu Scheule, Rupert; Capurro, Rafael; Hausmanninger, Thomas (Hrsg.): Vernetzt gespalten. Der Digital Divide in ethischer Perspektive. München: Fink, 2004, S.17.

4 Der Kompetenzbegriff geht u.a. auf den Linguisten Noam Chomsky zurück. Bei ihm bedeutet Kompetenz ('competence') die grundsätzliche Kenntnis der Regeln einer Sprache im Gegensatz zu Performanz ('performance'), als die konkrete Anwendung der Sprache. Chomsky, Noam: Aspects of the Theory of Syntax. Cambridge, Mass. : MIT Press, 1965.

Informationskompetenz auch Demokratiekompetenz und nicht bloß, dass man fähig ist, eine Suchmaschine zu benutzen, die, wie am Beispiel von *Google*, so wenig eine bloße Suchmaschine ist wie ein *smartphone* kein bloßes Telefon mehr ist. Ich glaube also, dass wir unterschiedliche Informationskompetenzen (im Plural!) in verschiedenen theoretischen und praktischen Bereichen thematisieren müssen, ohne diese Vielfalt auf eine Grundkompetenz reduzieren zu können. Jeder Bereich hat eigene Qualitätsstandards und Kreativitätspotentiale, so dass die Informationskompetenzen sich dynamisch und auf verschiedenen Ebenen entwickeln müssen. Schließlich sollten wir bedenken, dass Informationskompetenzen im Sinne von interaktiver Kommunikation zu verstehen sind. In einer globalisierten Welt sind diese Fähigkeiten mit ihren jeweiligen intellektuellen, technischen, rechtlichen und ethischen Dimensionen beinah überlebensnotwendig. Dazu gehören zum Beispiel das Sprachenlernen, sich mit Menschen anderer Kulturen auseinandersetzen und dabei die Sensibilität für unterschiedliche Normen und Werthaltungen zu schärfen. Die Aufgabe einer *éducation sentimentale* ist medial gesehen komplexer geworden. Die Formen und sozialen Konventionen Gefühle auszudrücken, in digitalen und nicht-digitalen Medien werden ständig neu definiert, auch im globalen Maßstab.

Hans Blumenberg untersucht in seinem Werk Die Lesbarkeit der Welt die hermeneutische Hintergrundmetapher des Buches. Welche Beziehungen bestehen ihrer Meinung nach zwischen Informationskompetenz und Literalität?

Der spanische Philosoph José Ortega y Gasset (1883-1959) hat einmal gesagt, dass die Demokratie „Tochter des Buches" ist. In seiner Schrift „Die Aufgabe des Bibliothekars" gehalten beim Internationalen Bibliothekarkongress am 20. Mai 1935 in Paris, schreibt er wörtlich:

„Die demokratische Gesellschaft ist Tochter des Buches, es handelt sich um den Erfolg des von einem Schriftsteller geschriebenen Buches über das von Gott verkündeten Buches sowie auch über das Gesetzesbuch der Autokratie".[5]

Die Frage ist nun, welche demokratiepolitischen Formen – denn es gibt ja nicht die *eine* Demokratie, sondern verschiedene Formen derselben – heute und in naher Zukunft die interaktiven digitalen Medien ermöglichen, jenseits der von den Massenmedien des 20. Jahrhunderts bestimmten Demokratieverständnisse und -ausformungen. Zur Pressefreiheit gesellt sich heute die Freiheit – manche sprechen sogar von einem Menschenrecht – des Zugangs zum Internet. Diese Frage wurde bekanntlich beim *World Summit on the Information Society* sehr kontrovers diskutiert. Wir befinden uns mitten in einer Umbruchzeit, in der die Möglichkeiten der politischen Partizipation mittels interaktiver Technologien sich zu entfalten beginnen. Mit anderen Worten, es geht heute nicht mehr nur um die „Lesbarkeit der Welt", sondern um die Mitteilungskompetenz in einer Welt, die nicht nur lesbar, sondern auch gestaltbarer und digital ‚in-formierbar' geworden ist. Auch hier, bei der digitalen Informierbarkeit der Welt, befinden wir uns in den Anfängen vergleichbar vermutlich mit der Gutenbergzeit. Die Menschen wollen nicht nur informiert und ‚in-formiert' werden, sondern sie wollen selber informieren und ‚in-formieren' so wie damals als sie nicht nur die ‚maßgebenden' Bücher lesen, sondern auch selber Bücher schreiben wollten.

Das Begriffsaufkommen von Informationskompetenz verhält sich korrelativ zu der Entwicklung des WWW oder der digitalen Medien- und Kommunikationstechnik generell. Daher stellt sich die Frage wie relevant das Trägermedium oder die Form generell für den Inhalt ist. Was sagen Sie in diesem Kontext zu McLuhans viel zitierter These „the medium ist the message"?

Das schließt sich an das vorher Gesagte an. Ein Dualismus zwischen Inhalt und Medium lässt sich nicht aufrecht erhalten. Schon seit Platons Zeiten als er im Dialog „Phaidros" die erste Medienkritik des Abendlandes verfasste mit dem Paradoxon, dass im Mythos von der Erfindung der Schrift durch den ägyptischen Gott Theut, der dem griechischen Hermes entspricht, Platon die Schrift gegenüber dem mündlichen *logos* abwertet, unter anderem mit der Begründung, dass Schriften sich nicht gegenüber dem Leser verteidigen können da ihr „Vater" abwesend ist. Das zielt auf den Vorteil des lebendigen Dialogs, in dem der Philosoph als Erzieher sich auf das individuelle Vorverständnis des anderen sowie auf seine – es waren vor allem athenische männliche Bürger an diesen Dialogen beteiligt – Lernentwicklung einstellen kann. Der französische Philosoph Jacques Derrida hat diese Abwertung der Schrift als „Logozentrismus" kritisiert. Was man also mündlich sagt, hat nicht denselben Wert, dieselbe Qualität und vor allem dieselbe Wirkung als wenn man es schriftlich mitteilt. Ein bestimmtes Medium ermöglicht nicht nur unterschiedliche Formen der Verbreitung, sondern auch der Deutung. Dabei muss man nicht an ein hierarchisches Modell wie bei Platon denken. Die Differenz zwischen Inhalt und Träger ist uns durch die griechische Philosophie mitgegeben. Wir haben inzwischen gelernt, sie nicht unkritisch zu akzeptieren. McLuhans Satz „the medium is the message" steht im Kontext seiner Erörterungen von „heißen" und „kalten" Medien im Hinblick auf die größere oder kleinere Arbeit, die der Rezipient beim Verstehen einer Botschaft leistet, etwas was in den Erörterungen McLuhans manchmal kontraintuitiv wirkt. Es kommt letztlich weniger auf den Inhalt der Botschaft, so McLuhan, als eben auf das Medium selbst an. McLuhan war nicht nur ein Medien- sondern auch ein Botschaftstheoretiker. Hinter seiner Medientheorie verbirgt sich eine Boten-/Botschaftstheorie, wobei die Frage was Boten und Botschaften sind, sowohl in der Medienwissenschaft als auch in der Informationstheorie m.E. kaum beachtet wurde. Man versteht Medien wie ein Fluidum – *‚medium'* ist das lateinische Wort für Mitte – zwischen einem Sender und einem Empfänger. Eine andere Deutung von Medium besagt, dass es sich weniger um eine Mitte als um ein Mittel oder Werkzeug handelt. In beiden Fällen bleibt aber jene Dimension des Bringens einer Botschaft ungefragt, die in Luhmanns Theorie der Kommunikation mit dem Ausdruck „Mitteilung" oder „Sinnangebot" thematisiert wird. Ich halte Luhmanns Kommunikationstheorie mit der Differenz zwischen „Mitteilung", „Information" und „Verstehen" für die anspruchsvollste Kommunikations- und Medientheorie, über die wir zur Zeit verfügen.

Im Kontext meiner Magisterarbeit[6] habe ich die verschiedenen Dimensionen des Informationsbegriffs als theoretische Grundlage für das Konzept der Informationskompetenz untersucht. Laut ihrem „Trilemma" ist es unmöglich, einen einheitlichen Informationsbegriff zu formulieren. Wie beurteilen Sie die Versuche zur Schaffung eines einheitlichen Informationsbegriffs?

Die Frage lautet zunächst was „einheitlich" heißt und dann wer einen solchen einheitlichen Informationsbegriff braucht. In vielen anderen Bereichen arbeiten wir mit unterschiedlichen nicht-einheitlichen Begriffen wie zum Beispiel der Arbeitsbegriff selbst oder der Begriff der Kraft oder der Masse oder der Geschwindigkeit und so weiter, die im Laufe der Entwicklung von wissenschaftlichen Theorien vor allem seit der Neuzeit gegenüber dem alltäglichen Gebrauch neu oder anders bestimmt wurden als der alltägliche

5 „La sociedad democrática es hija del libro, es el triunfo del libro escrito por el hombre escritor sobre el libro revelado por Dios y sobre el libro de las leyes dictadas por la autocracia." Siehe Ortega y Gasset, José: Misión del bibliotecario. Madrid: Revista de Occidente, 1962, S.33.

6 Die Arbeit mit dem Titel *Das Konzept Informationskompetenz. Ein Beitrag zur theoretischen und praxisbezogenen Begriffsklärung* wurde am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin eingereicht und wird 2011 in der Reihe *Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft* veröffentlicht.

Wortgebrauch. Wortgebrauch und wissenschaftliche Definitionen in eingeschränkten Bereichen und im Rahmen einer bestimmten Theorie müssen also nicht einheitlich sein. Es gibt aber ein Sich-gegenseitig-Bedingen vom Alltagsgebrauch und präziser Definition, worauf zum Beispiel Carl Friedrich vom Weizsäcker im oben erwähnten Vortrag hingewiesen hat. Hans-Georg Gadamer, der Gründer der philosophischen Hermeneutik in Anschluss an Martin Heidegger, hat immer wieder darauf hingewiesen. Gerade die angebliche Ungenauigkeit und Veränderung des Alltagsgebrauchs bietet oft der Wissenschaft Grundlage für neue Definitionen, die über eine verkrustete wissenschaftliche Sicht hinausgehen. So war es zum Beispiel mit dem durch Claude Shannon geprägten Informationsbegriff, der dem Alltagsgebrauch entgegengesetzt ist aber im Rahmen der Shannonschen Theorie gute Dienste leistet. Das Problem entsteht aber dann, wenn andere Wissenschaften, unter ihnen auch die Informationswissenschaft selbst, mit diesem Informationsbegriff wenig anfangen können da sie andere Phänomene im Blick haben, bei denen zum Beispiel die semantische und pragmatischen Aspekte, die Shannon ausschloss, im Vordergrund stehen. So ist also im Laufe der letzten fünfzig Jahren eine Fülle von Informationsbegriffen entstanden, die sich teilweise überlagern teilweise aber äquivok sind. Das gab meinen Wiener Kollegen Peter Fleissner und Wolfgang Hofkirchner Anlass vom „Capurroschen Trilemma" zu sprechen, das eigentlich auf die aristotelische Unterscheidung zwischen univoken, analogen und äquivoken Bedeutungen basiert.

Es war nämlich Aristoteles derjenige, der musterhaft zu Beginn beinah einer jeden Untersuchung sich fragte, in welchen verschiedenen Bedeutungen ein Begriff gebraucht wird und in welchem semantischen Zusammenhang sie miteinander stehen. Am bekanntesten und für die Philosophie einflussreichsten Analyse der verschiedenen Bedeutungen eines Begriffs gilt seine Aussage, dass der Seinsbegriff sich vielfältig aussagen lässt („to ón légetai pollachós") etwa im Sinne von Potentialität und Aktualität oder von Materie und Form oder von Ursache und Wirkung und so weiter. Er setzte damit eine Diskussion im Gang, die bis heute anhält.

Man kann das Trilemma dadurch auflösen, dass eine bestimmte Definition, die in einem Wirklichkeitsbereich angesiedelt ist, als die erste und maßgebende nimmt – das nennt man dann das *primum analogatum* – und alle anderen Anwendungen und Definitionen nur im Vergleich, also im analogischen Sinne versteht. Das ist die von Peter Janich befürwortete Lösung, indem er den Informationsbegriff im zwischenmenschlichen Bereich primär ansiedelt und dadurch seine analogische Anwendung zum Beispiel in der Biologie nur in einem eingeschränkten und eher irreführenden Sinne zulässt. Das hat den Nachteil, dass man eine Hierarchie von Bedeutungen schafft, die für andere nicht-soziale Bereiche nicht annehmbar oder nicht nützlich ist, da man ungern mit Analogien wissenschaftlich argumentiert. Eine andere Form der Auflösung des Trilemmas besteht darin, die verschiedenen Begriffe in den unterschiedlichen Wissenschaften ‚einfach' Nebeneinander stehen zu lassen. Das kann, muss aber nicht zu Äquivokationen führen. Ich ziehe eine Auflösung nach Art der „Sprachspiele" und „Familienähnlichkeiten" von Wittgenstein vor. Die abendländische Philosophie führt am Beispiel des Begriffs „forma" (Griechisch „idéa" oder „eidos" oder „morphé") vor, welche Vor- und Nachteile die Verwendung eines abstrakten Formbegriffs hat, worauf vor allem Platons Schüler Aristoteles hingewiesen hat. Die platonischen und aristotelischen Bestimmungen des Formbegriffs sind, wie von Weizsäcker bemerkte, mit der Problematik des heutigen Informationsbegriffs eng verwandt. Ich bin diesen Weg in meiner Dissertation gegangen später aber skeptisch geworden, da mir diese Lösung zu sehr unter den Vorgaben der Metaphysik stehen schien.

In Bezug auf das Trilemma: Inwiefern wäre es zur Begriffsklärung des Informationsbegriffs hilfreich die pragmatische Maxime von Peirce anzuwenden?

Wenn wir das vorher Gesagte bedenken, würde ich eher sagen, dass der jeweilige Sprachgebrauch nicht metaphysisch am Verhältnis von Sache und Zeichen orientiert ist, sondern eine dritte Instanz, nämlich den von Peirce eingeführten „Interpreten" einschließt, womit das „semiotische Dreieck" entsteht. Dies macht möglich, von einer Ebene auf die nächste zu wechseln. So kann zum Beispiel in der Informationswissenschaft Information als Zeichen verstanden werden, welches auf der Objektebene auf die Bedeutung einer Mitteilung in Relation auf

Rafael Capurro

Bevor Rafael Capurro 1972 nach Deutschland kam, hatte er in Chile und Argentinien als Mitglied des Jesuitenordens Geisteswissenschaften (Altgriechisch, Latein, Rhetorik, Kunstgeschichte, Literatur, Psychologie, Theologie) und Philosophie studiert und mit einem „Lizentiat", etwa dem Meister entsprechend, abgeschlossen. Er verließ die Jesuiten 1970 und arbeitete in der Dokumentationsabteilung einer Forschungsinstitution in Buenos Aires. Dort begann sein Interesse für Dokumentationsmethoden. Deshalb bewarb er sich im Rahmen der wissenschaftlich-technischen Zusammenarbeit zwischen Deutschland und Argentinien für eine zweijährige theoretische und praktische Ausbildung am Lehrinstitut für Dokumentation (LID) in Frankfurt am Main sowie an der Zentralstelle für Atomkernenergie-Dokumentation (ZAED), einer Vorgängereinrichtung des heutigen FIZ Karlsruhe. Aufgrund der politischen Wirren in Argentinien nach der Rückkehr des ehemaligen Präsidenten Juan Perón blieb er in Deutschland und arbeitete als Dokumentar in der ZAED.

Der Informationsbegriff wurde seinerzeit in der Philosophie in Deutschland kaum beachtet bis auf die Aktivitäten der Professoren Alwin Diemer und Norbert Henrichs im Zusammenhang mit der philosophischen Dokumentation an der Universität Düsseldorf. Rafael Capurro wandte sich 1976, angeregt durch die Schriften Carl Friedrich von Weizsäckers, an Norbert Henrichs mit der Frage, ob er bei ihm mit einer ideengeschichtlichen Arbeit über den Informationsbegriff promovieren könnte. Seine Arbeit mit dem Titel „Information. Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs" (München: Saur 1978) wurde mit „magna cum laude" bewertet. Einige Jahre später versuchte er einen systematischen Zugang zur philosophischen Begründung der Informationswissenschaft mit dem Titel „Hermeneutik der Fachinformation" (Freiburg i.Br. 1986). Die Arbeit wurde 1989 als Habilitationsschrift von der Universität Stuttgart angenommen. Von 1986 bis 2009 arbeitete Capurro als Professor für Informationswissenschaft und Informationsethik an der Hochschule der Medien (HdM) bzw. an der Vorgängereinrichtung.

Von 1989 und 2004 war er Privatdozent für Praktische Philosophie am Philosophischen Institut der Universität Stuttgart, von 1980 bis 1986 persönlicher Referent des wissenschaftlich-technischen Geschäftsführers des FIZ Karlsruhe, Dr. Werner Rittberger. Während dieser Zeit entstand sein Interesse an ethischen Fragen der Informationswissenschaft und der aufkommenden Informationsgesellschaft, die aber Mitte der achtziger Jahre fast genau so exotisch klangen wie die philosophische Beschäftigung mit dem Informationsbegriff. 1999 gründete er das International Center for Information Ethics (ICIE).

den Interpreten (Informationserzeuger, -vermittler oder -rezipient) bezogen ist. Aber auf der nächsten Ebene, sagen wir die eines Dokuments, bedeutet Information (als Objekt) das Mitgeteilte, sofern dies in Dokumenten niedergeschrieben wurde. Denen stehen andere Formen von Zeichen zur Verfügung, wie etwa die Speicherung in einer Bibliothek oder einem Computer und auch spezielle Formen von Erzeugern (Autoren), Vermittlern (Dokumentaren) und Nutzern. Was also hier ‚einheitlich' ist, ist nicht *ein* Informationsbegriff im Rahmen eines Anwendungsgebietes, sondern ein formales triadisches Verhältnis das auch zur Bestimmung eben dieses Gebietes dient. In ähnlicher Form hat die Hermeneutik das Verhältnis von Autor, Text und Interpret reflektiert. Ich habe diese Ansätze in meiner „Hermeneutik der Fachinformation" (S. 61-67) anhand der von Norbert Henrichs für die Informationswissenschaft eingeführten semiotischen Ansätzen thematisiert. Wenn es also bei Peirce heißt:

> „Überlege, welche Wirkungen, die denkbarerweise praktische Relevanz haben könnten, wir dem Gegenstand unseres Begriffs in unserer Vorstellung zuschreiben. Dann ist unser Begriff dieser Wirkungen das Ganze unseres Begriffes des Gegenstandes."[7]

dann bedeutet das, dass durch das Dreieckverhältnis ein Seinsgebiet oder eine Regionalontologie, wie man sie in der Philosophie nennt, konstituiert wird. Peirce argumentiert in der Kantischen metaphysikkritischen Tradition, indem er die „Wirkungen" von Begriffen „nur" auf den durch die Sinnlichkeit bestimmten Bereich einschränkt und keine metaphysischen Objekte zulässt. Die „praktische Relevanz" bezieht sich also auf diese Möglichkeit der Anwendung von Begriffen im Bereich der Sinnlichkeit oder der sinnlichen „Anschauung" wie Kant es ausdrückte. Dem steht eine im Kantischen Sinne sozusagen rein spekulative Relevanz gegenüber, die Kant lediglich im Bereich der Moral mittels des kategorischen Imperativs zulässt. Letzteres setzt aber wiederum einen anderen Begriff von Pragmatismus oder von Pragmatik voraus, als der von Peirce anvisierte.

Im sogenannten DIKW-Modell[8] wird der Informationsbegriff in einer hierarchischen Beziehung mit Daten, Wissen und Weisheit gesehen. In welcher Relation steht dann der Zeichenbegriff zum Informationsbegriff? Sind diese beiden Begriffe eindeutig voneinander trennbar und wenn ja wie?

Eine solche Hierarchie ist im vielfachen Sinn fragwürdig. Sie suggeriert, das aus etwas irgendwie etwas anderes entstehen kann. Man gibt Daten in den Computer ein und irgendwie wird daraus Information. Und aus vielen Informationen entsteht dann irgendwie Wissen und dann, wenn man viel weiß, wird man sogar weise. Ich parodiere dieses Modell, um zu zeigen, dass hier mit angeblich klaren Begriffen argumentiert wird, während in Wahrheit, die Sachverhalte, um die es geht, höchst problematisch sind. Bekanntlich definiert sich sogar die Philosophie nicht als Weisheit, sondern als Liebe zur ihr, so dass sie aus Wissen allein nie hervorgehen kann, ja alles Wissen immer unzureichend bleibt und derjenige, der sich als Weise ausgibt, entweder ein Gott oder ein Scharlatan ist. Die griechische Philosophie kennt eine Wissenstaxonomie, die zwischen Meinung (‚*doxa'*), praktischem Wissen (‚*empeiría'*), Wissen, wie man etwa herstellt (‚*techné'*), wissenschaftlichem Wissen (‚*episteme'*), ethischem Wissen (‚*phronesis'*) und Wissen der ersten Prinzipien (‚*sophia'*) unterscheidet. Diese klassische Taxonomie wurde im Laufe der Jahrhunderte verändert, sogar teilweise umgedreht wie etwa im Falle von Kant und Peirce, die das ‚höhere' Wissen oder Weisheit nicht auf einer metaphysischen Ebene für sinnvoll oder möglich hielten. Für Kant bedeuten die „Daten" der Erfahrung etwas anderes als für Platon oder für einen heutigen Physiker, der solche sinnliche Daten erst durch die Vermittlung einer komplexen Apparatur gewinnt und nur so wahrnehmen kann.

Eine gute phänomenologische Erörterung des Zeichenbegriffs findet sich in Heideggers *Sein und Zeit* und zwar in § 17 „Verweisung und Zeichen"[9]. Heidegger erörtert dort diese Begriffe in Zusammenhang mit der Weise, wie die Dinge als „Zuhandenes", das heißt, in ihrer praktischen Verwendbarkeit, sind. Man kann zwar, so Heidegger, den Zeichenbegriff formalisieren im Sinne von „Zeichensein für…" so wie Husserl es tat. Aber Heidegger argumentiert, dass Zeichen „zunächst" im Sinne von „Zeuge" oder „Zeigen" zu deuten sind, etwa von Wegmarken, Signale oder Trauerzeichen. Sie verweisen also auf etwas. Es gibt aber andere Arten von Verweisungen wie zum Beispiel Symbole, Ausdruck oder Bedeutung. Ferner gibt Heidegger zu bedenken, dass Verweisungen zwar immer Beziehungen sind, aber der Beziehungsbegriff formalerer Art ist. Alle Verweisungen sind Beziehungen, aber nicht umgekehrt so wie nicht jede Verweisung ein Zeichen ist. Ich zitiere Heidegger:

„Unter den Zeichen gibt es Anzeichen, Vor- und Rückzeichen, Merkzeichen, Kennzeichen, deren Zeigung jeweils verschieden ist, ganz abgesehen davon, was je als solches Zeichen dient. Von diesen „Zeichen" sind zu scheiden: Spur, Überrest, Denkmal, Dokument, Zeugnis, Symbol, Ausdruck, Erscheinung, Bedeutung. Diese Phänomene lassen sich auf Grund ihres formalen Beziehungscharakters leicht formalisieren."[10]

Als Beispiel für das besondere „Verweisen" von Zeichen führt Heidegger den „roten, drehbaren Pfeil" bei Kraftwagen („Sein und Zeit" erschien im Jahr 1927!). Dieses Zeichen ist ein „Zeug" mit dem Charakter des „Um-zu" also mit einer bestimmten Dienlichkeit: „Es ist zum Zeigen". Wichtig ist aber dabei, dass ein Zeigen des Zeichens (in diesem Fall) „im ganzen des Zeugzusammenhang von Verkehrsmitteln und Verkehrsregelungen" steht (ebda.). Mit anderen Worten, „ein" Zeichen gibt es nicht, sondern ein Zeichen befindet sich immer in einem bestimmten praktischen Weltzusammenhang oder einer „Zeugganzheit". Das worauf das Zeichen verweist, ist also der Sinn oder die Bedeutung des Zeichens innerhalb einer Zeugganzheit für denjenigen (oder für diejenigen menschlichen Kulturen), die einen Unterschied zwischen „Zeug" und „Zeichen" machen, was zum Beispiel im Falle eines „Fetisch" oder „Zauber" nicht zutrifft.

Das Entdecken des Zeichens als Zeug bedeutet also, dass Dinge im „alltäglichen Besorgen" in der Seinsart des „Zuhandenen" wahrgenommen werden. Dies erlaubt wiederum, dass beim Wegfallen diese Verwendbarkeit, aus welchen Gründen auch immer, die Zeichen als bloß „vorhanden" losgelöst von ihrem praktischen Zusammenhang gesehen werden. Man kann daraus schließen, dass wenn „Welt" oder „Weltlichkeit" im Heideggerschen Sinne durch einen solchen Bedeutungs- und Verweisungszusammenhang konstituiert wird, das Mitteilen von solchen Bedeutungen vor dem Hintergrund einer gemeinsamen „Welt" erst möglich und sinnvoll wird. Dieses Mitteilen lässt sich dann auch im alltagssprachlichen Sinne als Informieren und das Mitgeteilte als Information bezeichnen.

In der Semiotik wird der Zeichenbegriff durch Relationen definiert und als Prozess (oder wie Peirce formuliert durch „semiosis") bestimmt. Könnte man ähnlich dazu schon Informationen selbst als Prozesse begreifen?

Man müsste aber darauf achten, dass der Zeichenbegriff nicht verabsolutiert wird und daraus eine Zeichenmetaphysik aufgebaut wird. Diese ist zwar möglich, indem alles was ist, in irgendeiner Weise ALS Zeichen verstanden wird. Aber da-

7 Peirce, Charles S.: How to make our ideas clear. In: Popular Science Monthly, Vol. 12 (1), 1878, S.286-302. Online verfügbar unter: http://www.peirce.org/writings/p119.html.

8 Data-Information-Knowledge-Wisdom, siehe dazu Ackoff , Russell L.: From Data to Wisdom. In: Journal of Applied Systems Analysis, Volume 16, 1989, S.3-9.

9 Siehe Heidegger, Martin: Sein und Zeit. Tübingen : Niemeyer, 1976.

10 Heidegger, Martin: Sein und Zeit. Tübingen : Niemeyer, 1976, S.78.

durch werden andere Möglichkeiten, nämlich die Dinge nur (!) ALS Zeichen wahrzunehmen, versperrt. Wir sind zwar, wie Octavio Paz schreibt, „grammatische Affen“[11], das heißt, wir haben die Fähigkeit, alles was uns begegnet, ALS Zeichen für etwas wahrzunehmen aber, so Octavio Paz, „Die Baumallee hat keinen Namen und diese Bäume sind keine Zeichen: sie sind Bäume“[12]. Wenn wir also, aufgrund unserer „grammatikalischen“ Verfasstheit, Bäume ALS Zeichen wahrnehmen, sie Benennen und dadurch auch Sinn erzeugen – und den Sinn an andere „grammatische Affen“ mitteilen, sie also darüber informieren – dann bedeutet das immer eine beschränkte Sicht der Dinge sowie auch des jeweiligen Bedeutungszusammenhang oder der „Welt“, von wo aus wir die Bäume ALS „etwas für…“ deuten. Zugleich ist es aber so, dass wir nicht auf eine bestimmte „Welt“ fixiert sind, sondern diese überschreiten oder endlich transzendieren können. Mit anderen Worten, wenn wir etwas ALS etwas wahrnehmen, dann tun wir das in einem Horizont von Unbestimmtheit jenseits dessen, was die Dinge für uns in ihrer Konkretheit bedeuten, und somit jenseits dessen, was wir denken, was sie ‚eigentlich' sind. Dieser Horizont von Unbestimmtheit nennt Heidegger „Sein“ (auch „Seyn“), wobei hier das Paradoxon entsteht, dass etwas, was eigentlich kein ‚etwas' ist, sprachlich gekennzeichnet werden soll. Die Welt ist zwar die jeweilige Ganzheit von Bedeutungs- und Verweisungszusammenhängen, in der wir als In-der-Welt-sein eingebettet sind, aber sie lässt die Fragilität unserer Deutungen erscheinen, indem die Dinge nie in einem bestimmten ALS erschöpft werden können. Wir leben immer in *einer* Welt.

Es scheint also so zu sein, dass für dieses „Es gibt“ (Heidegger) der Welt kein Bezeichnen möglich ist, denn Deutungen und Zeichen beziehen sich immer auf Dinge im Wozu ihrer Zuhandenheit. Die Welt selbst ist aber kein Ding, sowenig wie wir als diejenigen die, soweit wir wissen, von diesem „Un-Ding“ sozusagen angesprochen (eher angefühlt) werden. Wittgenstein schrieb am Schluss des „Tractatus“: „Wovon man nicht sprechen kann, darüber muss man schweigen“. Aber indem er das sagt und schreibt, spricht er „davon“, obwohl er nicht „darüber“ sprechen will und kann. Heidegger hat auf diese Unterscheidung zwischen „Sprechen von“ und „Sprechen über“ aufmerksam gemacht und nach einer Art des Sprechens und vor allem des Schweigens gesucht jenseits der philosophischen und wissenschaftlichen Begrifflichkeit, die man zum Beispiel in der Dichtung oder in der Musik finden kann.

Ist durch Ihre Ablehnung eines einheitlichen Informationsbegriffs, auch die Möglichkeit einer „universellen Informationswissenschaft“ nach Søren Brier, die eine empirische Wissenschaft der Kognition und Kommunikation werden soll, ausgeschlossen?

Man kann Kognition und Kommunikation auf der Grundlage einer „cyber-kybernetischen“ Informationstheorie, wie sie Brier in der Nachfolge von Peirce entwickelt, erschließen und empirisch analysieren. So etwas findet man auch bei Luhmann, wenngleich auf Systeme bezogen, die Sinn erzeugen können. Die großen philosophischen Systeme, wie etwa die von Aristoteles, Thomas von Aquin oder Hegel, haben so etwas angestrebt. Heidegger nennt solche Versuche, „das Seiende im Ganzen“ zu erfassen, Metaphysik. Kant meinte, dass wir nicht unterlassen können, solche metaphysischen Fragen zu stellen, aber zugleich uns bewusst sein müssten, wo die Grenzen der reinen Vernunft bei der Beantwortung solcher Fragen liegen. Auch Peirce war sich darüber bewusst in Bezug auf die „sinnlichen“ Wirkungen der Begriffe.

Ich glaube, dass unsere Epoche durch die Phänomene der Artifizialität und der Kommunikation bestimmt ist, und nicht mehr von Natur und Geschichte, wie das im 19. und im 20. Jahrhundert der Fall war. Zugleich bedeutet diese Faszination, dass wir die Grenzen der künstlichen Herstellbarkeit nicht mehr sehen und uns der Naturbegriff abhanden gekommen ist. Das geschieht heute zum Beispiel im Zusammenhang mit der Nanotechnologie und der synthetischen Biologie. Auch die Faszination der Kommunikation verführt uns dazu, die Endlichkeit unseres persönlichen und sozialen Lebens nicht mehr wahrzunehmen und uns auf die reine Gegenwart zu beziehen, ohne die Dimensionen der Vergangenheit und Zukunft mehr richtig zu würdigen und in und von ihnen zu leben. Das kann sowohl individuell als auch gesellschaftlich zu schweren psychischen und sozialen Störungen führen. Man nennt das *information overload*.

Wo sehen sie die zukünftigen Aufgaben der Informationswissenschaft? Kann sie eine ähnliche Entwicklung nehmen wie andere Disziplinen zwischen Natur- und Kulturwissenschaften? Welche Anforderungen stellen sich an diese Disziplin in Zukunft?

Ich glaube, dass die Informationswissenschaft sich verstärkt den gesellschaftlichen Fragen widmen und sich von ihrer Jahrzehnte andauernden Fixierung auf das Information Retrieval befreien sollte. Ansonsten wird sie sich kaum von einer angewandten Informatik unterscheiden. Sie müsste intermedial, interkulturell und epochenvergleichend jene sozialen und materiellen Phänomene untersuchen, die mit der Erstellung, Bewahrung, Erschließung, Weitergabe und Nutzung von Wissensformen aller Art zu tun haben. Das ist zwar dem Anschein nach in der klassischen Definition enthalten, aber das Gebiet wurde vor allem auf das digitale Medium eingeschränkt. Es entstand eine Informationswissenschaft als Information-Retrieval-Wissenschaft. Diese Verengung ist weder sinnvoll noch zeitgemäß und sie passt auch nicht in einen universitären Rahmen, in der die Informatik fest angesiedelt ist.

Ich plädiere deshalb dafür, eine künftige Informationswissenschaft in der Fakultät für Geistes- und Sozialwissenschaften anzusiedeln, wobei dies nicht heißen soll, dass viele Phänomene nicht mit den ‚harten' Methoden der Naturwissenschaften sowie auch mit Hilfe der Informatik behandelt werden könnten und sollten. Man könnte sich für diese Neubestimmung an die „médiologie“ von Régis Debray orientieren, die den Schwerpunkt auf die Materialität der Träger – das nannte man früher „Dokumentation“ – sowie auf die Vermittlungsinstitutionen legt. Ich meine aber auch, dass die Medienwissenschaft und das, was ich „Angeletik“ nenne, also eine empirische Wissenschaft, die sich mit dem Phänomen von Boten und Botschaften auseinandersetzt, zum Kern dieser neuen Informationswissenschaft gehört.

Ferner glaube ich, dass in Zukunft eine immer größer werdende Vielfalt von digitalen und ‚hybriden' Gegenständen aller Art das Leben ganzer Gesellschaften bestimmen werden. Das bedeutet, dass eine künftige Informationswissenschaft sich mit sozialen Fragen der Robotik, Bionik oder Nanotechnologie beschäftigen muss.

Schließlich muss man bedenken, dass es seit Jahren eine interdisziplinäre Gruppe von Wissenschaftlern gibt, die nach einer umfassenden *Information Science* oder *Science of Information* oder *Information Studies* strebt. Das Profil einer gesellschaftlich orientierten Informationswissenschaft müsste deutlich hervorgehoben werden. Dazu können solche interkulturellen, intermedialen und epochenvergleichenden Studien dienen, die bisher vorwiegend auf Bibliotheken und Archive beschränkt blieben. Eine enge Verzahnung mit Medien- und Kommunikationswissenschaft ist die Folge. Die Fokussierung auf die gesellschaftliche Ebene bedeutet auch, dass die ethischen, rechtlichen und politischen Dimensionen der Wissensvermittlung zum Kern dieser Disziplin gehören.

Vielen Dank, Herr Capurro, für das Interview.

11 Paz, Octavio: El mono gramático, Barcelona : Seix Barral, 1974, S.97.
12 Paz, Octavio,1974, S.97.

Literatur

Ackoff , Russell L.: From Data to Wisdom. In: Journal of Applied Systems Analysis 16(1989).

Blumenberg, Hans: Die Lesbarkeit der Welt. Frankfurt am Main: Suhrkamp, 1986.

Brier, Søren: The foundation of LIS in information, science and semiotics. Introducing the conflict between informational and semiotic paradigms. In: LIBREAS, Library Ideas (2006)4, S. 1-27.

Capurro, Rafael: Information. Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs. München: Saur, 1978.

Chomsky, Noam: Aspects of the Theory of Syntax. Cambridge, Mass.: MIT Press, 1965.

Flusser, Vilèm: Die Informationsgesellschaft. Phantom oder Realität? Audio-CD, Supposé: 1996.

Heidegger, Martin: Sein und Zeit. Tübingen : Niemeyer, 1976.

Ortega y Gasset, José: Misión del bibliotecario. Madrid: Revista de Occidente, 1962.

Paz, Octavio: El mono gramático. Barcelona : Seix Barral, 1974.

Peirce, Charles S.: How to make our ideas clear. In: Popular Science Monthly 12(1878)1, S.286-302. Online verfügbar unter: http://www.peirce.org/writings/p119.html.

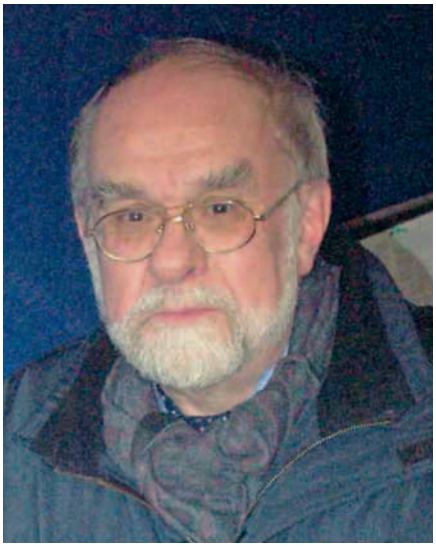
Scheule, Rupert; Capurro, Rafael; Hausmanninger, Thomas (Hrsg.): Vernetzt gespalten. Der Digital Divide in ethischer Perspektive. München: Fink, 2004, S.17.

Weizsäcker, Carl F. v.: Sprache als Information. In: ders.: Die Einheit der Natur. München: DTV, 1974.

Linda Treude
Studentische Mitarbeiterin
Humboldt-Universität zu Berlin
Projekt Dimekon: MLZ und Berufliche Weiterbildung
Telefon (030) 2093 5308
linda.treude@gmx.de

Prof. Dr. Jiri Peter Panyr †

Am 3. November 2010 starb Prof. Dr. Jiri Peter Panyr im Alter von 68 Jahren.
Jiri Panyr wurde am 4. Juli 1942 in Prag geboren. Er bestand 1959 das Abitur mit Auszeichnung, konnte aber aus politischen Gründen zunächst kein Hochschulstudium aufnehmen (Arbeit als Fräser in einer Automobilfabrik). Von 1961 bis 1966 studierte er Mathematik (Nebenfach Informatik) an der Karls-Universität in Prag und schloss dieses mit der Diplomprüfung ab (Dipl.-Math.). Im Rahmen eines anschließenden Studienaufenthalts am Geophysikalischen Institut der Tschechoslowakischen Akademie der Wissenschaften (CSAV, 1966 – 1968) entwickelte er neben seiner Forschungstätigkeit in der Gravimetrischen Abteilung u. a. ein neues Verfahren zur elektronischen Erschließung bibliographischer Daten. Auf Grund der Ereignisse in der Tschechoslowakei emigrierte er 1968 nach Deutschland (Einbürgerung 1978).
Von 1969 bis zu seiner Pensionierung war er bei der Fa. Siemens in München beschäftigt, wo er schwerpunktmäßig mit informationswissenschaftlichen und -praktischen Forschungen und Entwicklungsarbeiten im Umfeld der elektronischen Sprachdatenverarbeitung und Informationsverarbeitung befasst war.
Bei seiner beruflichen Tätigkeit bei der Siemens AG hat er häufig Themen bearbeitet, die erst später anderswo als Forschungsgegenstand aufgegriffen wurden. So hat z.B. seine interaktive Retrievalkomponente auf der Basis einer pseudohierarchischen Deskriptoren- und Dokumentenklassifikation eine Präzisierung der Retrievalergebnisse entlang der Deskriptorenhierarchie ermöglicht. Die Entwicklungen im Rahmen des Pro-



jekts „EWIDAT“ (= Extraktion von Wissen aus Daten, 1987) waren die Vorläufer erst später aufgegriffener Methoden des Data Mining. Beim Visualisierungsprojekt „KOAN“ (= Kontextanalysator) wurde ein bis dahin wenig erforschtes Gebiet der Visualisierung der komplexen informellen Zusammenhänge erschlossen. Über seine Forschungsergebnisse berichtete er damals regelmäßig in den Nachrichten für Dokumentation.
Nach einem informationslinguistischen und informationswissenschaftlichen Ergänzungsstudium in Regensburg und Saarbrücken wurde er 1985 an der Philosophischen Fakultät der Universität des Saarlandes im Fach Informationswissenschaft mit der Gesamtnote ‚magna cum laude’ promoviert (Dr. phil.). Die Dissertation „Automatische Klassifikation und Information Retrieval – Anwendung und Entwicklung komplexer Verfahren in Information-Retrieval-Systemen und ihre Evaluierung“ behandelte ein für die Informationswissenschaft zentrales Thema und gilt inzwischen als ein Standardwerk. Die Methode ist mathematisch-statistisch basiert, das Anwendungsfeld sind Texte bzw. Textcorpora (Sprachdaten).
Im Jahr 2003 wurde er zum Honorarprofessor der Fakultät 5 „Empirische Humanwissenschaften“ an der Universität des Saarlandes ernannt.
Prof. Dr. Panyr war vielseitig engagiert. So war er z. B. Mitglied des internationalen P.E.N.-Klubs, Zentrum der Exilschriftsteller (u.a. Veröffentlichung von Gedichten), wo er auch wegen seiner umfassenden Kenntnisse und Interessen (Literatur, Geschichte, Philosophie) geschätzt wurde.
Neben seiner beruflichen Tätigkeit suchte und fand Prof. Panyr immer wieder (und seit WS 1988/1989 kontinuierlich) als Lehrbeauftragter die Gelegenheit zur Lehrtätigkeit, vorwiegend an der Universität des Saarlandes im Fach Informationswissenschaft, aber auch an den Technischen Universitäten Kosice (Slowakei) und Ilmenau. Für die Tätigkeit als Gastprofessor in Kosice wurde ihm 2002 im Rahmen der Feier zum 50-jährigen Bestehen der TU Kosice die ‚Goldene Medaille’ verliehen. Die vielfältigen Anknüpfungspunkte, die sich zwischen seiner Forschungsarbeit bei der Siemens AG und dem Bereich der Wissenspräsentation, einem zentralen Bestandteil der Informationswissenschaft, ergaben, machten seinen kontinuierlichen Beitrag in der Lehre – vor allem bzgl. ihrer innovativen Inhalte - außerordentlich wertvoll.

Heinz-Dirk Luckhardt, Saarbrücken

Ein XML-basierter Ansatz für Austausch, Speicherung und Retrieval von bibliographischen Metadaten im Bereich Planen und Bauen

Dimitri Busch, Stuttgart

Es wird ein Ansatz zum Austausch, zur Speicherung und zum Retrieval von bibliographischen Sätzen in ICONDA®Bibliographic (International Construction Database) beschrieben. Nach einer Studie bestehender Ansätze wurde das Bibliographic Metadata Format (BMF) entwickelt, das auf XML basiert ist und Datenelemente nach ISO 2709 modelliert. Das neue Format wird nicht nur für den Austausch, sondern auch für die Speicherung der Sätze in der Datenbank verwendet. Der Artikel stellt einen Mechanismus für das Retrieval der XML-basierten Sätze mit dem Microsoft SQL Server dar.

XML based approach for the exchange, the storage and the retrieval of bibliographic metadata in the area of planning and building construction

The purpose of the paper is to describe a XML based approach to the exchange, the storage and the retrieval of bibliographic records in ICONDA®Bibliographic (International Construction Database). After the studies of existing approaches, we have developed the Bibliographic Metadata Format (BMF) that is based on the XML, and emulates data elements according to ISO 2709. The new format is used not only to exchange bibliographic records, but also to store the records in the database. The article presents a mechanism for the retrieval of XML based records with Microsoft SQL Server.

1 Einführung

Das Ziel dieses Artikels besteht in der Beschreibung eines XML-basierten Ansatzes zum Austausch, zur Speicherung und zum Retrieval von bibliographischen Metadaten in der ICONDA®Bibliographic (International Construction Database). ICONDA ist eine englischsprachige Online-Datenbank für den Nachweis und die Beschaffung von Fachliteratur im Bereich Planen und Bauen. Die Datenbank enthält mehr als 803000 Einträge, welche auf Zeitschriftenartikel, Bücher, Forschungsberichte, Konferenzbeiträge und andere Quellen verweisen. Die Datenbank wird von einem internationalen Konsortium aus mehr als 20 Organisationen produziert und vom Fraunhofer-Informationszentrum Raum und Bau (IRB) sowie den Online-Anbietern (Hosts) Dialog und Ovid online angeboten. Die Informationsproduzenten liefern dabei ihre Informationen an das Fraunhofer IRB, in dem die Informationen zentral aufbereitet und an die anderen Hosts weitergeleitet werden. Jeder Host betreibt sein eigenes Information-Retrievalsystem, das die Informationen indexiert und wiederauffindbar macht.

Die ICONDA-Datenbank wird seit 1986 produziert [Bauer/Acevedo-Alvarez 2006]. Das Fraunhofer IRB verwendete von 1999 bis 2009 das Information-Retrievalsystem Verity Search'97 Information Server für das Indexieren und das Wiederauffinden von ICONDA-Einträgen. Obwohl das Retrievalsystem seinerzeit als leistungsfähig galt, war es inzwischen für den wachsenden Datenumfang nicht schnell genug.
Als die Datenbank konzipiert wurde, wurde ein Metadatenformat, ICONDA Communication Format (ICF), für den Informationsaustausch zwischen den Informationsproduzenten und Hosts festgelegt. Das ICF ist auf dem Common Communication Format (CCF) von UNESCO basiert und unterstützt eine Satzstruktur nach der ISO-Norm 2709. Obwohl das ICF eine feine bibliographische Beschreibung von verschiedenen Dokumententypen ermöglicht, ist es nicht einfach, die Daten zu erfassen und zu bearbeiten, weil die Satzstruktur zu kompliziert ist. Das Format hat auch Einschränkungen bei der Darstellung der Originaltitel der Dokumente, weil es keine Umlaute und Sonderzeichen unterstützt. Wegen dieser Probleme liefern die Informationsproduzenten ihre Informationen an das Fraunhofer IRB nicht im ICF sondern in anderen Formaten, z.B. RIS, EXodic usw. Das Fraunhofer IRB muss die Informationen in das ICF konvertieren bevor sie an Hosts weitergeleitet werden.
Die genannten Probleme beim Informationsaustausch können mit Hilfe des XML-Formats gelöst werden, da das Format einfacher zu lesen ist, mit standardisierten Parsern bearbeitet werden kann und verschiedene Zeichensätze unterstützt. Man kann XML nicht nur für den Datenaustausch, sondern auch für die langfristige Speicherung der bibliographischen Daten verwenden. Moderne Datenbank- und Information-Retrivalsysteme unterstützen auch Suche in XML-Dokumenten. Im Folgenden werden verschiedene Ansätze zum Austausch von bibliographischen Metadaten, zum Wiederauffinden und zur Verwaltung von XML-Dokumenten betrachtet. Nach einer Studie der Ansätze wird ein XML-basiertes bibliographisches Metadatenformat und ein neues System für die Speicherung, Indexierung und das Wiederauffinden der Metadaten mit dem Microsoft SQL-Server dargestellt.

2 Stand der Wissenschaft und Technik

2.1 Austauschformate auf Basis ISO 2709

Das ICONDA Communication Format (ICF) unterstützt die Satzstruktur nach der ISO-Norm 2709. Austauschformate, welche die Satzstruktur nach ISO-Norm 2709 oder ähnliche Strukturen implementieren, wie MARC [Guenther 2004], CCF [UNESCO 1988], ICF [ICONDA Agency 1990] und MAB, werden seit vielen Jahren in vielen Bibliotheks- und Informationseinrichtungen weltweit eingesetzt. Im Folgenden wird die Satzstruktur nach ISO 2709 am Beispiel des ICF etwas ge-

nauer betrachtet. Ein ICF-Satz kann logisch in Datenfelder unterteilt werden. Ein Datenfeld enthält eine bestimmte Kategorie der bibliographischen Daten und wird mit einer Feldkennung (Tag) identifiziert. Der Tag 200 identifiziert beispielsweise ein Datenfeld, das einen Titel enthält, und der Tag 300 identifiziert ein Feld, das Namen einer Person beschreibt. Jedes Datenfeld enthält ein oder mehrere Unterfelder. Jedes Unterfeld wird mit einer Kennung identifiziert. Das Unterfeld mit der Kennung „A“ des Feldes 300 enthält beispielsweise einen Nachnamen einer Person. Die physische Struktur eines ICF-Satzes ist komplizierter und enthält außer Datenfeldern ein Label und ein Directory. Das Label enthält Informationen, die für die Identifizierung und für die korrekte Verarbeitung des Satzes notwendig sind. Im Directory werden für jedes Datenfeld seine Kennung (Tag), seine Länge und seine Position im Datensatz angegeben. Die **Abbildung 1** zeigt ein vereinfachtes Beispiel eines Satzes im ICF. Der Satz beschreibt ein Buch. Der Titel des Buchs lautet „Exchange Formats“ und der Autor des Buchs heißt „Max Mustermann“.
Die Zeichenkodierung in der ICF-Implemenierung erfolgt mit der britischen Version des Internationalen Alphabets Nr. 5 (ISO-Norm 646). Wenn ein Originaltitel nicht englische Buchstaben enthält, muss man diese Buchstaben umschreiben bzw. transliterieren.

Label	00100a m 2200054 452
Directory	200002100000003000025000 2100#
Datenfelder	01§AExchange Formats#00§AMustermann§BMax§E020#

Abbildung 1: Satz in ICF/ISO-2709.

2.2 XML-basierte Formate

Die ISO-Norm 2709 beinflusst die Entwicklung modernerer Formate, die auf XML basiert sind. Nach der Art der Strukturierung können XML-basierte Austauschformate in zwei Gruppen unterteilt werden. Zur ersten Gruppe gehören Formate, welche die Satzstruktur nach der ISO-Norm 2709 oder ähnlichen Strukturen modellieren, wie z.B. MARCXML [LoC 2010], MarcXchange [ISO 25577] und MABXML [Kett 2005]. Zur zweiten Gruppe gehören Formate, welche vom ISO 2709 unabhängig sind, wie z.B. MODS [Guenther 2004], BibteXML [Lechner 2006], EXodic [INIST 2010] und Dublin Core.
Die Wahl des Austauschformats kann davon abhängig sein, welche Erfahrungen eine Bibliothek oder Informationseinrichtung mit Austauschformaten hat. Wenn die Arbeitsabläufe bereits auf ISO 2709 basierte Formate orientiert sind, ist es einfacher, XML-Austauschformate zu wählen, welche die entsprechende Satzstruktur modellieren. Wenn die Einrichtung keine Erfahrungen mit ISO 2709 hat, ist es einfacher mit Formaten zu arbeiten, welche von der Norm unabhängig sind, weil bei solchen Formaten die Feldbenennungen aussagekräftiger sind.

2.3 XML und Information Retrieval

Für bibliographische Datenbanken ist das Textretrieval von besonderer Bedeutung. Es ist auch möglich, XML-Dokumente zu indexieren und in XML-Dokumenten zu suchen. Ferber [2003] unterscheidet zwischen textueller Indexierung, Feldindexierung und Strukturindexierung. Bei der textuellen Indexierung wird ein XML-Dokument als Volltext ohne die Berücksichtigung der durch XML vorgegebenen Struktur betrachtet. Man kann dabei zwischen Ansätzen unterscheiden, die die XML-Tags ignorieren, und solchen, bei denen auch nach Tags gesucht werden kann. Bei der Feldindexierung (field-based indexing) werden einzelne XML-Elemente identifiziert und als Felder behandelt. Inhalte der XML-Elemente können auch in eine Faktendatenbank extrahiert und z.B. in einer relationalen Datenbank gespeichert und suchbar gemacht werden. Die Strukturindexierung (structured indexing) nutzt die volle Strukturinformation eines XML-Dokuments zur Indexierung. Der wichtigste Fall der Strukturindexierung ist die Pfadindexierung, bei der die Position eines Elements in der Baumstruktur der Elemente des XML-Dokuments genutzt wird. Die Suche in XML-Dokumenten kann nach verschiedenen Modellen, z.B. nach dem Booleschen Modell oder nach dem Vektorraummodell, durchgeführt werden. Genaueres über XML-Retrieval kann man bei Ferber [2003], Manning/Raghavan/Schütze [2009] und Fuhr/Lalmas [2007] erfahren.

2.4 XML und Datenbanksysteme

Es bestehen Möglichkeiten, XML nicht nur zum Datenaustausch, sondern auch für die langfristige Speicherung zu verwenden. Für die Verwaltung der XML-Dokumente kann man entweder ein natives XML-Datenbanksystem oder ein „XML-enabled“ Datenbanksystem verwenden. Native XML-Datenbanksysteme speichern XML-Dokumente unverändert im Dateisystem. Zu nativen XML-Datenbanksystemen gehören u.a. Tamino, eEXist und Xindice. XML-enabled Datenbanksysteme sind herkömmliche relationale Datenbanksysteme, die durch Erweiterungen für die XML-Verarbeitung ergänzt werden. Zu diesen Erweiterungen gehören u.a. native XML-Datentypen für die Speicherung von XML-Inhalten, interne XML-Parser, eine Unterstützung von XML-Abfragetechniken, wie XPath und XQuery, sowie die Unterstützung textueller Indexierung und Boolescher Suche der XML-Dokumente. Beispiele der XML-enabled Datenbanksysteme sind Microsoft SQL-Server, Oracle, IBM DB2 [vgl. Haufler et al. 2008]. Die Wahl zwischen einem nativen und einem XML-enabled Datenbanksystem hängt u.a. davon ab, ob die Anwendung nur mit XML-Daten arbeitet oder auch Nicht-XML-Daten verwendet. Faktoren wie Neubeschaffungskosten und Einarbeitungszeit sind ebenso zu beachten [Müller-Spitzer/Schneider 2009].

3 Bibliographic Metadata Format (BMF)

Bei der Entwicklung des neuen Metadatenformats wurde davon ausgegangen, dass das neue Format Datenfelder und Unterfelder von ICF/ISO 2709 zu modellieren hat, um den Konvertierungsaufwand zu minimieren. Nach der Studie der o.g. Ansätze zum Informationsaustausch, wurde Bibliographic Metadata Format (BMF) entwickelt, das auf XML basiert ist und bereits existierende Elemente vom ICONDA Communication Format (ICF) und ISO 2709 unterstützt. Eine kurze Einführung in das BMF wurde in [Busch 2009] gegeben. Im Folgenden wird das Schema für BMF-Satz genauer beschrieben (Abbildung 2).
Das BMF verwendet das XML-Element <record> zur Beschreibung des bibliographischen Satzes. Das Element <record> enthält einen Satz von Kinderelementen <datafield>, welche Datenfelder beschreiben. Ein Datenfeld wird anhand eines Tag-Attributes gekennzeichnet. Jedes Element <datafield> enthält ein oder mehrere Elemente <subfield>, welche Unterfelder beschreiben. Ein Element <subfield> hat ein Attribut <code>, das ein Unterfeld kennzeichnet. Das Unterfeld „A“ des Feldes 300 enthält beispielsweise einen Nachnamen einer Person. Die Feld-Tags und Unterfeldcodes sind mit ICF kompatibel. Ein BMF-Satz kann logisch in Segmente unterteilt werden. Die Segmente sind für Publikationen von Bedeutung, die Teile anderer Publikationen sind. Die Segmente werden anhand des Attributes „segment“ des Elements <datefield> gekennzeichnet.

Ein Beispiel des BMF-Satzes, der einen Buchartikel beschreibt, ist in der Abbildung 3 dargestellt. Der Satz enthält zwei Segmente (0 und 1). Das erste Segment

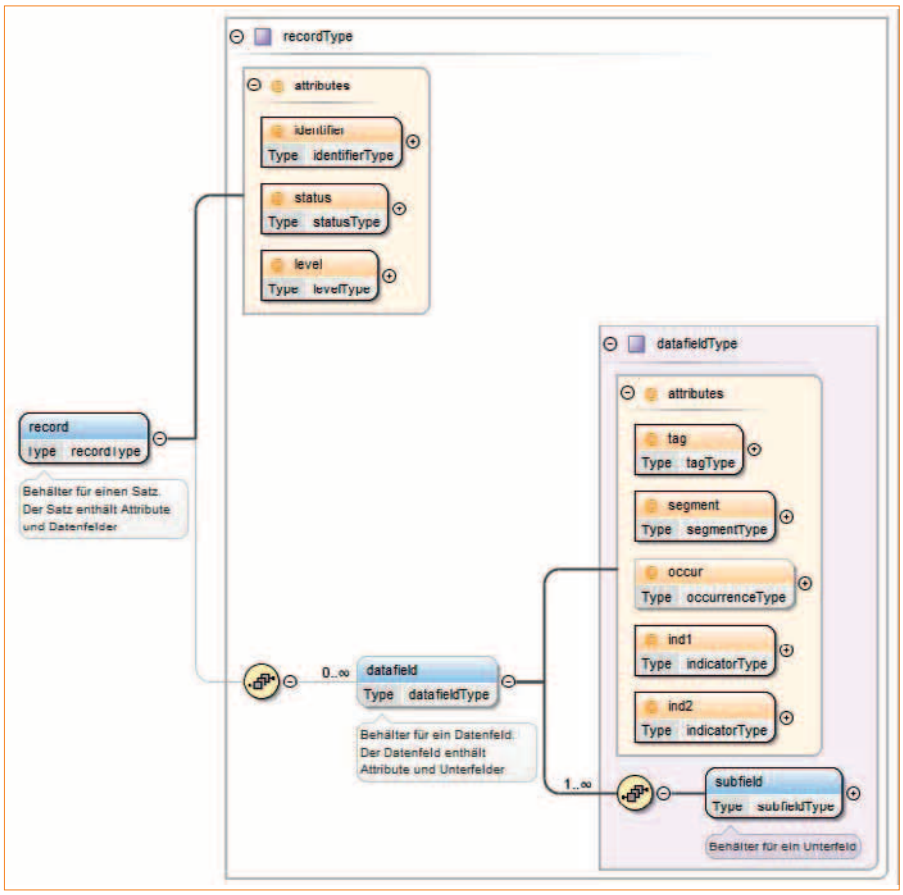


Abbildung 2: Schema eines Satzes im Bibliographic Metadata Format (BMF).

```
<record identifier="id001" status="a" level="a">
.....
        <datafield tag="200" segment="0" ind1="0" ind2="1">
                <subfield code="A">Modern Architecture</subfield>
        </datafield>
.....
        <datafield tag="200" segment="1" ind1="0" ind2="1">
                <subfield code="A">Architectural Styles</subfield>
        </datafield>
        <datafield tag="300" segment="1" ind1="0" ind2="0">
                <subfield code="A">Mustermann</subfield>
                <subfield code="B">Max</subfield>
                <subfield code="E">020</subfield>
        </datafield>
.....
</record>
```

Abbildung 3: Beispiel eines Satzes im Bibliographic Metadata Format (BMF).

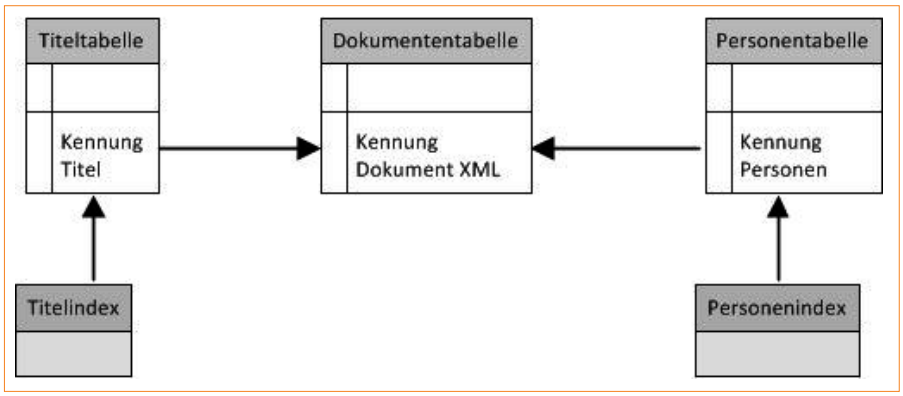


Abbildung 4: Beispiel der Seitentabellen für die Suche in BMF/XML-Dokumenten.

beschreibt den Artikel. Der Artikel hat den Titel „Modern Architecture". Das zweite Segment (1) beschreibt das Buch, das den Artikel enthält. Das Buch hat den Titel „Architectural Styles" und den Autor „Max Mustermann".

4 Speicherung und Retrieval der BMF-Daten

Für die Speicherung und das Wiederauffinden der Einträge wird der Microsoft SQL-Server verwendet. Die Wahl fiel u.a. deshalb auf den SQL-Server als Datenbanksystem, weil das Fraunhofer IRB das System schon seit Jahren für eCommerce- und Wirtschaftsanwendungen benutzt. Der SQL-Server unterstützt Felder vom Datentyp XML zur Speicherung von XML-Daten. Für die Arbeit mit den XML-Feldern können SQL-Anfragen durch XQuery-Anfragen ergänzt werden. Darüber hinaus wird die Volltextsuche mit iFTS (Integrated Full Text Search) unterstützt. iFTS ermöglicht eine Volltextindexierung und die Volltextsuche in Textfeldern (char, varchar usw.), in XML- und CLOB- Feldern. Die Volltextsuche in XML hat jedoch eine Einschränkung. Es ist möglich, in ganzen XML-Dokumenten zu recherchieren, aber es ist unmöglich, in separaten XML-Elementen mit iFTS zu suchen. Eine Lösung des Problems besteht in der Verwendung der Seitentabellen [Lapis 2005]. Jede Seitentabelle enthält Daten von einem oder mehreren XML-Elementen, in welchen die Suche durchgeführt wird. Die **Abbildung 4** zeigt ein Beispiel der Verwendung der Seitentabellen. Die Dokumententabelle enthält in der Spalte „Dokument" BMF-Sätze. Die Dokumententabelle ist mittels eindeutiger Kennung mit den Seitentabellen, „Titeltabelle" und „Autorentabelle", verknüpft. Die Spalte „Titel" der Seitentabelle „Titeltabelle" enthält für jeden Satz (<record>) Daten aus allen Datenfeldern (<datafield>) mit dem Tag 200 (Title). Die Spalte „Personen" der Seitentabelle „Personentabelle" enthält für jeden Satz Daten aus allen Datenfeldern mit dem Tag 300 (Name of Person). Für jede Seitentabelle wird ein separater Volltextindex erzeugt.

In der Abbildung 5 ist ein Beispiel einer SQL-Abfrage dargestellt, welche für die Suche von Dokumenten verwendet wird. Auf diese Abfrage werden Dokumente gesucht, welche die Terme „building" und „construction" im Titel enthalten. Die Abfrage verwendet eine Unterabfrage für die Seitentabelle "Titeltabelle".

Die ICONDA ist als eine serverseitige Web-Anwendung implementiert. Bei der Implementierung der neuen Anwendung wurden die Programmiersprache Java,

```
select * from Dokumententabelle
        where Kennung in
                (select Kennung from Titeltabelle
                        where contains(Titel,'building and construction')
```

Abbildung 5: Abfrage für die Suche in Titeln.

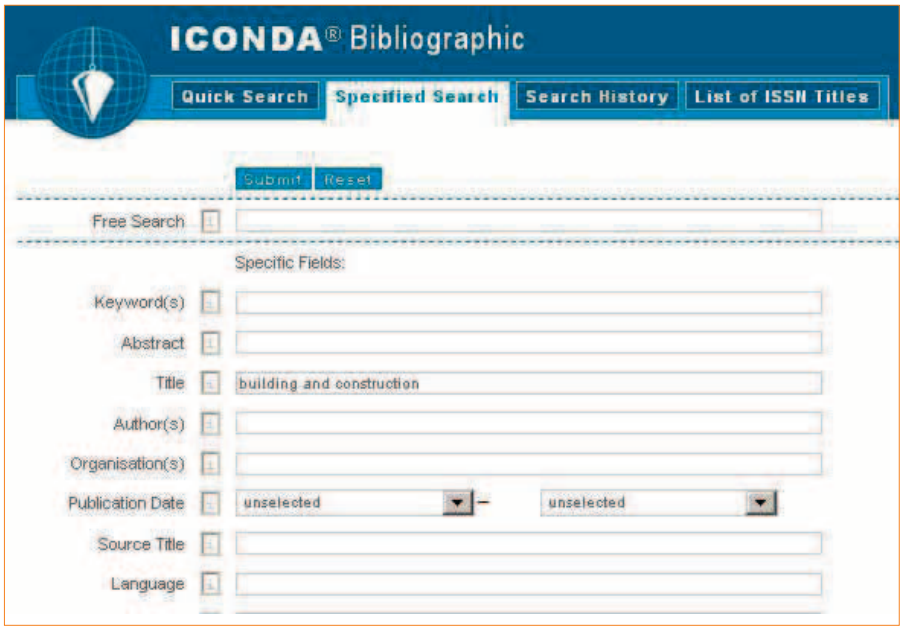


Abbildung 6: Suchformular mit der Anfrage an den Titel.

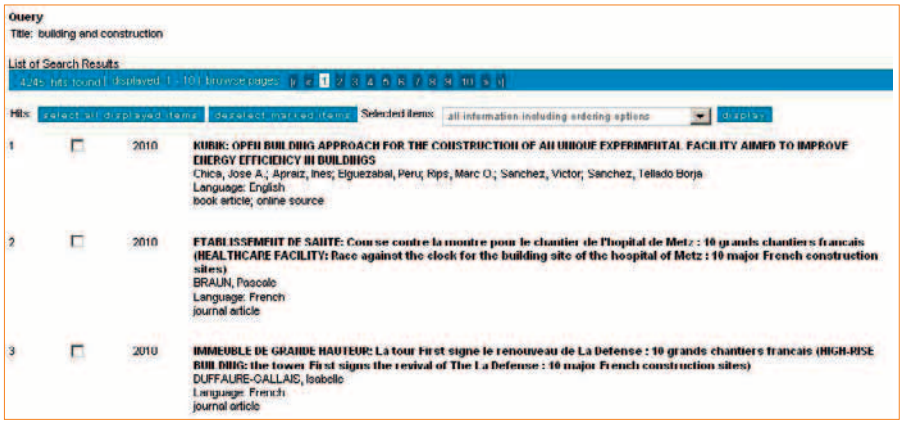


Abbildung 7: Teil der Ergebnismenge für die Anfrage aus Abbildung 6.

die Web-Programmiersprache JavaServer Pages, das Datenbanksystem Microsoft SQL-Server 2008 und der Servlet-Container Tomcat verwendet. Die Anwendung funktioniert unter dem Betriebssystem Microsoft Windows. Die meisten Implementierungsarbeiten wurden in den Jahren 2008 und 2009 durchgeführt.

Die ICONDA-Datenbank ist mittels eines Webbrowsers über das Internet zugänglich. Abbildung 6 zeigt das Suchformular von ICONDA. Das Suchformular enthält eine Abfrage, welche der SQL-Abfrage von Abbildung 5 entspricht. Eine Liste der Suchergebnisse kann man in Abbildung 7 sehen.

5 Migration und Konvertierung der Daten

Das Fraunhofer IRB verwendet seit Jahren selbst entwickelte Programme für die Konvertierung der Daten aus unterschiedlichen Eingangsformaten in das ICF-Format. Um den Programmieraufwand zu reduzieren, werden die meisten dieser Programme weiterverwendet, sodass die Datenumwandlung in das BMF in zwei Schritten durchgeführt wird. Zuerst werden Eingangsdaten mit den bereits vorhandenen Programmen in das ICF-Format umgewandelt. Dann werden die Daten aus dem ICF in das BMF konvertiert. Für die Konvertierung der ICF-Daten in das BMF-Format wurde ein neues Programm in Java entwickelt. Da das BMF-Format Felder und Unterfelder des ICF unterstützt, war es nicht kompliziert, das Konvertierungsprogramm zu entwickeln. Im Folgenden werden die Datenumwandlungsvorgänge genauer beschrieben.

Vor dem Neustart der neuen ICONDA wurden alle Einträge der alten Datenbasis in das ICF-Format exportiert, in das BMF-Format konvertiert und in die neue Datenbank gespeichert. Für den Datenexport in das ICF wurden die bereits vorhandenen Progamme verwendet. Für die Datenkonvertierung vom ICF in das BMF wurde das o.g. neuentwickelte Programm eingesetzt.

Die neuen Sätze der Informationsproduzenten werden momentan in die neue Datenbank, ähnlich der Datenmigration, in zwei Schritten importiert. Zuerst werden die Sätze aus ihren Eingangsformaten, wie z.B. RIS, EXodic usw., in das ICF-Format umgewandelt. Dann werden die ICF-Sätze in das BMF-Format konvertiert und in die Datenbank gespeichert.

Die neue Datenbankanwendung unterstützt auch einen direkten Datenimport vom BMF, d.h. wenn ein Informationsprovider seine Daten im BMF-Format liefert, können diese Daten direkt in die Datenbank importiert werden.

Für den Informationsaustausch mit Hosts unterstützt die ICONDA den Import sowohl in das ICF als auch in das BMF-Format. Es besteht auch die Möglichkeit, Einträge aus der Datenbank nach dem Eingangsdatum über das Internet herunterzuladen.

6 Zusammenfassung und Ausblick

Die BMF-Sätze können auf einfache Weise erstellt und bearbeitet werden, weil sie auf XML basiert sind. Es ist nicht kompliziert, die bibliographischen Sätze von ICF nach BMF zu konvertieren, da das BMF bereits existierende Datenelemente vom ICF unterstützt. Das neue Metadatenformat kann nicht nur zum Austausch von bibliographischen Sätzen, sondern auch zur Speicherung der Sätze in der Datenbank verwendet werden. Es ist möglich, ein schnelles Wiederauffinden der Sätze durchzuführen. Die Datenbank ist mittels Webbrowser über das Internet unter http://www.irb.fraunhofer.de/iconda zugänglich. Der vorgestellte Ansatz bietet folgende Vorteile:

- Die mehrjährige Erfahrung mit dem Austausch der ICONDA-Daten zeigt, dass es für Informationsprovider sehr problematisch ist, die Daten im ICF-Format aufzubereiten, da die Daten eine zu komplizierte Struktur haben. Die Verwendung des neuen BMF-Formats kann den Aufbereitungsaufwand seitens der Informationsproduzenten reduzieren, weil die Satzstruktur ein-

facher ist und für die Aufbereitung und die Bearbeitung der XML-Dokumente viele fertige Tools, z.B. XML-Editoren und XML-Parser, existieren.

- Die Informationslieferung im BMF-Format kann den Konvertierungsaufwand seitens der zentralen Aufbereitungsstelle (Fraunhofer IRB) reduzieren.
- Obwohl ICONDA englischsprachig ist, können Einträge der Datenbank Originaltitel in anderen Sprachen enthalten. Da das ICF-Format nur die britische Variante des Internationalen Alphabets Nr. 5 [ISO 646] unterstützt, müssen nicht-englische Buchstaben, z.B. Umlaute, umgeschrieben oder transliteriert werden. Wenn Informationen im BMF-Format geliefert werden, können die Originaltitel ohne die Umschreibung und ohne die Transliteration angegeben werden, weil XML Zeichensätze, wie beispielsweise UTF-16, unterstützt, die für die Darstellung von Informationen in verschiedenen Sprachen geeignet sind.
- Die Verwendung des Standarddatenbanksystems, Microsoft SQL-Servers, reduziert den Verwaltungs- und Aktualisierungsaufwand und trägt zur Flexibilität der Anwendung bei, da man auch Nicht-XML-Daten, z.B. Protokollierungstabellen, einbinden kann.
- Im Vergleich mit der alten ICONDA-Anwendung, die auf Verity Search'97 basiert wurde, ist die Suche in der neuen ICONDA viel schneller.

Die ICONDA-Datenbank kann zukünftig Metadaten, wie Abstracts und Deskriptoren, nicht nur in English sondern auch in anderen Sprachen enthalten. Ein Satz, der ein deutsches Buch beschreibt, könnte z.B. ein englisches und ein deutsches Abstract enthalten. Dies ist jedoch im ICF-Format unmöglich, weil ein ICF-Satz nur ein einziges Abstract enthalten darf. Das neue BMF-Format unterstützt dagegen mehrere Abstracts in einem Satz. Nicht-englische Metadaten können ebenfalls indexiert und wiederauffindbar gemacht werden, weil der Microsoft SQL-Server das sprachspezifische Indexieren und die sprachspezifische Suche, u.a. für Deutsch, Französisch und Spanisch unterstützt [Coles/Cotter 2009].

Die ICONDA-Datenbank enthält u.a. viele Sekundärdokumente, welche auf Volltexte der Primärdokumente im PDF-Format verweisen. In der aktuellen Anwendung werden nur Sekundärdokumente (Metadaten) und keine Primärdokumente indexiert und wiederauffindbar gemacht. In der Zukunft kann man auch die Primärdokumente wiederauffindbar machen, indem man sie z.B. in binäre LOB-Felder der SQL-Server-Tabellen importiert und von dort mit iFTS indexiert.

Literatur

Bauer, M.; Acevedo-Alvarez, J. (2006): ICONDA: History and Prospects of an international information system. In: Proceedings of the Joint International Conference on Computing and Decision Making in Civil and Building Engineering. June 14-16. 2006-Montreal, Canada, S. 3905-3913.

Busch, D. (2009): XML based format for the exchange and the storage of bibliographic records. Expanded Abstract. In: Deconstructing Babel: Sharing Global Construction Knowledge. Proceeding of the 5th CIB conference Information and knowledge management in Buiding. June 17-19, 2009, Rio de Janeiro, S. 24-26.

Coles, M.; Cotter, H. (2009): Pro Full-Text Search in SQL-Server 2008. Berkley: Apress.

Enderle, J. (2001): XML in relationalen Datenbanken. In: Informatik Spektrum 24(2001)6, S. 357-368.

Ferber, R. (2003). Information Retrieval. Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web. Heidelberg: dpunkt.

Fuhr, N.; Lalmas, M. (2007). Advances in XML retrieval: the INEX initiative. In: IWRIDL'06: Proceedings of the 2006 international workshop on Research issues in digital libraries. New York: ACM, S. 1-6.

Guenther, R. (2004): New and traditional descriptive formats in the library environment. DC2004 IFLA session, 2004. URL: dc2004.library.sh.cn/english/prog/ppt/ifla.ppt [3.12.2010].

Haufler, A.; Staudenecker, A.; Wobser, A. (2008): Vergleich von XML-Datenbanken. Universität Stuttgart. URL: http://elib.uni-stuttgart.de/opus/volltexte/2008/3538/ [3.12.2010].

ICONDA Agency (1990): ICONDA communication format: format for the exchange of records in the frame of the International Construction Database. Fraunhofer- Informationszentrum Raum und Bau (IRB). 5.ed. Stuttgart: IRB Verlag, 1990.

INIST(2010): Exodic format d'échange de l'INIST. Institut de l'Information Scientifique et Technique. Nancy (France). URL: http://exodic.inist.fr/ [3.12.2010].

ISO 646 Standard. Information processing – ISO 7-bit coded character set for information interchange. Geneva: ISO, 1991.

ISO 25577 Standard. Information and documentation. MarcXchange. International Organization for Standardization. Geneva: ISO, 2008.

ISO 2709 Standard. Information and documentation. Format for information interchange". International Organization for Standardization. 3.ed. Geneva: ISO, 1996.

IRB (2009): Bibliographic Metadata Format. Fraunhofer Informationszentrum Raum und Bau IRB. Stuttgart. URL: http://www.irb.fraunhofer.de/download/BMF_09-10-20.pdf [3.12.2010].

Kett, J.(2005): MABxml. Ein XML-Schema für das MAB2-Format. Die Deutsche Bibliothek. URL: http://www.d-nb.de/standardisierung/pdf/mabxml_1_dok.pdf [3.12.2010].

Lapis, G. (2005). XML and relational storage. Are they mutually exclusive? IBM Corporation. In: Proceedings of XTech 2005.

Lechner, Ch. (2006): Aufbau einer Webdatenbank für Bibliographien auf der Basis von PHP, MySQL und XML. München: GRIN-Verlag, 2006.

LoC (2010): MARCXML. MARC 21 XML. The Library of Congress. URL: http://www.loc.gov/standards/marcxml/ [3.12.2010].

Manning, C.; Paghavan, P.; Schütze, H. (2009): Introduction to Information Retrieval, Cambridge University Press, 2008.

Müller-Spitzer, C.; Schneider, R. (2009): Ein XML-basiertes Datenbanksystem für digitale Wörterbücher. In: it-Information Technology 51(2009)4, S. 197-205.

UNESCO (1988): CCF: The Common Communication Format. Paris: UNESCO.

Datenbank, Indexierung, Information Retrieval, XML, ICONDA

Der Autor

Dr. Dimitri Busch

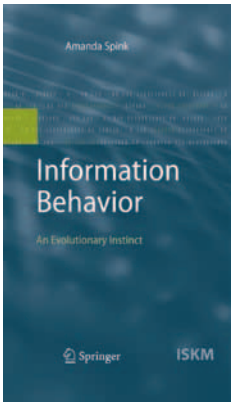
ist im Fraunhofer-Informationszentrum Raum und Bau als Softwareentwickler tätig. Er studierte Wirtschaftsinformatik an der Hochschule für Volkswirtschaft Kiew (Ukraine) und Informationswissenschaft an der Universität Konstanz. An der Universität des Saarlandes wurde er zum Dr. phil. im Fach Informationswissenschaft promoviert. Seine fachlichen Interessen sind Softwareentwicklung; Informationsmanagement; Information Retrieval; Textmining.

Fraunhofer-Informationszentrum
Raum und Bau IRB
Nobelstraße 12
70569 Stuttgart
Telefon 0711 9702565
dimitri.busch@irb.fraunhofer.de

www.dgi-praxistage.de

Information Behavior: An Evolutionary Instinct

Amanda Spink - (Information Science and Knowledge Management, Vol. 16) Heidelberg [u.a.]: Springer, 2010. ISBN: 978-3-642-11496-0. XXVIII, 85 S., 85,55€



Amanda Spink ist eine der weltweit bekanntesten im Feld der Informationswissenschaft tätigen Personen. Diese Bekanntheit hat sie insbesondere durch ihre mehr als zahlreichen Publikationen zur Analyse von Suchmaschinen-Logfiles erlangt, aber auch durch mehrere von ihr (mit-)herausgegebene Werke, darunter „New Directions in Human Information Behavior“ (2006) und „Web Search: Multidisciplinary Perspectives“ (2008). Nach Stationen in den USA und Australien ist sie seit Herbst 2010 Chair of Information Science an der Loughborough University (England).

In Ihrem neuen Buch unternimmt Amanda Spink den Versuch, Informationsverhalten (*information behavior*) aus der Perspektive des evolutionären Fortschritts zu erklären. Der Ansatz ist insofern weit gewählt, als dass eine solche Betrachtungsweise die Kombination der Erkenntnisse aus unterschiedlichen Disziplinen erfordert.

Die Betrachtung ist anhand von drei Fragen geleitet: Einerseits geht es darum, die unterschiedlichen Betrachtungsweisen auf das Informationsverhalten zusammenzuführen. Zum zweiten geht es um die Frage, welche Rolle Instinkte gegenüber der Umgebung für das Informationsverhalten spielen. Und schließlich soll die Frage beantwortet werden, inwieweit sich Fähigkeiten des Informationsverhaltens im Laufe der Zeit verändert bzw. entwickelt haben.

Bei dem beschriebenen Vorhaben handelt es sich in der Tat um ein anspruchsvolles Programm, welches aber entscheidende Fortschritte in der Untersuchung des Informationsverhaltens verspricht. So konstatiert Spink auch, dass der bisherige Fokus der Informationswissenschaft auf der Verbesserung von Informationssystemen liege, wobei grundlegende Fragen des Informationsverhaltens ignoriert werden würden. Diese Kritik bezieht sich vor allem darauf, dass die Informationswissenschaft Ergebnisse aus den Sozial- und Verhaltenwissenschaften sowie aus der Psychologie noch zu wenig aufgegriffen habe.

Das Buch ist in neun Kapitel gegliedert, wobei im grundlegenden zweiten Kapitel ein Modell zum menschlichen Informationsverhalten präsentiert wird, dessen unterschiedliche Ebenen dann in den weiteren Kapiteln „abgearbeitet“ werden. Als Ausgangspunkt nimmt Spink grundlegende evolutionäre Gegebenheiten (wie Sprache, Instinkte, usw.), welche wiederum zu bestimmten kognitiven und sozialen Verhaltensweisen führen. Konkreter zum Informationsverhalten wird es auf den weiteren Stufen, die sich mit den Unterprozessen des Informationsverhaltens (wie suchen, organisieren, benutzen), den „Information Grounds“ (physikalischer oder virtueller Ort) sowie dem Informationsverhalten auf verschiedenen Lebensalterstufen beschäftigen.

Das Modell vereint also unterschiedlichste Aspekte und vermag es sicher, diese Aspekte zu sammeln – mir ist bei der Lektüre allerdings unklar geblieben, ob es sich eben nur um eine Sammlung handelt oder ob man tatsächlich von einem stringenten Modell sprechen kann. Leider vermag der weitere Text diese Unklarheiten auch kaum zu beseitigen, denn die Erläuterungen sind – vor allem, wenn man sie einmal dem doch sehr hohen Anspruch und umfangreichen Arbeitsprogramm gegenüberstellt – doch recht knapp ausgefallen. Das Buch hat gerade einmal 80 Seiten (!); wenn man Trennseiten und die umfangreichen Literaturlisten, die am Ende jeden Kapitels stehen, abzieht, kommt man auf einen Text von gerade einmal 60 Seiten. Man könnte sagen, dass hieraus schon ersichtlich wird, dass das Buch bei weitem nicht halten kann, was es verspricht. So ist auch die einleitende Aussage ”This book takes the reader on a long journey“ (S.1) einerseits verständlich, wenn man die unterschiedlichen Ebenen und dazu die historische Perspektive des Ansatzes in Betracht zieht, bezogen auf den Umfang des Werks verleitet das Zitat dann aber doch eher zum Schmunzeln.

Natürlich ist der Textumfang nicht das entscheidende Kriterium für die Bewertung eines Buchs, im vorliegenden Fall wird der Mangel allerdings an zahlreichen Stellen deutlich, wo man sich als Leser genauere Erläuterungen gewünscht hätte. In der vorliegenden Form ist der Text zu knapp (und damit oft zu ungenau), um dem Leser eine fundierte Bewertung des Geschriebenen zu ermöglichen. Das Buch wirkt in weiten Teilen unfertig; auch finden sich einige Teile, in denen unklar ist, warum nun genau eine von mehreren vorgestellten Richtungen verfolgt wird, zum Beispiel im Kapitel „Lifetime Development“, in dem zwar unterschiedliche Modelle vorgestellt werden, sich die Autorin dann aber ohne hinreichende Begründung einem Modell anschließt. Es folgt dann nochmals die wörtlich aus dem vorangegangenen Text übernommene (knappe) Beschreibung des Modells, welche um einen einzigen Satz zum Thema Informationsverhalten ergänzt wird.

Es ist sehr schade, dass die Autorin aus diesem interessanten und für die Zukunft des Fachs sicherlich entscheidenden Thema nicht mehr gemacht hat. Gerade bei einem Thema, das noch nicht intensiv beackert wurde, ist eine ausführliche Darstellung von großer Bedeutung. Auch in Hinblick auf die Unmenge an Literatur, die in diesem Buch zitiert wird, erscheint die Form der Darstellung in keiner Weise angemessen. Ebenso unangemessen wirkt der Preis von 85 Euro für dieses schmale Werk, welches auch gut in der Form von einem oder zwei längeren Aufsätzen hätte veröffentlicht werden können.

Dirk Lewandowski, Hamburg

Patentinformetrie: Analyse und Verdichtung von technischen Schutzrechtsinformationen

Eine Herausforderung bei der Patentinformations-Recherche in Patentdatenbanken ist die immense Treffermenge, die – kombiniert mit der Komplexität der Dokumente und einer hohen Informationsdichte – einen erheblichen Zeitaufwand bei der Weiterverarbeitung bedeutet. Hier leistet die Informetrie als informationswissenschaftlicher „Methodenkasten“ zur Analyse von Datenbankinhalten gute Dienste. Mit ihren Verdichtungsmöglichkeiten bietet sie Instrumente, um verborgene Informationen in Dokumentmengen aufzuspüren und dabei auch die Informationsmenge zu verkleinern.

Jasmin Schmitz: Patentinformetrie: Analyse und Verdichtung von technischen Schutzrechtsinformationen · Frankfurt am Main, 2010, 328 Seiten, DGI Schrift (Informationswissenschaft 12) ISBN 978-3-925474-69-9, EUR 29,80 (für DGI-Mitglieder EUR 23,90) · Bestellungen an die DGI-Geschäftsstelle: Windmühlstraße 3, 60329 Frankfurt am Main, Fax 069 4909096, E-Mail: publikation@dgi-info.de

eSciDoc Community sorgt für nachhaltige Nutzung der Projektergebnisse

Fachleute für eScience-Strukturentwicklung aus neun Ländern folgten der Einladung von MPDL und FIZ Karlsruhe zu den eSciDoc Days 2010 nach Kopenhagen

Vera Münch, Hildesheim

Der architektonisch beeindruckende Neubau der Royal Library in Kopenhagen ist direkt an das wunderschöne alte Bibliotheksgebäude aus dem Jahr 1906 angebaut und mit ihm zweckmäßig verbunden. Die gelungene Verbindung passte gut zum Thema der eSciDoc Days 2010, auf denen es um einen möglichst reibungslosen Übergang von der Vergangenheit in die Zukunft geht. © Bundgaard, Royal Library

„There is no one size fits all" erteilte Birte Christensen-Dalsgaard, stellvertretende Direktorin der gastgebenden Royal Library Kopenhagen, in ihrem Hauptvortrag gleich zu Beginn der eSciDoc Days 2010 der Illusion eine Absage, dass es eine Lösung gibt, die zu allen Wissenschaften passt: Die verschiedenen Disziplinen brauchen eine generische Infrastruktur, auf der die fach- oder projektspezifischen Anwendungen aufbauen. In ihrem weitsichtigen Vortrag stellte Christensen-Dalsgaard Chancen und Risiken der wissenschaftlichen Kommunikation im Wandel von print zu digital dar und zeigte Konsequenzen und anstehende Probleme auf; beispielsweise die Frage, wie man damit umgehen kann, dass sich die Bedeutung von Information im Laufe der Zeit verändert. „Das semantische Web braucht eine Struktur mit Bedeutung", konstatierte Christensen-Dalsgaard, doch es sei „sehr teuer, die Masse und Vielzahl der digitalen Informationsobjekte mit kontextbezogenen Beschreibungen zu versehen".

Metadaten spielen anstatt Solitär!

Großes Interesse löste der von ihr präsentierte Vorschlag von Luis van Ahn aus, zum Aufbau von beschreibenden Metadaten für Informationsressourcen ein Online-Spiel einzusetzen. Der in Guatemala geborene, als Unternehmer und Professor für Computerwissenschaften an der Carnegie Mellon University in Pittsburgh, USA, tätige Erfinder des Verifizierungsverfahrens CAPTCHA schlägt „Games with a Purpose" (Spiele mit Zweckbestimmung) vor, um Aufgaben, die weder Menschen, noch Computer alleine lösen können, gemeinsam von beiden lösen zu lassen. Google hat van Ahn's Idee für sein ESP Game, das er selbst als „Human Computing", andere als „Crowdsourcing" bezeichnen, für den Google Image Labeler übernommen (http://images.google.com/imagelabeler/). Die Google-Software nutzt den Spieltrieb von Menschen zur weltweiten Gewinnung von Bilder beschreibenden Metadaten. Wer sich anmeldet, bekommt einen Mitspieler aus dem Web zugeteilt. Zwei Spieler sehen und beschreiben dann das gleiche Bild. Stimmen die Begriffe (Himmel, blau, Vogel, fliegend…) überein, gibt es dafür Punkte - und Google hat eine qualitätsgesicherte und tippfehlerfreie Beschreibung des Bildmotivs. Laut van Ahn haben Computernutzer im Jahr 2003 weltweit neun Billionen Stunden damit verbracht, Solitär zu spielen. Der Bau des Panama-Kanals verschlang 20 Millionen Personen-Stunden, also ein Äquivalent von weniger als einem Tag. Warum also nicht ab und zu „Metadaten-Game" spielen?

„Linked Data" verlangen neues Denken

Christensen-Dalsgaard erklärte zudem, dass es im Bezug auf „Linked Data", also die miteinander verbundenen Daten und Informationen im semantischen Web, notwendig sei, „total anders zu denken", was Urheberschaft, Bearbeitungs- und Publikationsbefugnisse betreffe (authority question). Sowohl im Bezug auf das Semantic Web, als auch auf den Aufbau von eScience Libraries ist die stellvertretende Bibliotheksdirektorin der festen Überzeugung, dass als nächstes „Jemand gebraucht wird, der die Verantwortung für die Sicherung des zukünftigen Zugangs zu Information übernimmt und deren korrekte Nutzung und Interpreta-

tion unterstützt". Wer das sein könnte, blieb offen. Doch dürfte sich die eSciDoc Community in ihren Aktivitäten bestätigt gefühlt haben.

eSciDoc Software wird international eingesetzt

2008 haben FIZ Karlsruhe und die Max Planck Digital Library (MPDL) mit den ersten eSciDoc Days in Berlin die eSciDoc Community ins Leben gerufen, um nach dem Auslaufen der Projektförderung durch das Bundesministerium für Bildung und Forschung Mitte 2009 eine nachhaltige Weiternutzung ihrer Entwicklungsergebnisse aus dem Projekt eSciDoc zu erreichen. Außerdem vereinbarten sie die eigenfinanzierte Weiterführung ihrer Zusammenarbeit über das Projektende hinaus. 2009 fanden die zweiten eSciDoc Days in Karlsruhe statt *(IWP berichtete in 60(2009)6-7,395-397)* und nun sieht es so aus, als sei das Vorhaben gelungen, die Ergebnisse der Forschungs- und Entwicklungsarbeit zum Aufbau von eScience Digital Libraries und eResearch-Umgebungen zu etablieren und kontinuierlich weiterentwickeln zu können. Neben MPG-Instituten in Deutschland und mehreren europäischen Ländern, dem Geoforschungszentrum (GFZ) Potsdam, der Leibniz Open Access-Initiative und dem baden-württembergischen Forschungsprojekt BW-eLabs setzen weitere wissenschaftliche Institute und Universitäten in Japan, Spanien und Dänemark die Projektergebnisse ein und entwickeln dabei Anwendungen mit Benutzeroberflächen in ihren Landessprachen. Gefördert von der Europäischen Union, der Deutschen Forschungsgemeinschaft, von Bundes- und Landesministerien sind in Deutschland oder mit deutscher Beteiligung zahlreiche weitere spannende Projekte angelaufen, die Erkenntnisse und Software aus eSciDoc weiternutzen. Rainer Stuike-Prill, Bereichsleiter Marketing und Vertrieb von FIZ Karlsruhe, konnte zu den dritten eSciDoc Days Mitte November 2010 in der Royal Library in Kopenhagen 52 Wissenschaftlerinnen und Wissenschaftler aus neun Ländern begrüßen.

Die eSciDoc Projektleiter Malte Dreyer, MPG (links) und Matthias Razum, FIZ Karlsruhe. erläuterten in Kopenhagen die weltweit laufenden eSciDoc-Aktivitäten. 52 Teilnehmerinnen und Teilnehmern aus neun Ländern hörten vor Ort zu. © FIZ Karlsruhe

Open Source Systembaukasten für eResearch-Umgebungen

Im Projekt eSciDoc haben FIZ Karlsruhe und die MPDL umfassende Grundlagen und modulare Software für den flexiblen Aufbau von eScience-Libraries/eReseach-Arbeitsumgebungen[1] geschaffen. Als neue, Multimedia-fähige Publikations- und Forschungsdatenmanagementsysteme sollen diese Arbeitsumgebungen der Wissenschaft eine neue Art der Wissensnutzung und der Handhabung von Informationsressourcen bringen. Die eSciDoc-Software ist ein Systembaukasten, der nach dem Prinzip einer Serviceorientierten Architektur (SOA) grundlegende Infrastruktur-Services für Speicherung, Verwaltung, Sicherung, Zugangskontrolle und vielfältige Nachnutzung der Daten bereitstellt. Die eSciDoc-Infrastruktur mit der Repository-Software Fedora Commons sowie die darauf aufsetzenden eResearch-Anwendungen können als Module kostenlos unter der Common Development and Distribution License (CDDL) nachgenutzt werden. Ausführliche Informationen zu den verfügbaren Bausteinen für Infrastruktur- und Anwendungsentwicklung sind auf der Webseite www.escidoc.org bereitgestellt (Navigationspunkte eSciDoc Applications, eSciDoc Services, eSciDoc Infrastructure).

Sicherheit der Daten und Weiterverwendung von Anwendungen sind gefragt

Nach Christensen-Dalgaards spannendem Hauptvortrag gaben die eSciDoc Projektleiter Matthias Razum (FIZ Karlsruhe) und Malte Dreyer (Max Plack Digital Library, MPDL) in mehreren Konferenzbeiträgen einen einführenden Überblick über die Software, berichteten über die Entwicklungen im vergangen Jahr und leiteten am zweiten Tag Workshops für Entwicklerinnen und Entwickler, die eSciDoc in der Praxis einsetzen. Razum hob in seinem Überblick zur Infrastruktur besonders hervor, dass Sicherheitsfragen eine immer größere Rolle spielen. Deshalb müssen eScience-Systeme durch zuverlässige, ins System eingebettete Authentifizierungs- und Autorisierungswerkzeuge geschützt werden. FIZ Karlsruhe hat hierfür Shibboleth-Nutzerattribute und Kollaborator-Rollen neu in die eSciDoc Software integriert. Dreyer stellte den aktuellen Stand der eSciDoc-Anwendungen bei der MPG vor. Neben eSciDoc PubMan für durchgängiges Publikationsmanagement hat die MPDL Forschungsdatenmanagementsysteme für Bilddaten und Retrodigitalisate auf Basis der eSciDoc Software realisiert, die jetzt in mehreren Anwendungsszenarien unterschiedlicher Wissenschaftsgebiete zum Einsatz kommen. So wird zum Beispiel die Anwendung FACES, entwickelt zum Erkennen und Interpretieren von Fotos der menschlichen Gesichtsmimik, vom japanischen National Institut for Material Science (NIMS) für die Dokumentation und Bereitstellung von Kristallstrukturen und anderen Bildern aus der materialwissenschaftlichen Forschung eingesetzt. In diesem Zusammenhang wird FACES gerade zur neuen eSciDoc-Anwendung IMEJI weiterentwickelt, die den gesamten Workflow zur Verwaltung und Veröffentlichung von Bilddaten disziplinübergreifend berücksichtigt.

Verknüpfung von Forschungsdaten mit Forscherprofilen

Schon bei den eSciDoc Days 2009 hatte Mikiko Tanifuji von NIMS mit ihrem Vortrag über die Entwicklungsarbeiten zum Einsatz der eSciDoc-Software an ihrem Institut stark beeindruckt. Ihr Vortrag in Kopenhagen zeigte, dass NIMS das neue, webbasierte Publikations- und Forschungsdatenmanagement ganz konsequent weiterentwickelt hat: Alle Informationsquellen der materialwissenschaftlichen Forschung, ob Publikationen, Bilder, Grafiken, Forschungsdaten, persönliche Daten oder Informationen zu den einzelnen Wissenschaftlerinnen und Wissenschaftlern, werden Schritt

1 eScience (enhanced Science) bezeichnet eine durchgängige digitale Wissens- und Informationsinfrastruktur, durch die Forschungsarbeit schneller und besser werden kann. eResearch bezeichnet die neue Art der Forschung auf Basis digital verfügbarer Daten und Informationen, die überhaupt erst durch die Digitalisierung von Informationsquellen und ihre Verbindung mit spezifischen Softwareinstrumenten möglich wird

für Schritt zur „NIMS eSciDoc Digital Library“ zusammengeführt. Die werkstoffwissenschaftliche Bibliothek „NIMS Library Collection in Material Science“ ist auf Basis der eSciDoc Software PubMan (Publication Manager) bereits vollständig zu einem „NIMS Research Publication Archive“ umgestellt. Derzeit wird daran gearbeitet, die Bilddatenbank, das „NIMS Research Image Data Archive“ auf IMEJI umzustellen und mit dem Publikationsarchiv zu verknüpfen. Die Bilddatenbank enthält sowohl Open Access (OA) als auch anderweitig urheberrechtlich geschützte Inhalte. Zentrale Aspekte für NIMS sind eine zuverlässige Zugriffsverwaltung und das Rechtemanagement, welche mit Funktionen der eSciDoc-Infrastruktur realisiert werden können. Präsentationsfunktionen zur Erhöhung der Sichtbarkeit der NIMS-Forschung sind ebenfalls ein wichtiger Gesichtspunkt. Es gibt zum Beispiel eine Verbindung zu „ResearcherID.com“, einem Service, den Thomson Reuters der wissenschaftlichen Community anbietet. Der Identifier dient der Zuordnung von Informationen aus unterschiedlichen Datenbeständen zu einer Person: Die persönlichen Daten der Autoren zu ihren Publikationen in der eSciDoc Digital Library, von dort weiter zu den Forschungsdaten, die der Arbeit zugrunde liegen. Die japanische eScience-Library soll also nicht nur eine neuartige, forschungsfördernde Arbeitsumgebung für die Forschung am Institut werden, sondern auch die Forschungsleistungen des Instituts und seiner Wissenschaftler weltweit besser sichtbar machen. Wie das praktisch geht, zeigte Tanifuji in Kopenhagen live und wer möchte, kann es auf der eSciDoc-Homepage auch selbst nachvollziehen: Die japanische Wissenschaftlerin hat ihren Konferenzbeitrag dort unter der von NIMS adaptierten eSciDoc-Anwendung abgelegt.

In Japan hat das Nationale Institut für Materialwissenschaften (NIMS) beim Aufbau der eScience-Infrastruktur eine federführende Rolle übernommen. Projektleiterin Mikiko Tanifuji präsentierte ein hoch interessantes Konzept weltweiter Vernetzung von Wissenschaftsquellen und ist mit den darauf basierenden eigenen Entwicklungen schon sehr weit. © FIZ Karlsruhe

Japanische Radiologen der Forschungsorganisation NIRS sind von dem System der Materialwissenschaftler so angetan, dass sie eine Kooperation mit NIMS planen, um ihre Datenbestände ebenfalls in eSciDoc zu migrieren.

eSciDoc PubMan in mehreren Sprachen

Die größte spanische Universität, die Universidad Nacional de Educación (UNED), setzt eSciDoc-PubMan ein, um über ihre „Bibliotheksbestände zu informieren und die verfügbaren Publikationen zu verwalten“. Luis Zorita von der Bibliothek der UNED stellt in Kopenhagen die rhetorische Frage: „Warum haben wir uns für PubMan entschieden?“ Seine Antwort: Weil sich die Rolle der Bibliotheken mit der Entwicklung der bisherigen Bibliotheken hin zu ‚eResearch Libaries‘ ändere. UNED setze eSciDoc-PubMan ein, um neuen Anforderungen wie

- der dauerhaften Bereitstellung digitaler Publikationen,
- der Unterstützung und der Einbringung von Innovationen in den wissenschaftlichen Publikationszyklus,
- der Messung des internationalen Einflusses eigener Publikationen,
- der Zugriffsverwaltung in einer Open-Access-Umgebung und
- der Registrierung und Bereitstellung von Forschungsprimärdaten

begegnen zu können. Die spanischen Wissenschaftler haben für ihre eSciDoc-Anwendung eine spanische Oberfläche entwickelt. Das japanische NIMS-Archiv wurde sogar mit einer mehrsprachigen Oberfläche ausgestattet: zur Zeit japanisch, deutsch und englisch. NIMS hat vor, alle Publikationen im eigenen Repository mehrsprachig bereitzustellen, auch Aufsätze, die NIMS-Forscher in renommierten internationalen Fachjournalen publizieren. Diese sollen nach der Veröffentlichung original und (zurück) übersetzt(!) in die eScience-Bibliothek integriert werden. Über diese sind sie dann wieder mit den Experimentaldaten, Bildern zu den Forschungsarbeiten und den Lebensläufen der Forscherinnen und Forscher verknüpft.

Auch im Inneren imponiert die königliche Bibliothek durch ihre spannende Verbindung von Architektur und Funktion. © FIZ Karlsruhe

Clarin.dk baut eine Infrastruktur für die Geisteswissenschaften

Mit Clarin.dk, einem Projekt, das in enger Beziehung zum europäischen Entwicklungsprojekt CLARIN (Common Language Resources and Technology Infrastructure) steht, wurde damit begonnen, eine digitale Informationsinfrastruktur für die Geisteswissenschaften in Dänemark zu entwickeln. Dafür wird ein erweiterbares Repository für Texte, Audio- und Videodaten, Wörterbücher, Bilder und Daten ohne syntaktische Zuordnung gebraucht. Anders Conrad von der Royal Library Kopenhagen begründete die Entscheidung für die eSciDoc-Software: „Weil damit ein flexibles Modellieren von Content möglich ist“ und „ein ausgefeiltes Zugriffs- und Rechte-Management zur Verfügung steht“.

Schon zuvor in der Begrüßungsansprache hatte der Generaldirektor der gastgebenden Royal Library, Erland Kolding Nielsen, auf den steigenden Bedarf zur Abdeckung des kompletten Lebenszyklus von digitalen Objekten und der Schaffung von Möglichkeiten zur netzbasierten Kollaboration mit externen Partnern hingewiesen und zum Ausdruck gebracht, dass sich die Royal Library „auf die Teilnahme an der eSciDoc Community freut und hofft, neue Partnerschaften zu etablieren“. Mit den eSciDoc Days 2010 wurde das Fundament dafür gelegt.

Virtuelle und fernsteuerbare Labore mit automatischer Datenerfassung über eSciDCoc

Klaus Zimmermann vom Materialforschungszentrum der Universität Freiburg präsentierte in Kopenhagen die Planung einer eResearch-Infrastruktur, die deutlich über eine eScience Library hinausgeht. Im Projekt BW-eLabs entwickeln die Universität Stuttgart, die Universität Freiburg, die Hochschule der Medien in Stuttgart und FIZ Karlsruhe eine digitale Infrastruktur für virtuelle und fernsteuerbare Labore. Ein Entwicklungsblock von BW-eLabs beschäftigt sich mit der Steuerung des Fernzugriffs auf echte Laborgeräte und der Nutzung virtueller Laborsimulationen; der zweite mit dem

dazugehörigen Informationsmanagement zur Unterstützung des gesamten wissenschaftlichen Arbeitsablaufs, der Sicherung eines dauerhaften Zugriffs auf die im Forschungsprozess erarbeiteten Daten sowie dem Aufbau einer Infrastruktur für den Datenaustausch. In Zukunft sollen Forscherinnen und Forscher dadurch Laborexperimente auch vom Schreibtisch aus durchführen können. Das Informationsmanagement des eLabs auf Basis der eSciDoc-Software soll dabei alle zu einem Experiment gehörende Daten (Kalibrierungsdaten, Umgebungsbedingungen, eingesetzte Apparate, Simulations- und Berechnungssoftware, Messwerte etc.) weitgehend automatisch dokumentieren. Das Projekt wird vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert.

Wissenschaftlerinnen und Wissenschaftler aus verschiedenen Fachrichtungen diskutierten in Kopenhagen über die Entwicklung virtueller Arbeitsumgebungen für die Forschung.
© FIZ Karlsruhe

MPDL baut eResearch-Arbeitsumgebungen zügig auf

In weiteren Vorträgen stellten Vertreterinnen und Vertreter der MPG laufende und geplante Anwendungsentwicklungen in MPG-Instituten vor. Gisela Lausberg vom Max-Planck-Institut (MPI) für evolutionäre Anthropologie, Leipzig, berichtete, dass an einer Open Access Digital Library zur Beschreibung des Sprachgutes gearbeitet wird (Language Description Heritage). Diana Mewes vom MPI für chemische Ökologie, Jena, stellte den Einsatz von PubMan zum Management der Institutspublikationen mit eigenen fachspezifischen Klassifikations- und Dokumentationsformularen für die Verlinkung der Publikationslisteneinträge mit Autorendaten und weiteren Informationen vor. Dirk Wintergrün berichtete, dass das MPI für Wissenschaftsgeschichte die Nutzung natürlicher Sprachtechnologien sowie XML-Werkzeuge für eSciDoc entwickelt.

eSciDoc hat auch Aspekte zur langfristigen Bewahrung digitaler Inhalte im Blick

In einem abschließenden Extrakt fassten die Projektleiter Dreyer und Razum spannende Projekte zusammen, die vor kurzem angelaufen sind oder 2011 anlaufen und an denen die eSciDoc-Entwickler in größerem oder kleinerem Umfang beteiligt sind. Sie werden von der Europäischen Union, der Deutschen Forschungsgemeinschaft (DFG), Bundes- und Landesministerien gefördert oder von den Projektpartnern selbst finanziert. Neu arbeiten die eSciDoc-Partner mit den Zentralbibliotheken TIB Hannover und ZBMed zusammen, um beim Aufbau des Leibniz Open Access Repository zu unterstützen. Weiter haben sie folgende Projekte mit unterschiedlichen Entwicklungszielen in Angriff genommen:

- Digitization Lifecycle
- Astronomer's Workbench
- Jus CMS
- TEI Repository
- Lifecycle Logger
- SCAPE
- e-Kinematix
- BW-eSci(T)
- DILPS

Die genannten Projekte werden in dem eSciDoc-Vortrag „At a Glance: Further eSciDoc Projects" kurz vorgestellt. In die ersten fünf Projekte sind insgesamt 14 Max-Planck-Institute involviert; sie untersuchen den Digitalisierungs-Lebenszyklus (Digitization Lifecycle) eines Informationsobjekts von den Rohdaten bis zur Publikation und weiterführend über ihre gesamte digitale Existenz (Lifecycle Logger). Sie entwickeln Lösungen, in denen eine eResearch-Arbeitsumgebung für die astronomische Forschung (Astronomer's Workbench) entsteht und die sich mit der Integration von Content Management Systemen (CMS) in eSciDoc Pubman befassen. Die MPDL ist auch an dem DARIAH Teilprojekt TEI Repository beteiligt, das ein zentrales System für die Identifikation, Beschreibung und Verwaltung von Wissenschaftlern und Zitierungen werden soll. FIZ Karlsruhe ist am EU-geförderten Infrastrukturprojekt SCAPE (Scalable systems and services for preserving digital content) beteiligt, das mehr als zehn internationale Partner aus der Wirtschaft und der Wissenschaft in mehreren europäischen Ländern zur Entwicklung einer Infrastruktur für die nachhaltige Bewahrung von digitalen Inhalten verbindet, darunter auch Microsoft. Mit e-Kinematix fördert die DFG den Aufbau einer integrierten Informationsinfrastruktur für Maschinenbauer, die von den drei Projektpartnern RWTH Aachen, TU Ilmenau und FIZ Karlsruhe realisiert wird. BW-eSci(T) strebt mit Fördermitteln des Ministeriums für Wissenschaft, Forschung und Kunst Baden-Württemberg den Aufbau einer eResearch-Arbeitsumgebung am Beispiel der Computerlinguistik an. Diese soll so aufgesetzt werden, dass sie auf andere Fachdisziplinen übertragbar ist. FIZ Karlsruhe steuert hier Services und Werkzeuge für die Administration bei und unterstützt die Installation, die Programmierung und den Betrieb der eResearch Umgebung. Projektpartner ist die Universität Tübingen. DILPS/M (Distributed Image Library Processing System) ist ein DFG-gefördertes Projekt des Karlsruher Institut für Technologie (KIT) zusammen mit der Hochschule für Gestaltung Karlsruhe, das die Entwicklung einer Digitalen Bibliothek auf Basis einer virtuellen eResearch-Infrastruktur für die Verwaltung, Archivierung und Recherche von multimedialen Objekten zum Ziel hat. FIZ Karlsruhe unterstützt das Projekt als Kooperationspartner.

Betrachtet man diese hochspannenden Vorhaben so komprimiert, wird unübersehbar, dass die Wissenschaft in Deutschland, aber auch in vielen anderen Ländern die Bedeutung einer funktionsfähigen digitalen Informationsinfrastruktur und eines zuverlässigen Forschungs- und Forscherdatenmanagements erkannt und damit begonnen hat, sie auf breiter Ebene zu entwickeln. Bis zu den eSciDoc Days 2011 wird hier ganz sicher einiges passieren. Termin und Ort der Veranstaltung werden noch bekannt gegeben.

Die Folien zu den Vorträgen in Kopenhagen sind über die Seite www.escidoc.org verfügbar (Stichwort: eSciDoc Days).

Tagung, wissenschaftliches Arbeiten, Zusammenarbeit, Rechnernetz, Förderung. Projekt

Die Autorin

Vera Münch

Jahrgang 1958, ist freie Journalistin und PR-Beraterin mit Schwerpunkt Wissenschaft und Forschung. Seit vielen Jahren beschäftigt sie sich mit elektronischer Information und Kommunikation (Naturwissenschaften, Technik, Patente, Wirtschaftsinformationen) sowie Informatik und Software-Themen.
PR+TEXTE
Leinkampstr. 3 · 31141 Hildesheim
Telefon: (0 51 21) 8 26 13
Telefax: (0 51 21) 8 26 14
vera.muench@t-online.de

Aus der Praxis der Patentinformation

Die Entwicklung der elektronischen Medien und Dienstleistungen bei den Patentbehörden und Internetprovidern

Dieter Geiß, Castrop-Rauxel

Die 38. Arbeitssitzung der AGM (Arbeitsgruppe Elektronische Medien in der Patentinformation) beim Technischen Informationszentrum des Deutschen Patent- und Markenamt in Berlin am 4. November 2010 stand wie in jedem Jahr unter dem Motto „Fortschritte in der Patentinformation". Die AG hat rd. 60 Mitglieder aus Wirtschaft und Behörden. In diesem Jahr kamen 30 Fachleute nach Berlin, um sich mit Neuerungen und der zukünftigen Entwicklung der Patentinformation vertraut zu machen.
Lücken in den digitalen Datenbeständen der Ämter wurden geschlossen und jeder Provider setzt alles daran, seine Daten und Recherchetools noch benutzerfreundlicher zu gestalten.
Vor etwa 20 Jahren wurden die ersten Patentdokumente in digitaler Form auf CD gebrannt. Noch heute ist eine rasante Weiterentwicklung beim Aufbau von Patendatenbanken und Recherchemöglichkeiten feststellbar. Auch heute ist noch kein Ende dieser Entwicklung abzusehen. Der Informationssuchende muss sich daher notwendigerweise ständig weiterbilden, um sich mit den automatischen Übersetzungen aus andern Sprachen, dem „Cross File Searching" oder auch den semantischen Analysemethoden und Darstellungsweisen auszukennen.

An dieser Stelle ergeht , wie in jedem Jahr, der Appell an die Nutzer, nicht zu zögern und den Ämtern bzw. Providern mitzuteilen, wenn sie bei ihrer Arbeit Fehler oder Ungereimtheiten bei den Rechercheergebnissen oder Softwaretools feststellen sollten. Nur durch eine ständige Überwachung und Mitarbeit ist sicherzustellen, damit in der Welt der komplexen Datenstrukturen Fehler beseitigt und benutzerfreundliche Lösungen gefunden und erarbeitet werden können.

Patentbehörden

Europäisches Patentamt

Dienstleistungen des EPA

Die Patentinformationsdienste des Europäischen Patentamtes werden ständig an die Bedürfnisse der Nutzer und an die technische Entwicklung angepasst.
Durch die Einführung des neuen Expert Services sind für erfahrene Rechercheure verschiedene Dienste über einen gemeinsamen Einstiegspunkt erreichbar. So werden Synergien durch eine gemeinsame Verwaltung und ein einfacher, transparenter Zugang über eine einzige Plattform sichergestellt.
https://data.epo.org/expert-services/start.html
Der Expertendienst bietet die in Abbildung 1 aufgeführten Datenbanken an.

Im Bereich der PATSTAT Datenbank gibt es einen „Public mode", der mit verschiedenen vorgefertigten Statistiken für bestimmte Industrien (44) gratis genutzt werden kann. Die Parameter sind fest eingestellt und man kann keine eigenen Auswertungen vornehmen.
Der zweite Modus ist der „Private Mode". Dieser ist für Abonnenten vorgesehen, die Rechercheergebnisse als Eingangsdaten für eigene Statistiken verwenden. Sie können eigene Auswertungen vornehmen und die Möglichkeit nutzen, mit entsprechenden Parametern in weiteren Bereichen der Datenbank zu arbeiten.
Die Datenbank für statistische Auswertungen stellt als Werkzeug die „Cross Reference Chart" bereit. Diese eröffnet viele Kombinationsmöglichkeiten von Feldern wie IPC, Anmelder, Erfinder usw. zu Aussagen über den Wissenstransfer und bietet eine Zitatanalyse.

Das EPA hat mit der Firma Init AG einen Vertrag über die Bearbeitung der INPADOC-Daten geschlossen. Vorgesehen ist die Vervollständigung der bestehenden Sammlung mit den Daten der wichtigsten „hotspots" (Japan, Indien u.a.) mit zusätzlichen Rechtsständen aus EPÜ-Ländern. Hinzu kommt eine verbesserte Benutzerunterstützung auf der Webseite.

Das EPA hat auch in diesem Jahr seine Patentdatenbanken ergänzt. Zuletzt wurden vom Chinesischen Patentamt die Inhaberdaten von 1985 bis 1990 und von 1995 bis 2008 übernommen. Die geplanten Ergänzungen im Jahre 2011 umfassen Inhaberdaten von 1991 bis 1994 und den Frontfile ab 2009. Hinzugekommen ist der vorherige und neue Inhaber in englischer Sprache.
Vom Russischen Amt ist ab 2005 die vollständige Historie der Patenterteilungsverfahren verfügbar.
Vom Brasilianischen Patentamt stehen Daten über den „PCT Eintritt in die nationale Phase" bereit. Patentdaten aus Südamerika wurden ergänzt: LATIPAT ist jetzt voll in Betrieb mit einem starken Fokus auf Rechtsstandsdaten. Dies wurde durch eine enge Zusammenarbeit zwischen dem EPA, dem Österreichischen Amt und der WIPO erreicht.
Ergänzte Daten vom US-Patentamt bezogen sich auf die Beschreibung der US-Rechtsstandscodes und auf generelle Informationen zum US-Patentverfahren.
Im EPA ist vorgesehen, die XML-Konvertierung aller EP-A-Dokumente bis Ende 2010, die XML-Konvertierung der EP-B Dokumente bis Ende 2011 und das PDF Archiv der EP-B und EP-A Dokumente bis Ende 2010 fertig zu stellen.
Die PDF/A-Dokumente enthalten inzwischen Meta-Daten, die es Web-Suchmaschinen erlauben, bessere „snippets" zu erzeugen.

	Database name 1 ▲	Database ID	Year 2 ▼	Volume 3 ▼
	Bibliographic data and legal status data for	XBUL2010001	2010	001
	EPO granted patents	XABD2010002	2010	002
	EPO granted patents	XABD2010001	2010	001
🔒	European Patent Application	EPAF2010439	2010	439
	European Patent Application	XEPA2010439	2010	439
	European Patent Application	XEPA2010039	2010	039
	Global Patent Index	XGPI2010038	2010	038
	Granted European Patents	XEPB2010439	2010	439
	Granted European Patents	XEPB2010039	2010	039
	PATSTAT	PATSTAT	2010	000
	PATSTAT-201004	PATSTAT-201004	2010	000
	PATSTAT-PUBLIC	PATSTAT-PUBLIC	2010	000

Abbildung 1: Datenbanken im Expert Service des EPA. Die mit dem Schloss gekennzeichnete Datenbank ist nur zugänglich, wenn ein Nutzungsvertrag mit dem EPA abgeschlossen wurde.

Inzwischen hat das EPA dafür Sorge getragen, dass „smart search“ in der Datenbank EspaceNet voll implementiert ist und eine Hilfe zu „smart search“ zur Verfeinerung der Recherche bereitsteht. Die ursprüngliche Begrenzung auf vier Suchterms ist auf 20 je Suchfeld mit bis zu 20 Booleschen Operatoren und 20 Trunkierungen (mindestens zwei Zeichen je Ausdruck) erhöht worden. Eine Markierung der Suchterme ist möglich. Eine Sortierung der Treffer nach Ladedatum, Prioritätsdatum, Erfinder, Anmelder und ECLA ist möglich.
Volltexte der EP-A- und PCT-Dokumente in DE, EN und FR sind direkt verfügbar.
Die Übersetzungswerkzeuge in ESPACENET wurden verbessert, so dass jetzt sowohl von EN ins FR, DE, IT, ES und PT wie auch von jeder der oben genannten Sprachen eine Patentdokument ins Englische übersetzt werden kann.
Ein Tagging neuer Technologien mit Y-Klassen im ECLA Klassensystem zur Patentanalyse wird ermöglicht. Die XP-Dokumente sind mit der Quelle verlinkt und die Hitliste kann bis zu 500 Treffer anzeigen. Folgende Verbesserungen sind geplant:

- Die bessere Navigation in der Trefferliste (einfaches Umschalten zwischen den Seiten der Trefferliste),
- Übernahme der Dokumente in „my patents“-Liste mit Mausklick,
- Herunterladen der Titelseiten für „my patents“-Liste,
- RSS feed zu Suchanfrage mit Benachrichtigung, wenn Liste erweitert wird und der
- Export der vollständigen Trefferliste in den Formaten CSV und XLS.

Nach der Integration in die neue EPO-Webseite (Abbildung 2) wird es eine „quick help“-Leiste mit „tool Tipps“ und eine Neuauflage des „electronic assistant“ geben.

Das EPA hat Verbesserungen für den Dienst Register Plus vorgesehen:

- Fehlermeldeformular für Benutzer mit Fehlerverfolgung,
- Sortieren der Einspruchsunterlagen nach Ursprung (Inhaber und einsprechende Partei(en),
- verschiedene Anmeldenummernformate sind bei der Eingabe erlaubt (z. B. 00101010, 00101010.7, EP00101010, EP20000101010, EP2000101010),
- ein neues Icon zum Öffnen des Hilfeglossars und
- eine „quick help“ Leiste mit Antworten auf häufig gestellte Fragen.
- Sortieren der Trefferliste nach (für Trefferliste < 200 Treffer) Anmeldenummer, Veröffentlichungsnummer, Anmelder und IPC.

Nach der Integration in die EPA-Webseite gibt es weitere Verbesserungen:

- tool tips zu Eingabeformaten,
- quick help und die Rechercheoberfläche für Recherchen über Register Plus oder EspaceNet sind gleich.

Die Daten des Europäischen Patentamtes im Open Patent Services werden ST.36

Vorschau

Abbildung 2: EPA Vorschau auf die neue Webseite für die Patentrecherche in Espacenet.

entsprechen und mit einem eigenen DTD/Schema verteilt. Sie sind auf drei verschiedene Weisen erhältlich:

- als Datendownload über Register Plus GUI,
- als Datendownload über RESTful OPS Webservices und
- als „bulk download" der Register Datenbank.

Über zwei Dienste werden die Register Daten auffindbar sein (derzeit öffentlicher Beta-Test). Einmal über Register Daten zu Anmelde- oder Veröffentlichungsnummer und zum anderen als Register-Suche (Ende 2010 / Beginn 2011) mit der CQL Abfragesprache. Die INPADOC Familien Information (OPS family service) sind nicht enthalten. Die Anzeige des Akteninhaltes unter „all documents" (file inspection) ist auch nicht möglich. Verbesserungen sind aber für die Zukunft noch vorgesehen.

Übersicht über die Web-Dienste des EPA

EPA Publikationsserver:

https://data.epo.org/publication-server/doc/European_publication_server_webservice_doc_v1_1.pdf

Open Patent Services – OPS:

http://ops.epo.org

Register Plus:

http://ops-i.epo.org/2.6.2/rest-services/register/application/docdb/EP.00101010.*.20050101/biblio,events

http://ops-i.epo.org/2.6.2/rest-services/register/publication/docdb/EP.1500000.20000101

http://ops-i.epo.org/2.6.2/rest-services/published-data/search?q=applicant%3DIBM&Range=1-100.

Das EPA bietet für Expertenrecherchen weitere Hilfsmittel an. Die Benutzer-API für die MIMOSA-PIES-Oberfläche in der JTask Version 2 GTI V5 und V6 ist eine installierbare Version und soll jährlich aktualisiert werden. Es handelt sich um einen Browser-Versionsdienst für den PI-Experten.

Die Espacedienste werden wie bisher weiter gepflegt. Geplant ist, die A und B Dokumente des Amtes ab 1978 als Volltext im Internet bereitzustellen. In den Espaceserien wurden neue Felder eingeführt. So z. B. Felder über das Erstveröffentlichungsdatum auf DVD: FFD und die Veröffentlichungsdaten der DVDs: PUA - A1/A2, PUA3, PUB1, PUB2 und PUB3.

Bei BULLETIN gibt es nach wie vor die Browser Version wöchentlich mit der kompletten Datenbank. Die Bulletin DVD erscheint nur jährlich.

Eine mächtiges Rechercheinstrument ist weiterhin der Global Patent Index mit der Datenquelle = DOCDB XML. Erfasst werden die Publikationsdaten von mehr als 95 Patentämtern mit mehr als 73 Millionen bibliographische Datensätze. Er wird wöchentlich aktualisiert (montags 14:00) mit der Recherchesoftware MIMOSA PIES, spezifischen Suchmöglichkeiten (über 100 Suchfelder) zum weltweiten Stand der Technik mit der Möglichkeit zur Patentüberwachung sowie einem einfachen Zugang zu den Inhalten der Suchfelder.

Die recherchierbaren bibliographischen Daten sind verfügbar über die Felder: Klassifikationen (IPC, ECLA, ICO, nationale Klassif.), Entgegenhaltungen (Patentliteratur, NPL, Kategorien), Zusammenfassungen (Proximity Suche) und EPA einfache Patentfamilie.

Über den Open Patent Service sind zusätzlich die Felder verfügbar: repräsentative Zeichnung, die INPADOC Patentfamilien und die INPADOC-Rechtsstandsdaten. Die Dokumentenanzeige ist anpassbar und herunterladbar (XML) und enthält GPI- (DOCDB) und OPS-Daten.

Die Trefferliste ist ebenfalls zu bearbeiten und herunterladbar (CSV). Ein Hyperlink zum Volltextdokument über EspaceNet und eine Recherche nach Patent- und Nichtpatentzitaten runden die Recherchemöglichkeiten ab. Der GPI wird ständig ergänzt, Recherchen können gespeichert und bei Bedarf wieder geladen werden.

Die Nutzungsgebühr beträgt jährlich 710 Euro pro Nutzer mit ID-Kontrollen und Password.

DEUTSCHES PATENT- und MARKENAMT

Dienstleistungen des DPMA

Für das DPMA wurde ein neues Logo konzipiert. In Zukunft werden alle Dokumente oben links mit einem senkrechten Balken mit den Farben schwarz/ rot/ gold bedruckt. Das Patentblatt bekommt ein blaues Feld und das Markenblatt ein rotes Kennzeichnungsfeld.

DEPATISNET

Die Datenbank DEPATISnet wurde um die Anzeige vollständiger japanischer Dokumente ergänzt, und dies ab 1994 in PDF-Format. Suchbar sind japanische Schriftzeichen und englische Zusammenfassungen.

Die Trefferliste erlaubt nun ein Highlighting der Suchbegriffe.

Die Anzeige der recherchierbaren Volltexte aller DE und DD Dokumente 1877ff ist ab sofort möglich. Dabei kann die Anzeige der Treffer auf maximal 1000 Treffer erhöht werden.

Inzwischen ist auch eine Recherche nach Datumsbereichen möglich:

- M01-2009 Monat,
- KW12-2009 Kalenderwoche,
- Q3-2009 Quartal und
- J-2009 Jahr.

Für die Expertenrecherche wurden die Felder für die Klassifikationssymbole neu konzipiert:

- < IC = E03C1-04 >
- < IC = ‚E03C 1/04' >
- < IC= E03C0001040000 >

Bei der Einsteigerrecherche gab es verschiedene Änderungen. Das Anmeldedatum entfällt. Bei der Recherche nach Titel und Volltext werden eingebundene Sonderzeichen (Bindestrich, Schrägstrich, usw.) automatisch in den Nachbarschaftsoperator „W" umgewandelt. Damit folgt Suchbegriff 2 direkt auf Suchbegriff 1. Die bibliographische IPC, die reklassifizierte IPC und die Prüfstoff IPC wurden zu dem Sammelfeld „Alle IPC-Felder" zusammengefasst.

Bei der Veröffentlichungsnummer wurde ein Sammelfeld für Dokument ID (Länderkennzeichen, Nummer und Dokumentenart) geschaffen und ein Sammelfeld für Anmelder/Inhaber/Erfinder (PA+IN). Auch der Anmelder wird zu Anmelder/Inhaber (PA).

Die Hilfetexte sind entsprechend überarbeitet worden.

DPMA Publikationen

Die Datenabgabedienste des Amtes mussten nach den neuen Bestimmungen des Bundesdatenschutzgesetzes überarbeitet werden. Die Wiederaufnahme der Dienste erfolgte im Mai 2010. Die Anpassung bezog sich auf entsprechende Hinweise über die Datenweitergabe auf der Homepage des DPMA sowie die explizite Nennung aller Datenempfänger. Die Hinweise über Datenweitergabe sind inzwischen auf allen Anmeldeformularen aufgedruckt.

Auch die DEPATIS Connect Dienste mussten angepasst werden. Alle neun Alt-Verträge wurden zum Ende des Jahres 2010 gekündigt. Ein Neuabschluss kann zum 1. Januar 2011 erfolgen.

In den Jahren 2011 und 2012 soll DPMAconnect eine XML-Schnittstelle zu DPMAregister zur kostenfreien und publikationsunabhängigen Abfrage von Schutzrechtsdaten erhalten, wobei auch eine Überleitung von DEPATISconnect in DPMAconnect erfolgen soll.

Die Zusammenführung von DPINFO und DPMApublikationen zum DPMA-Register ist weiter fortgeschritten.

Die erste Projektphase mit dem Datenbestand Marken wurde abgeschlossen. Der Betrieb war am 28. April 2009 aufgenommen worden. Dies gilt auch für die zweite Projektphase mit dem Datenbestand Geschmacksmuster. Der Betrieb begann am 22. Februar 2010. Die dritte Projektphase mit dem Datenbestand Patente/Gbm

kann erst 2011 abgeschlossen werden. Der Start des Dienstes für Patente und Gebrauchsmuster musste auf Februar 2011 verschoben werden, solange ist DPINFO und DPMApublikationen noch für Patente und Gebrauchsmuster in Betrieb.

DPMA Connect

Die Inbetriebnahme der DPMAconnect web service Schnittstelle ist für das 1. Quartal 2011 geplant. Sie wird XML-basiert sein und Aktenzeichenabfragen oder komplexe Suchfragen erlauben. Die Übermittlung von kompletten Registerauszügen erfolgt in XML-Format. Dazu ist ein Nutzungsvertrag erforderlich mit Nutzerkennung und Passwort. Die Nutzung ist weiterhin unentgeltlich. Eine Freischaltung sollte im zweiten Quartal 2011 für alle Schutzrechtsarten möglich werden.
Die Webservice-Schnittstelle DPMAconnect wird die Nutzung durch kommerzielle Informationsdienste und firmeninterne Informationssysteme ermöglichen. Ferner ist eine Anbindung von TMview (HABM), das föderierte europäische Patentregister (EPA) und PatentScope der WIPO eingeplant.

Elektronische Akteneinsicht

Die elektronische Akteneinsicht hat im DPMA noch einige Hürden zu nehmen. Basis ist die „Elektronische Schutzrechtsakte“ (Elsa) im DPMA. Die Bereitstellung der elektronischen Akteneinsicht für Patente/Gebrauchsmuster erfolgt voraussichtlich erst zu Beginn des Jahres 2012, die Marken voraussichtlich erst 2013 und die Geschmacksmuster erst in den Jahren 2013/2014. Der Aufruf der Akte soll aus der Trefferliste des DPMAregisters möglich werden.

World Intellectual Property Organisation

Die WIPO hat ein neues Portal für alle WIPO Informationsdienstleistungen eingerichtet (Abbildung 3).

PatentScope

Die traditionelle PatentScope Benutzeroberfläche für die Recherche nach PCT-Dokumenten mit den Möglichkeiten, nach einer Volltextsuche in PCT-Anmeldungen über Beschreibung und/oder Ansprüche, unbegrenzte Stichwörter und farbig markierte Schlüsselwörter zu Suchen, wurde erweitert mit Informationen über den Eintritt der PXCT-Anmeldung in die nationale Phase. Dazu erhielt PatentScope eine neue Benutzeroberfläche.

WIPO GOLD

Abbildung 3: Portal für WIPO-IP-Informationsdienstleistungen.

Ab sofort kann nicht nur nach PCT-Dokumenten, sondern auch nach nationalen Dokumenten gesucht werden.
Eine „Simple Search“ Benutzeroberfläche wurde eingeführt.
Die neue Suchmaschine (open source) hat folgende neue Module:
Eine flexiblere Suchsprache (Lucene), eine Recherche nach dem Wortstamm (electric, electric-al, electric-ity, ...) mit stapelbarer Trunkierung, eine Unterstützung der Recherche in mehreren Sprachen, eine automatische statistische Analyse der Suchergebnisse, ein besseres Ranking nach Relevanz, soweit nötig eine Maschinenübersetzung von Titeln und Zusammenfassungen und auch Volltexte mit Beschreibungen sowie Ansprüchen und eingebetteten Zeichnungen.
Weitere PCT-Dokumente und nationale Patentdokumente von Argentinien, Brasilien, Israel, Kuba, Korea, Marokko, Mexiko, Singapur, Spanien, Südafrika und Vietnam wurden in die Datenbank aufgenommen. Es folgen noch die Daten der Philippinen und von Kenia.
Eine tabellarische Darstellung der Suchergebnisse wird ermöglicht und eine Direktanalyse von tausenden von Suchergebnissen nach bestimmten Kategorien, so z. B. nach Patentamt, IPC Klassen, Anmeldern, Erfindern und nach dem Anmeldedatum.
Eine Übersetzung mit Hilfe von Google kann genutzt werden. Möglich ist inzwischen auch eine nutzerdefinierbare Proximitätssuche mit dem Tilde Operator “~”.
Das Cross-lingual-search-tool (CLIR) bietet dem Rechercheur folgende Möglichkeiten:
Ein Verwendung von Suchbegriffe/Stichworte in EN, DE, ES, FR oder JP, mit einer automatischen Synonymerweiterung in der ursprünglich eingegebenen Suchsprache und eine automatische Übersetzung der Stichwörter/Schlüsselwörter in andere Sprachen. Hinzu kommt ein automatischer oder „supervised“ Modus, ein „Sliding bar“ (precision vs. recall), eine Begrenzung/Eingrenzung der Suche auf die vom Nutzer gewählten Technologiefelder und passenden Synonyme. Aufgebaut wird diese Recherchemöglichkeit auf zweisprachigen Wörterbüchern, die mit statistischen Methoden aus Patentdokumenten zusammengestellt wurden. (Siehe auch www.wipo.int/patentscope/search/en/search.jsf)

IPC Green Inventory

Das „Green Inventory“ ist ein IPC-Klassenverzeichnis auf der Grundlage der UNFCCC, („United Nations Framework Convention on Climate Change“) der Rahmenübereinkunft der Vereinten Nationen über den Klimawandel (s. auch http://unfccc.int/ttclear/jsp/Technology.jsp).
2008 wurde von GB ein Vorschlag an die IPC-Arbeitsgruppe der WIPO herangetragen, ein IPC-Indexing-Schema auf Grundlage von UNFCCC zu erstellen, um den Zugang zu EST („environmentally sound technologies“) zu verbessern, das den Technologietransfer, die statistischen Analysetools, das Monitoring, die Identifizierung von Technologietrends und die Erstellung von „Patent landscapes“ beinhalten sollte. Die Mehrzahl der Patentämter hat dem Vorschlag nicht zugestimmt, weil es keine anerkannte Definition für den Begriff „umweltfreundlich“ gibt.
Wenn nötig, sollten reguläre IPC-Klassensymbole geschaffen werden.
Mit Hilfe der WIPO sollte aber das neue „IPC Green Inventory Tool“ mit einem zusätzlichen IPC-Stichwortverzeichnis im Januar 2011 verfügbar sein. Das „IPC Green Inventory“ ist in Stand-alone-Form über die IPC-Homepage www.wipo.int/classifications/ipc/en/est zugänglich.

Internetprovider für Patentinformationen

LexisNexis

Die Total Patent Datenbank von Lexis Nexis umfasst die Patentdaten von 100 Ämtern und 27 Volltext-Sammlungen von wichtigen Ländern wie US, EP, WO, DE, FR, GB und JP (Patent Abstracts) einschließlich der Daten von Russland, China und Brasilien. Gespeichert sind die Dokumente in der Originalsprache mit einer Maschinenübersetzung ins Englische. Die Erfassung und Einbindung weiterer Länder ist geplant.
In den Jahren 2009 und 2010 sind die BRIC Länder (BR, IN, RU CN) hinzugekommen.
2010 wurden die Patentdaten für verschiedene Länder wie Japan, Korea und China erheblich ergänzt. Es konnten auch Daten aus Argentinien bereitgestellt werden. Mit den VISTA-Staaten wird noch mit dem Ziel verhandelt, dass auch diese Daten erfasst werden.
Bei den Maschinenübersetzungen aus dem Japanischen wurde besonders auf eine hohe lexikalische Genauigkeit geachtet, um sicherzustellen, dass eine Recherche auch erfolgreich ist. Es wird großer Wert auf die korrekte Lesbarkeit der Volltexte (Beschreibung und Ansprüche) gelegt.
Ziel einer semantischen Recherche ist es, eine Möglichkeit zu schaffen, um noch vollständigere und präzisere Resultate zu generieren.
Die semantische Recherche wird als die Recherchetechnologie der nächsten Generation angesehen. Mit der semantischen Suchmaschine in Total Patent soll dies erreicht werden.
Was dürfen Rechercheure bei der semantischen Recherche erwarten:
Ein Verständnis der Rechercheinhalte bis ins Detail. Die Fähigkeit zu erwerben, eine Überprüfung vornehmen zu können über das, was zu löschen ist, was bei der Zusammenstellung recherchelevanter Begriffen (concepts) erweitert oder eingeschränkt werden muss und die Fähigkeit zu entwickeln, die Skalierbarkeit und Lernkapazität der Recherchetermes (brains), die in der Datenbank Total Patent wie auch in anderen freien Patentdatenbanken genutzt werden können, gezielt einzusetzen.
Verknüpfungen zu Nichtpatentliteratur werden über Elsevier's Scopus® Preview bereitgestellt. Sie ermöglichen dem Nutzer, relevante Zitate und Abstracts aus anderen Zeitschriften zu finden und in die Recherche zu integrieren.
In Total Patents werden auch die Patentinhaberdaten standardisiert, so dass Rechercheure leicht erkennen, welche Patentinhaber Gruppen zusammengehören.
Mit dem Tool BizInt Smart Charts können Datensätze von internen Ordnern weiter verwendet, analysiert und als Total Patent Daten präsentiert werden, um die Brauchbarkeit und Flexibilität des Recherchetools nachzuweisen.
Das Recherchetool PatentOptimizer Analytics erlaubt eine detaillierte und rasche Analyse des Patentportfolios mit Hilfe von Anspruchselementen und Begriffen, die Concepts (Begriffsbündelungen) beschreiben.
Darüber hinaus hilft der Optimizer, eine konsistente, präzise und gut durchdachte Patentanmeldung zu planen, sie zu analysieren und für eine Online-Anmeldung vorzubereiten. Details sind bei marcus.stewart@lexisnexis.com zu erfragen.

Minesoft

PatBase

Minesoft Ltd. hat neue Funktionen und Verbesserungen im Design von PatBase vorgenommen. Im März und November kommt jeweils ein neues Release auf den Markt. PatBase umfasst die Familiendaten und den Volltext von Dokumenten aus 13 Staaten, so z. B. auch Daten aus Taiwan, China und Korea.
Dazu kommen Klassensymbole von Klassifikationsrevisionen der Ämter US, EP und JP, Zitate, spezielle Indexe, Korrekturen von Familiendaten und Standards. Jeder einzelne Datensatz ist verknüpft über die Patentfamilie. Das Design wurde verbessert. Es gibt jetzt ein klares und deutliches Layout, ein leichteres Navigieren, ein schnelles Interface, neue Help Files, Guides und Training-Videos und bessere „Offline" Folder- und Alertreports.
Der Rechercheur kann bei bestimmten Recherchevorgängen Suchbegriffe in Ergebnislisten durch selbst gewählte Farben hervorheben.
Zur Optimierung der Recherche werden vom System selbst verwaltete Technologiebereiche oder andere erwünschte Zieldaten vorgegeben, die mit in die Recherche einfließen.
Neue Funktionen unterstützen ein erweitertes Highlighting für alle Operatoren. Es können auch Trunkierungen und Abstandsoperatoren verwendet werden und dies über alle Ansichten. Neue Hit Maps zeigen genau die Suchworte im Dokument an. Mit einem separaten Feld „ex=term" kann eine Suche nach Beispielen im Volltext erfolgen. Ein Multi Window Support mit vielen Fenstern steht bei Bedarf zur Verfügung.
Verbesserte Alerts stehen ebenfalls bereit, so auch RSS Feeds. Mehrere Update codes können in einem Alert zusammengefasst werden. Eine Linkstrunkierung ist möglich. Ein neuer Alert Viewer wurde erstellt. Die Empfänger des Alerts können im Ergebnis suchen und die Daten an Kollegen weitergeben.
Ein verbessertes Layout wurde eingeführt, dazu „Analysis Snapshot" und die „Export Alerts". Ergebnisse können direkt in ein Excel und PDF Format umgewandelt werden. Die Software und die Inhalte des Systems werden für die professionelle Recherche ständig weiterentwickelt.
Weitere Details sind bei jl@minesoft.com zu erfragen.

Thomson Scientific

Thomson Scientific stellt eine Vielzahl von IP-Diensten und Werkzeuge für Experten bereit. Thomson Innovation ist eine Datenbank, die Patentdaten und Nichtpatentliteratur zusammenfasst und recherchierbar macht.
Der WPI von Derwent umfasst derzeit rd. 20 Millionen Patentfamilien, wird aus 44 Quellen gespeist und steht für Recherchen und die Volltextversorgung bereit.
Der Derwent Citation Index wurde neu aufgebaut.
Es gibt erheblich wachsende Probleme mit der Dokumentenflut. TS fordert daher eine verbesserte Zusammenarbeit aller Beteiligten.
TS möchte mit seinem IP-Manager den Kunden von der Idee bis zur technologischen Umsetzung dieser Idee betreuen. Dazu stehen eine Menge spezieller Werkzeuge bereit u.a. auch das „Work File Sharing".
Durch die globale Vernetzung kommt es immer häufiger zu Patentverletzungen und Prozessen. Auch auf diesem Gebiet hilft TS. Ein Rechtsstreit über eine Patentverletzung kann in den USA Kosten von bis zu fünf Millionen US$ verursachen. TS stellt hierfür die Westlaw Patents Datenbank bereit, um im Vorfeld schon analysieren zu können, wie die Ansprüche aufgebaut sind und was sich technologisch verändert hat. TS gibt Hilfen zu der Frage, wie man Patentstreitprozessen aus dem Wege geht.
Ein weiteres Feld der Betätigung von TS ist die Wertschöpfung aus Patentportfolios. TS hält Tools für die Bewertung vor und erkennt wertlose Schutzrechte. Mit einer strategischen Aufbereitung des Portfolios lassen sich Stärken und Schwächen der Schutzrechte erkennen.
Die Geschwindigkeit der technologischen Entwicklung und der Innovationen hat zugenommen und hat erhebliche Auswirkungen auf das Patentwesen. Thomson Innovation hilft dem Kunden durch „Cross search collections" bei der Auslese und Weiterverfolgung von der Idee bis zur Anmeldung.
2010 erfolgte alle drei Monate ein Update. Es wurden eigene Templates für Alerts und Recherchen entwickelt. Inzwischen sind auch chemische Strukturen in

der Datenbank T Innovations verfügbar. Analysen können mit bis zu 60.000 Dokumenten durchgeführt werden. Dazu kommen eigene Kundenfelder.
Im DWPI konnte die Vollständigkeit der gespeicherten Daten weiter verbessert werden. 1.336.000 Datensätze mit Neuanmeldungen gab es 2010 bisher weltweit.
Bei TS wird das Qualitätsmanagement groß geschrieben. Die Rohdaten durchlaufen verschiedene Prozesse, ehe sie in die Datenbank aufgenommen werden. Auch für die Dokumente aus China mit einem manuellen Übersetzungsdienst wird eine höchstmögliche Datenqualität zugesichert. Der DWPI wurde um Indices mit grünen Technologien erweitert.
Weitere Informationen sind bei peter.fischer@thomsonreuters.com zu erfragen.

Questel Orbit

Die UNYCOM Patentdatenbank mit dem Patent Pilot wurde von Questel übernommen. Bis zum Jahresende2010 sollen alle LexisNexis Daten durch Questel Daten ersetzt werden. Es erfolgt eine Anbindung an das Questel PDF Archiv. In den Jahren 2011 bis 2012 erfolgt auch die Integration von PatentPilot Funktionalitäten in Orbit.
Über das Portal Orbit.com sind Patent- und Designrecherchen möglich.
Eine optimale Nutzung des neuen Portals wird in Aussicht gestellt mit besten Suchmöglichkeiten, Anzeigen, Auswahlvarianten, Verteilung, Analyse und Bewertung von Patentinformationen, um sicher, schnell und umfassend die entscheidenden Informationen zu finden. Die Recherche wird unterstützt durch Suchassistenten und Verlinkungen.
Die Dienstleistungen sind einzigartig. Recherchen, Patentverletzungen, Unterstützung bei Anmeldevorhaben und EP Auswertungen werden angeboten. Mit „key contents“ und semantischen Tools sind für einfache und komplexe Recherchen alle Voraussetzungen gegeben. Der „Search xpress search“ ist gedacht für unerfahrene Nutzer. Rechercheprofile sind in die entsprechenden Felder einzutragen und können auch in die Command Line integriert werden mit einem optimalen Management von Alerts und Ergebnissen der Recherche. Noch nach zwölf Monaten kann man seine Recherchen wieder aufrufen. Eine „Enduser-freundliche Ansicht“ der Ergebnisse wurde eingerichtet, die nur auf den technischen Inhalten basiert ohne die „patentrechtlichen“ Informationen. Möglich ist auch die Markierung von Suchbegriffen unabhängig von den Suchwörtern und eine Speicherung von „Highlighting Profilen“.
Die Suche nach nationalen Klassen, ECLA, ICO, F-Terms, FI Codes, auch in Kombination untereinander, ist möglich. Dabei unterstützt der „Classification Assistant“ den Rechercheur mit der Übernahme von wichtigen relevanten Klassen in die Recherche.
Firmenbezeichnungen werden in so genannten „Corporate Trees“ dargestellt und ermöglichen eine gute Übersicht über das Zusammenwirken von Firmengruppen und Konzernen. Angeboten wird auch die Suche nach Patentanwälten mit Arbeitsgebiet und Anzahl der Anmeldungen in den verschiedenen Ländern.
Über jeden beliebigen Datensatz werden in der **Datenbank FAMPAT** alle Publikationsstufen einer FAMPAT-Patentfamilie ermittelt. Angezeigt werden bei Bedarf alle bibliografischen Daten, die Abstracts in der Originalsprache und in Englisch mit nationalen und internationalen Klassensymbolen, mit Maschinenübersetzungen für Abstracts, Titel und für Volltexte für 22 Länder. Die Maschinenübersetzung von ermittelten Volltexten ist ebenfalls möglich. Hinzu kommen Zitate und Rechtsstand der gefundenen Dokumente.
Questel extrahiert anhand von statistisch linguistischen Auswertungen den „**Key Content“** und **die Questel „Concepts“,** die dann ebenfalls in einem FAMPAT-Datensatz enthalten sind.
Für den „**Key Content“** werden die Inhalte englischsprachiger Volltexte analysiert. Der Key Content extrahiert und strukturiert den Gegenstand der Erfindung neu, lässt Vorteile erkennen, Rückschlüsse ziehen und mögliche neue unabhängige Ansprüche formulieren.
Damit sind die Vorteile eines Key Contents erkennbar. Er gibt einen direkten und schnellen Überblick über die Erfindung, die Vorteile, die Nachteile, den Stand der Technik und die unabhängigen Ansprüche. Eine Suche in den Teilfeldern des Key Content ist möglich, aber auch eine Suche „quer“ über die wesentlichen Inhalte.
Die „**Concepts**“ werden durch statistisch linguistische Methoden erstellt, wie Häufigkeit des Vorkommens der Wörter, der Abstand von Wörtern zueinander, in welchen Abschnitten eines Patents die Wörter vorkommen usw. Die Gewichtung erfolgt auf der Basis der Häufigkeit der Begriffe im Dokument und in welchen Feldern sie vorkommen und dies mit der Frequenz von „multiword concepts“ mit höchstens 100 Wörtern.
Die Darstellung eines Begriffs in einem gefundenen Dokument bzw. Abstract ist farbig und mit größerer Schrift je nach Bedeutung und Gewichtung des Begriffs hervorgehoben. Damit ist der Inhalt des Treffers auf den ersten Blick sofort erkennbar.
Möglich ist auch eine Festlegung und Recherche mit „Top Concepts“.
Die Trefferliste wird nach Relevanz sortiert, wobei diese Relevanzsortierung ebenfalls auf statistisch linguistischen Methoden, Concepts, Key Content, Klassen und Zitaten basiert.
Nach einer Auswahl von Dokumenten einer manuell erstellten Liste, einer Recherche oder eines Alertergebnisses können über die Hauptklassen, die Concepts aber auch über das „IP Business Intelligence Modul“ statistische Auswertungen erfolgen.
Je nach Bedarf kann die Analyse Ergebnisse auf vielen Gebieten liefern, so z. B. über die Anzahl der Anmeldungen pro Jahr, über die Anmeldungen pro Land, über die Anmelder, über die Zusammenarbeit von Erfindern und Anmeldern auf diesem Gebiet der Technik, über die IPC-Klassen oder die semantische Analyse und über die Häufigkeit der Verwendung gleich lautender Begriffen bei einzelnen Anmeldern. Die Diagramme sind dynamisch und können angepasst werden.
Die Sammlung, Verteilung und Kommunikation von Patentinformationen im Unternehmen kann individuell gestaltet werden. Die Verteilung von Rechercheergebnissen ist auf viele Arten möglich, so mit Hilfe von einfachen Listen, Workfiles nach Selektion der Dokumente und Auswahl des Zielordners und dies mit vielfältigen Exportmöglichkeiten durch Generierung von Listen mit Permanent-Links, PDF-Sammlungen zum Import in interne Datenbanken oder mit Hilfe andere Programme wie Biznet oder VantagePoint. Dabei werden die Exporte gespeichert und können jederzeit wieder aufgerufen werden. Leserechte für ausgewählte Workfiles können auch an Nutzer vergeben werden, die keinen ORBIT-Zugang haben.
Über ORBIT können relevante Informationen mit anderen „geteilt“ werden. Es kann der Aufbau eines „eigenen“ Archivs mit relevanten, bewerteten Dokumenten erfolgen.
Für weitere Informationen und Anfragen zu einem Testzugang:
aschmitt-bender@questel.com
oder Telefon +49 8141 3154208.

Dialog

Dialog hat neue Patentinformationsdienste und Werkzeuge im Angebot. Im August 2010 wurde das neue Internet Portal **ProQuest Dialog** frei geschaltet. Dialog wird sich auf folgende drei Schwerpunkte konzentrieren und diese auch erweitern und verbessern:

1. Discover
   mit der Bereitstellung von naturwissenschaftlicher, technischer und medizinischer Literatur,
2. Validate
   mit dem Inhalt des gewerblichen Rechtschutzes und
3. Market
   mit der Bereitstellung von Rechts- und Marktdaten und Analyseprodukten.

There are easier ways of finding what you are looking for!

At PatBase we can help!

- PatBase is created by professional patent searchers for professional patent searchers
- Over 15,000 users worldwide
- Used by a range of IP professionals including: corporate IP departments, patent attorneys, information specialists, patent analysts, licensing, R&D
- Search across a comprehensive coverage:
  38.5m patent families
  25m Full-Text records
  8.7m Asian Full-Text records
  With translations

Call us to try PatBase free for 14 days!

minesoft

contact minesoft:
+44 (0)2 8404 0651 (UK)
+49 (0) 211 15924489 (Germany)
(401) 737 6499 (USA)
+81 3 54039191 (Japan)

info@minesoft.com
www.minesoft.com

PatBase

PatBase has been developed in partenership with RWS GROUP

Dialog bietet gewisse Vorteile mit einer User- und Experten-Schnittstelle an und verbindet den Inhalt der Dialogdatenbanken mit denen von **ProQuest** mit entsprechenden Schlüsselfunktionen.
Diese Schlüsselfunktionen umfassen eine ausgefeilte Rechercheschnittstelle für einfache und schwierige Recherchen, eine einfache Führung durch die Datenbanken und die Auswahl von Seiten, Vorschläge für Suchterme, Textauswertung und Relevanzanalyse, Linkstrunkierungen, Volltexte, Direktübersetzungen, Erkennung von Druckfehlern, graphische Auswertewerkzeuge, Rankingfilter, flexible Export- und Speicherfunktionen mit Werkzeugen für die Zusammenarbeit von Experten.
Recherchemöglichkeiten in der Pharmazie und Biomedizin werden ebenfalls angeboten.

Die Dialog Datenbank 325 umfasst 2,5 Millionen Patente und 1,6 Millionen Gebrauchsmuster des Chinesischen Patentamtes (SIPO). Die Daten sind von 1985 bis heute verfügbar, werden wöchentlich aktualisiert, und zwar etwa zwei Wochen nach der Veröffentlichung durch SIPO. Die Patentinformationen enthalten in intellektueller Übersetzung ins Englische den Titel, das Abstract und detaillierte Beschreibungen, Ansprüche, Prioritätsinformationen mit den laufenden Rechtsstandsinformationen. Eine einzelne Seite kann online aufgerufen werden. Links zur originalchinesischen A- und B-Volltextdokumenten im PDF-Format sind vorhanden. Im Vergleich zum DWPI-File 351 und INPADOC-File 345 bietet die Datenbank 325 bis auf den Familiendienst wesentlich mehr Informationen.
Über **SciSearch** von **ProQuest** können weiter innovative Daten aus China abgerufen werden.

Innography ist ein weiteres Tool von Dialog für die professionelle und analytische Recherchearbeit im Bereich des gewerblichen Rechtsschutzes. Sie fasst Patent-, Marken-, Geschäfts- und Patentverletzungsdaten zusammen.
Dieses Tool hilft nicht nur beim Aufbau und Schutz des Patentportfolios, sondern unterstützt den Aufbau neuer Technologiefelder und Innovationsgebiete eines Unternehmens, erstellt Trendanalysen und CI-Analysen, hilft bei der Auffindung von Kandidaten für Lizenzierungen und Partnern für M&A (deal due diligence) mit der erforderlichen Sorgfalt, unterstützt bei Rechtstreitigkeiten und bei der Abwehr und Vermeidung von Schutzrechtsverletzungen.
Für Patentverletzungen gibt es ebenfalls eine Datenbank, die Ergebnisse zu Unternehmen liefert, die in besonders viele Verletzungsstreitfalle verwickelt sind und erkennbar machen, ob es sich bei diesem Unternehmen eventuell um einen Patenttroll handelt. Dies ist wichtig für die Entscheidung, ob man das Risiko eingeht mit dieser Firma zusammen zu arbeiten.
Innography wird ständig ergänzt, so z. B. durch Datenbanken zu US-Marken (Proprietary SemaMap) und Daten von europäischen Finanzmärkten. Mit großer Sorgfalt wird die Datenbank @UltimateParent über weltweite Unternehmensdaten gepflegt. Über die Eingabe des Namens eines Unternehmens wird von diesem Unternehmen ein Übersichtsreport im Word-Format erstellt.
Algorithmen wurden verbessert, die eine Aussage über die Patentstärke und -bedeutung zulassen mit zusätzlichen Bewertungskriterien wie Patentlaufzeit, Verletzungshandlungen und Einsprüchen. Dazu kommen Datenbanken mit Quellen der Rechtsprechung, mit Analysewerkzeugen und neuen Visualisierungsmodulen. Aus Deutschland sind derzeit 226019, aus den USA 64311 und vom Europäischen Patentamt 80688 Fälle von höchsten Gerichten erfasst und recherchierbar.
Eine detaillierte und integrierte Online-Dokumentation wird angeboten.
Im Markenbereich stellt Innography verschiedene Werkzeuge bereit, um die Bedeutung von Marken im Wettbewerb darzustellen und zu analysieren. Ein Werkzeug steht bereit, das die Möglichkeit eröffnet, Marken zu finden, die für ein Patent eingetragen wurden, um auch hier die Bedeutung des Patents am Markt beurteilen zu können, mögliche Lizenznehmer zu finden oder Schutzrechtsverletzungen ausfindig zu machen.
Der Datensatz einer Marke beinhaltet die Anmeldenummer, den Namen des Inhabers und die Klassensymbole der internationalen Markenklassifikation. Innography ergänzt Symbole der IPC.

Bei den vorgesehenen erweiterten Referenzen geht es um die Einbeziehung

- von Ergebnissen von Verletzungsklagen weltweit vor Gerichtshöfen in Australien, Kanada, den USA, in Frankreich, Deutschland, England, China und Japan. Mit diesen Daten können kleine Datenbanken angelegt werden, die bei der Ergebnisfindung hilfreich sein können.
- von technischen Normen, die Bezüge zu den Patenten haben, wie z. B. die ETSI (European Telecommunications Standards Institute), die ISO Normen (International Organization for Standards), die ANSI Normen (American National Standards Institute). So auch die Normen IEC (International Electrotechnical Commission), IETF (Internet Engineering Task Force), IEEE (Institute of Electrical and Electronics Engineers), W3C (World Wide Web Consortium) und der OMA (Open Mobile Alliance).
- von pharmazeutischen Informationen das „Orange Book“, das „Green Book“ (Veterinary drugs) und die kanadische Wirkstoff-Liste.
- das Modul „research seats“ erlaubt es dem Nutzer festzustellen, wer Zugang zu den Patentrecherchen und Analysen hatte.
- das Modul „Label visualizations“ verknüpf die gefundene Klassifikationssymbole der IPC oder der US Codes mit den individuellen Klassifikationsterms des Nutzers und erweitert damit die Indexierung für den Nutzer auf eine einzigartige Weise.

Weitere Details sind bei r.wonner@dialog.com zu erfragen.

Tagung, Patentinformation, Host, Patentamt

Der Autor

BD Dipl.-Ing. Dieter Geiß

war seit 1969 Leiter des Pateninformationszentrums der Universitätsbibliothek Dortmund und Lehrbeauftragter an der Universität Dortmund für Patentwesen und Dokumentation. Von 1995 bis 2003 führte er den Vorsitz der APD Arbeitsgruppe „Elektronische Medien in der Patentinformation“ (AGM), jetzt Schriftführer der AGM und Berichterstatter zu Patenfragen der DGI.

Mulvanystraße 32
44575 Castrop-Rauxel
d.geiss@t-online.de

Qualifiziert in die Zukunft

Innovative, maßgeschneiderte Weiterbildung: Der Zertifizierungslehrgang „Technische Redakteurin/Technischer Redakteur (tekom)“ von Comet startet im Februar in München

Das Schulungszentrum der Comet Unternehmensgruppe bietet jetzt eine Karrierechance für alle, die sich bereits mit Technischer Dokumentation beschäftigen oder in diese spannende Branche einsteigen wollen: Am 18. Februar 2011 beginnt ein neuer berufsbegleitender Weiterbildungslehrgang, der gezielt auf die Zertifizierung „Technische Redakteurin/Technischer Redakteur (tekom)“ vorbereitet.

An wen richtet sich der Lehrgang?

Der dreimonatige Bildungsgang eignet sich hervorragend für Ein- und Umsteiger/-innen aus Ingenieur-, IT- und anderen Technikberufen, richtet sich aber genauso an berufserfahrene Redakteure, die Technische Dokumentation konzipieren und erstellen.

Wie wird der Lehrgang durchgeführt?

Eine innovative zeitgemäße Unterrichtsgestaltung zeichnet das Lehrgangsangebot aus. Mit einer Kombination aus acht kompakten Präsenzveranstaltungen am Wochenende und E-Learning-Modulen bietet der Lehrgang größte Flexibilität, neben dem Berufsalltag absolviert zu werden. Kleine Gruppen und ein erfahrenes Ausbildungsteam gewährleisten eine optimale, individuelle Betreuung während des Lehrgangs und bei der Abschlussarbeit.

Wer unterrichtet?

Die Teilnehmenden lernen von renommierten Dozentinnen und Dozenten praxisnah und zielgerichtet alles, was sie brauchen, um die Prüfung zur tekom-Zertifizierung erfolgreich zu meistern. Zu den Kursleiterinnen und Kursleitern zählen Professoren der bekannten Hochschulstudiengänge für Technische Redaktion, prominente tekom-Vertreter und Experten aus der Praxis. Durch eine intensive individuelle Betreuung der Abschlussarbeiten profitieren die Teilnehmenden des Zertfizierungslehrgangs zusätzlich von Spezialisten-Know-how.

Was wird geschult?

Der Lehrgang behandelt selbstverständlich all jene Themen, die von der tekom als Pflichtbausteine ausgewiesen sind. Im Kurs mit dem Titel „Rechtliche Anforderungen“ geht es um Ansprüche an die interne und externe Dokumentation, nationale und internationale rechtliche Normen sowie Informationsquellen und Recherchemöglichkeiten. Der Kurs, der tekom-Pflichtbaustein 2 zum Inhalt hat, behandelt die Phasen der Informationsentwicklung, Produkt- und Zielgruppenanalyse, Informationskonzepte, Projektplanung und Organisation. Die Kurse für den dritten Baustein beschäftigen sich mit Strukturieren, Standardisieren, XML sowie den Möglichkeiten verschiedener Redaktionssysteme. Baustein Nummer vier befasst sich mit der professionellen Verwendung deutscher Sprache: Bestandteil des betreffenden Kurses sind sowohl Rechtschreibungs- und Grammatikrichtlinien sowie Stilistik und Verständlichkeit von Texten als auch Redaktions- und Korrekturprinzipien.

Neben den tekom-Pflichtbausteinen sind im Lehrgang folgende Wahlbausteine enthalten: Der Kurs mit dem Titel „Mehrsprachige Dokumentationserstellung und Übersetzung“ setzt sich unter anderem mit interkultureller Kommunikation, computergestützter Übersetzung und dem Management von Projekten zur Erstellung multilingualer Dokumentation auseinander. An einem weiteren Kurstag werden die Grundlagen der Terminologielehre sowie Terminologiedatenbanken und -management vermittelt. „Optische Gestaltung/Layout“ behandelt neben Grundaspekten der Typografie auch Bildschirm- und Webseitengestaltung und führt in die Arbeitsweise von Layoutprogrammen ein. Unter anderem um Gestaltgesetze und Erfordernisse an die Druckvorstufe geht es in „Bildhafte Darstellung und digitale Bildbearbeitung“. Ein weiterer Kurs beschäftigt sich mit Online-Dokumentation – von Grundlagen und Konzepten über Formate, Tools und Techniken bis hin zu Online-Hilfen. „Multimediale Dokumentation“ beleuch-

Termin	Baustein	Dozenten
Präsenzveranstaltungen in München (Freitag/Samstag)		
18./19. Februar 2011	Professionelles Deutsch (tekom-Pflichtbaustein 4 und Aspekte der tekom-Wahlbausteine 7 und 9)	Prof. Dr. Anne Lehrndorfer Prof. Dr. Marita Tjarks-Sobhani
25. Februar 2011	Strukturieren (Teile des tekom-Pflichtbausteins 3)	Prof. Sissi Closs
26. Februar 2011	Informationsentwicklung (tekom-Pflichtbaustein 2)	Ralf Steiner
4. März 2011	Kommunikation, Rhetorik und Präsentation (tekom-Wahlbaustein 18)	Markus Boxleitner
5. März 2011	Rechtliche Anforderungen (tekom-Pflichtbaustein 1)	
18. März 2011	Usability (tekom-Wahlbaustein 13)	Jens Weinhold
19. März 2011	Standardisieren (Teile des tekom-Pflichtbausteins 3)	Prof. Jürgen Muthig
25. März 2011	Optische Gestaltung/Layout (tekom-Wahlbaustein 9)	Prof. Anja Grunwald
26. März 2011	Bildhafte Darstellung und digitale Bildbearbeitung (tekom-Wahlbaustein 10)	
1. April 2011	Terminologie (tekom-Wahlbaustein 8)	Prof. Dr. Petra Drewer
2. April 2011	Mehrsprachige Dokumentationserstellung und Lokalisierung (tekom-Wahlbaustein 7)	Prof. Dr. Petra Drewer Elke Grundmann
8. April 2011	Online-Dokumentation (tekom-Wahlbaustein 11)	Prof. Sissi Closs
9. April 2011	XML, Redaktionssysteme (Teile des tekom-Pflichtbausteins 3)	Prof. Sissi Closs Elke Grundmann
15. April 2011	Online-Hilfe-Tools (tekom-Wahlbaustein 11)	Robert Hinesley
16. April 2011	Multimediale Dokumentation (tekom-Wahlbaustein 12)	Prof. Martin Schober

tet interaktive Programme, moderne Autorenwerkzeuge sowie den zeitlichen und finanziellen Erstellungsaufwand von kompletten Multimedia-Projekten. In „Usability Testing“ wird nach dem Kennenlernen benutzerorientierter Gestaltung unter anderem in Praxisarbeit ein Usability-Test durchgeführt. Und im Kurs „Kommunikation, Rhetorik und Präsentation“ werden die Teilnehmenden schließlich in Sachen Gesprächstechnik und Verhandlungsgeschick auf den neuesten Stand gebracht und intensiv geschult.

Wie sieht das Berufsfeld aus?

Die Aufgabe von Technischen Redakteuren ist es, Produkte für Anwender verständlich darzustellen. Schnelle Entwicklungszyklen, komplexe Technik und strenge gesetzliche Rahmenbedingungen stellen Unternehmen zunehmend vor neue Herausforderungen. Um wettbewerbsfähig zu bleiben, sind Produkte immer stärker zielgruppen- und mediengerecht zu dokumentieren – und das zu vertretbaren Kosten. Die Arbeit in der Technikredaktion hat sich zu einer hochqualifizierten Dienstleistung entwickelt, die über die reine Erstellung und Pflege von Dokumenten hinaus anspruchsvolle Aufgaben wie Projektplanung, Strukturierung großer Informationsbestände und Übersetzungsmanagement umfasst. Technische Redakteure analysieren den Informationsbedarf der Zielgruppen, konzipieren und erstellen verständliche Informationsprodukte und gestalten diese für unterschiedliche Medien. Sie koordinieren Dokumentations- und Übersetzungsprozesse, planen System- und Produktionsabläufe und wirken bei der Gestaltung nutzergerechter Bedienoberflächen mit. Der Bedarf an Fachleuten, die in der Lage sind, diese Aufgaben kompetent zu erfüllen, ist groß. Das breite, vielseitige Betätigungsfeld bietet Abwechslung, Herausforderung und viele Möglichkeiten der Spezialisierung nach eigenen Neigungen und Talenten.

Gute Gründe für die Zertifizierung

Das Zertifikat „Technische Redakteurin/Technischer Redakteur (tekom)“ ist in der Branche als Garant für Fachkompetenz und Know-how bekannt, und bringt einen spürbaren Mehrwert für die Zertifizierten selbst sowie für die Unternehmen, in denen sie tätig sind: Die Teilnehmenden des Zertifizierungslehrgangs erwerben praxisnahe Kompetenzen, weiten ihren Aufgaben- und Verantwortlichkeitsbereich aus und sind nach erfolgreich absolvierter Prüfung mit Hochschulabsolventen gleichgestellt. Das führt nicht nur zur Steigerung des beruflichen Erfolgs sondern gleichzeitig auch zur Sicherung des Arbeitsplatzes. Selbstverständlich profitieren auch Unternehmen davon, wenn ihre Mitarbeitenden das tekom-Zertifikat in der Tasche haben: Der Zuwachs an Fachwissen bringt eine deutliche Qualitätssteigerung, verbesserte Performance und höhere Flexibilität, was wiederum zu klaren Wettbewerbsvorteilen führt.

Der berufsbegleitende tekom-Zertifizierungslehrgang der Comet Unternehmensgruppe ist daher eine Investition in eine qualifizierte berufliche Zukunft, die sich dauerhaft lohnen wird. Die Kosten betragen 4.800 Euro (zzgl. gesetzl. MwSt.). Es besteht die Möglichkeit, die Bildungsprämie der Bundesregierung einzulösen

Weitere Informationen finden Sie unter www.comet.de/schulung/zertifizierung. Comet Communication GmbH, Goethestraße 17, 80336 München, Telefon 089 59989260, Fax 089 59989269, www.comet.de

Die Comet Unternehmensgruppe

Als Full-Service-Dienstleister nehmen die Comet Computer GmbH und die Comet Communication GmbH in allen Bereichen der Technischen Dokumentation eine führende Marktposition ein. Seit über 20 Jahren konzipieren, realisieren und übersetzen die Dokumentationsprofis Technische Dokumentation für elektronische und gedruckte Medien wie Handbücher, Online-Hilfen, Gebrauchsanweisungen und Installationsanleitungen. Mit ihrem Expertenwissen über innovative Open-Source-Technologien (Wiki, DITA) und als Partner wegweisender Tool-Hersteller bietet die Unternehmensgruppe ihren Kunden individuelle Lösungen für alle Anforderungen. Zum breiten Tätigkeitsspektrum von Comet zählen zudem der Vertrieb von Redaktions-Tools sowie das bedarfsgerechte Einrichten von Dokumentationsumgebungen. Ein sehr umfangreiches Schulungsangebot zu aktuellen Themen rundet das Dienstleistungsportfolio der Comet-Gruppe ab.
www.comet.de

GBI-Genios präsentiert das neue GENIOS 2011

Im Februar ist es soweit: Dann wird „der Schalter umgelegt“ und GBI-Genios präsentiert unter www.genios.de das neue GENIOS 2011, das an die Stelle der alten Rechercheoberfläche treten wird. Wer die neue Oberfläche näher kennen lernen möchte, kann sich unter www.genios.de/roadshow direkt zu einem der Roadshow-Termine im Februar und März anmelden. Auf dem Programm stehen eine anwendungsorientierte Live-Präsentation sowie Hintergründe und Wissenswertes zum Relaunch (u. a. Informationen zur neuen Automatischen Indexierung). Im Anschluss besteht bei einem Business-Lunch die Gelegenheit zum Expertenaustausch.

Programm

09:30 Uhr
Anmeldung und Begrüßungskaffee

10:00 Uhr
Begrüßung und aktuelle GENIOS News

10:20 Uhr
Hintergründe und Wissenswertes zum Relaunch

10:40 Uhr
Live-Einführung in das neue GENIOS 2011

12:00 Uhr
Business-Lunch und Expertenaustausch

Termine & Orte

17.02.2011 München

22.02.2011 Düsseldorf

23.02.2011 Frankfurt

24.02.2011 Stuttgart

02.03.2011 Berlin

Kontakt:

GBI-Genios Deutsche Wirtschaftsdatenbank GmbH

info@genios.de

www.genios.de/roadshow

Internetnutzung in Freiheit und Fairness

Fragen an Dorothee Bär, MdB, von Ronald Kaiser und Prof. Dr. Wolfgang Ratzek

Über welche Quellen halten Sie sich über die Entwicklung in der IT-Welt, aber auch über Bibliotheken auf dem Laufenden?
Da gibt es ja ein großes Angebot – und vor allem im IT-Bereich ständig eine Fülle an Neuigkeiten. Natürlich informiere ich mich regelmäßig im Netz und über Fachzeitschriften. Aber auch in Gesprächen mit Experten und Bekannten aus meinem Freundeskreis, stoße ich immer wieder auf neue, interessante Details.

Welche Schlüsselqualifikationen wünschen Sie sich von zukünftigen Schülergenerationen?
Es zeigt sich immer deutlicher: Die Bereitschaft zum lebenslangen Lernen wird eine wesentliche Bedeutung für die Zukunftschancen der jungen Generation haben. Gerade im IT-Bereich verändert sich das Wissen ja rasant schnell. Neue Technologien revolutionieren unseren Lebensalltag. Allein um in diesem Bereich auf der Höhe der Zeit zu bleiben und den Überblick zu behalten, braucht es die Bereitschaft, immer wieder neu dazu zu lernen. Das heißt auch: Das Handwerkszeug, das wir unseren Kinder mit auf den Weg geben müssen, ist außer Lesen, Schreiben und Rechnen auch die Fähigkeit, mit den neuen Medien gut umzugehen.

In einer sich veränderten Wissensgesellschaft ist die Vermittlung von Informations- und Medienkompetenz wichtig. Wer sollte sich dafür verantwortlich zeigen?
Wir müssen schon im Kindesalter anfangen, unsere junge Generation auf den richtigen Umgang mit Medien vorzubereiten. Da sind natürlich in erster Linie die Eltern und die Familie gefragt. Aber auch in unseren Schulen muss die Vermittlung von Medienkompetenz besonderes Gewicht bekommen. Ich plädiere dafür, ein eigenes Schulfach Medienkunde einzuführen. Kinder und Jugendliche sind heute tagtäglich mit aggressiver Werbung konfrontiert und haben leichten Zugang zu gewaltverherrlichenden Darstellungen. Damit sie dagegen gerüstet sind, reicht es nicht mehr aus, das Thema Medien hier und dort im Unterricht anzusprechen. In einem eigenen Schulfach Medienkunde hätten Lehrer genügend Zeit, mit ihren Schülern über deren Erfahrungen im Umgang mit Medien zu sprechen und sie auf die Gefahren hinzuweisen. Mit dem fächerübergreifenden Lernziel der Medienerziehung leisten unsere Lehrer bereits heute sehr viel. Wir können diese Arbeit mit einem eigenständigen Schulfach Medienkunde noch weiter stärken. Aber das Thema Medienkompetenz hört nicht an der Schultür auf – und es gilt über Generationsgrenzen hinweg für alle in unserer Gesellschaft. Dieser Unterricht sollte in Zusammenarbeit mit externen Bildungsträgern z.B. Bibliotheken erfolgen.

Dorothee Bär, MdB
- Geboren am 19. April 1978 in Bamberg, aufgewachsen und wohnhaft in Ebelsbach, Landkreis Haßberge. Verheiratet, eine Tochter.
- 1999 bis 2005 Studium der Politikwissenschaft an der Hochschule für Politik in München und am Otto-Suhr-Institut der Freien Universität in Berlin, Diplom Journalistin bei verschiedenen Radiostationen, Tageszeitungen und Agenturen.
- Seit 2002 Mitglied des Deutschen Bundestages
- Stellv. Generalsekretärin der CSU und stellv. Vorsitzende der Jungen Union Deutschlands
- Familienpolitische Sprecherin der CDU/CSU-Bundestagsfraktion und Mitglied im Ausschuss für Kultur und Medien
- Mitglied u.a. in der Vorauswahljury des Deutschen Filmpreises, der Jury des Deutschen Computerspielepreises und im Kuratorium der Bundeszentrale für politische Bildung.

Wenn in den Medien von Bildung die Rede ist, dann stehen Kindergärten, Schulen und Hochschulen im Mittelpunkt. Dagegen bleiben Bibliotheken als wichtige informationsvermittelnde Institution ausgeblendet. Wie könnten Ihrer Meinung nach Bibliotheken stärker in das Bildungsspektrum integriert werden?
Bibliotheken sind ohne Frage eine außerordentlich wichtige, unersetzbare und unverzichtbare Bildungseinrichtungen. Sie bewahren unser kulturelles Erbe und unsere Geschichte, vermitteln Wissen und leisten darüber hinaus einen großen Beitrag zur kulturellen Integration. Eine bessere Einbindung der Bibliotheken ins Bildungsspektrum kann beispielsweise dadurch erfolgen, dass wir unsere Kinder und Jugendlichen im Bereich der schulischen Bildung gezielt an die Bibliotheken heranführen. Beispiele aus der Praxis zeigen, dass Schulbesuche der örtlichen Bibliothek eine ganz hervorragende Möglichkeit bieten, Kinder und Jugendliche mit den Einrichtungen vertraut zu machen. Ganz besonders gut funktioniert das dann, wenn die Bibliothek im selben Gebäude ist wie die Schule, der Kindergarten oder die Hochschule. Darüber hinaus sind Kooperationen zwischen Schulbibliotheken und öffentlichen Bibliotheken eine gute Möglichkeit zur besseren Integration ins allgemeine Bildungsspektrum.

Im Schlussbericht der Enquete-Kommission „Kultur in Deutschland" für den Deutschen Bundestag empfahl die Enquete-Kommission die Einführung eines Bibliotheksgesetzes. Wie stehen Sie dazu?
Ich befürworte das. So können die Aufgaben und die Finanzierung geregelt werden und es wird zugleich ein Zeichen der Wertschätzung für unsere Bibliotheken gesetzt. Es muss unser aller Ziel sein, dass die Kultur unseres Landes gelebt, gepflegt und geschützt wird. So ist es unsere Pflicht, und nicht nur reine Freiwilligkeit, öffentliche Bibliotheken zu schützen.

In diesem Schlussbericht wird auch der Vorschlag des Deutschen Bibliotheksverbandes aufgenommen, eine Bibliotheksentwicklungsagentur (BEA) nach dem Wegfall des Deutschen Bibliotheksinstituts (DBI) ins Leben zu rufen. Würden Sie diesen Vorschlag unterstützen?Das hängt von der konkreten Ausgestaltung dieser Agentur ab. Wichtig ist doch, dass ein mehr an Verwaltung vermieden wird und gleichzeitig unbürokratisch neue, innovative Wege gefunden werden.

Die Diskussion um die Buchpreisbindung ist im vollen Gange. Verstöße im Online-Buchhandel sind seit längerem bekannt. Wie schätzen Sie die Entwicklung national und international ein?
Generell halte ich die Buchpreisbildung für absolut richtig. Sie ist Garant der kulturellen Vielfalt und trägt damit zum Erhalt unserer Verlage und Buchhandlungen bei. Man sollte auch keinen Unterschied zwischen der realen und der virtuellen Welt machen, damit gilt die Buchpreisbindung selbstverständlich auch für Anbieter im Internet. Bei der Umsetzung warten hier allerdings sowohl international also auch national noch einige Herausforderungen auf uns. Meine Fraktion befindet sich daher in einem regen Austausch mit dem Börsenverein des deutschen Buchhandels.

Das WWW, vor allem die Web 2.0-Entwicklung tragen zur Demokratisierung mit allen gewünschten und unerwünschten Erscheinungsformen bei. Die ordnungspolitischen Instrumente verfehlen häufig ihre Wirkung. Nun hat der Bund Deutscher Kriminalbeamter (BDK) der Bundesregierung ein 15-Punkte-Programm zur Bekämpfung der Internetkriminalität vorgeschlagen. Darin werden unter anderem eine Ausweispflicht und ein Notschalter im Falle einer Cyber-Attacke thematisiert. Wie stehen Sie dazu?
Ich wünsche mir eine Internetnutzung in Freiheit und Fairness. Das Internet ist eine unglaubliche Erfolgsgeschichte als Ort der freien Entfaltung persönlicher, wirtschaftlicher und politischer Interessen. Zur Freiheit gehört aber auch Verantwortung. Wenn dieses Verantwortungsbewusstsein fehlt und etwa durch unfairen Gebrauch die Freiheit anderer gefährdet wird, dann muss man Regeln aufstellen und durchsetzen, die diese Freiheit sichern. Das ist nach unserem demokratischen Verständnis eine zentrale Aufgabe des Staates. Ich glaube wir sind uns einig, dass wir nicht alle Formen und Auftritte, die sich im Web tummeln, als kulturell hochwertig und geschmackvoll bezeichnen würden. Dort, wo Recht und Gesetz verletzt werden, muss der Staat einschreiten. Über die konkreten Wege muss man sicher diskutieren.

Abschließend noch eine Frage: Welches Buch haben Sie zuletzt gelesen?
Ich habe gerade wieder angefangen, mich in Hape Kerkelings Erfahrungen auf dem Jakobsweg reinzulesen. „Ich bin dann mal weg“ hat mir beim ersten Mal lesen schon gut gefallen, ich freue mich jetzt schon auf die zweite Runde!

CeBIT-Premiere des Elektronischen Lesesaals „MyBib eRoom“

Fraunhofer IAIS und ImageWare präsentieren den kompletten Workflow: Von der Digitalisierung über die automatische Bildbearbeitung und Indexierung bis zur Präsentation im geschützten Bereich wird die gesamte Verarbeitungskette vorgestellt.
MyBib eRoom ist eine von Fraunhofer IAIS und ImageWare entwickelte Präsentationsplattform für Digitalisate und Volltexte. Das System stellt digitalisierte Dokumente wie Bücher, Bilder oder Schriftdokumente in einem geschützten elektronischen Bereich bereit. Die CeBIT-Präsentation zeigt die Digitalisierung von Vorlagen mit dem „ScanRobot“, einem Scan-Roboter von Treventus. Dieses speziell für Massendigitalisierungsverfahren konzipierte Scansystem kann bis zu 2.500 Seiten/h bearbeiten. Die gescannten Images werden anschließend in die von ImageWare entwickelte Scan- und Capturingsoftware BCS-2® importiert. Dort erfolgt eine OCR-Erkennung und die Indexierung mit anschließendem Export der Daten in den elektronischen Lesesaal MyBib eRoom. Nur in diesem geschützten Bereich können die Digitalisate gelesen werden. MyBib eRoom schützt Eigentümer und Anwender vor Daten-Missbrauch durch Dritte. Authentifizierte Benutzer können an registrierten Arbeitsplätzen zum Beispiel auf Lehrbücher, Zeitschriften, Verträge, Pläne und andere sensible Dokumente zugreifen. Die CeBIT-Präsentation in Halle 9, Stand B36 simuliert Workflowszenarien aus dem Bibliotheks-, Archiv-, Verlags- und Wirtschaftsumfeld. Weitere Informationen unter www.imageware.de.

IMPRESSUM

Gegründet von H.-K. Soeken †
unter dem Titel Nachrichten für Dokumentation (NfD)
Herausgegeben von der Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis e.V. (DGI)
Präsident: Prof. Dr. Stefan Gradmann
Windmühlstraße 3
60329 Frankfurt am Main
Telefon: (0 69) 43 03 13
Telefax: (0 69) 4 90 90 96
mail@dgi-info.de
www.dgi-info.de
Mitteilungsblatt des Normenausschusses Bibliotheks- und Dokumentationswesen im DIN Deutsches Institut für Normung e.V., der Fachgruppe Dokumentation im Deutschen Museumsbund und der Arbeitsgemeinschaft der Spezialbibliotheken (ASpB)

Redaktionsbeirat
Dr. Sabine Graumann, München (Informationswirtschaft)
Prof. Dr. Hans-Christoph Hobohm, Potsdam (Management von Informationseinrichtungen)
Prof. Dr. Rainer Kuhlen, Konstanz (Informationswissenschaft)
Prof. Dr. Dirk Lewandowski, Hamburg (Suchmaschinen, Internet)
Dr. Philipp Mayr, Bonn/Darmstadt (Information Retrieval, Informetrie, Wissensrepräsentation)
Prof. Dr. Wolfgang Ratzek, Stuttgart (Informationspraxis)
Prof. Dr. Christian Schlögl, Graz (Metriken, Informations- und Wissensmanagement)
Prof. Dr. Ralph Schmidt, Hamburg (Newcomer Report, Medien)
Prof. Dr. Wolf G. Stock, Düsseldorf (Wissensrepräsentation und Information Retrieval)

Redaktion
Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V.
Marlies Ockenfeld (verantwortlich)
Viktoriaplatz 8, 64293 Darmstadt
Telefon: (0 61 51) 99 71 17
Telefax: (0 69) 4 90 90 96
ockenfeld@dgi-info.de
Daniel Ockenfeld (Redaktionsass.)

Verlag
Dinges & Frick GmbH
Greifstraße 4
65199 Wiesbaden
Postfach 1564
65005 Wiesbaden
Telefon: (06 11) 9 31 09 41
Telefax: (06 11) 9 31 09 43
Bankverbindung:
Wiesbadener Volksbank
BLZ 510 900 00, Kto-Nr. 714 22 26
Postbank Frankfurt
BLZ 500 100 60, Kto.-Nr. 267 204-606

Objektleitung
Erwin König,
e.koenig@dinges-frick.de

Anzeigenservice
Ursula Hensel Anzeigenservice
Hermann-Schuster-Straße 39
65510 Hünstetten-Wallbach
Telefon: (0 61 26) 57 08 82
Telefax: (0 61 26) 58 16 47
ursula.hensel@unitybox.de
Rocco Mischok
Verlag Dinges & Frick GmbH
Greifstraße 4
65199 Wiesbaden
Telefon: (06 11) 3 96 99-60
Telefax: (06 11) 3 96 99-30
r.mischok@dinges-frick.de

Gestaltung
Meinhard Zielke, Wiesbaden

Druck
Dinges & Frick GmbH
Greifstraße 4
65199 Wiesbaden
Postfach 2009
65010 Wiesbaden
Telefon: (06 11) 3 96 99-0
Telefax: (06 11) 3 96 99-30
df@dinges-frick.de
www.dinges-frick.de

Hinweis
Die Aufsätze stellen ausschließlich die Meinung der Autoren dar. Der Inhalt wurde sorgfältig und nach bestem Wissen erarbeitet. Die Originalbeiträge werden einem Begutachtungsverfahren durch die Mitglieder des Redaktionsbeirats und weitere international anerkannte Fachleute unterworfen. Dennoch kann von Verlag und Redaktion eine Gewährleistung auf Richtigkeit und Vollständigkeit nicht übernommen werden. Die Beiträge und die grafischen Darstellungen unterliegen dem Urheberrecht. Nachdruck und Vervielfältigung jeglicher Art bedürfen der Genehmigung des Verlages und der Autoren.
Einreichungen werden als Word-Dateien erbeten. Abbildungen, Fotos und Grafiken müssen eine Auflösung von 300 dpi haben und sollen als separate Dateien beigefügt werden. Abhandlungen müssen ein deutsches und ein englisches Abstract sowie einen deutschen und einen englischen Titel haben. Die Publikationssprache in der IWP ist grundsätzlich deutsch. Die Redaktion strebt eine möglichst umfassende Berichterstattung über neue informationswissenschaftliche Publikationen an und nimmt gerne Rezensionsangebote aus diesem Umfeld an.

Erscheinungsweise/ Bezugspreise
Sechs Hefte jährlich
(Doppelausgaben März/April und September/Oktober)
Jahresabonnement EUR 208,–
Schüler/Studenten EUR 148,–
Einzelheft EUR 40,–
inkl. Versandkosten/Porto im Inland,
Versand in Europa + EUR 28,–
Versand Welt / Luftpost auf Anfrage.
Das Abonnement gilt für mindestens ein Jahr und kann danach bis sechs Wochen zum Ende des Bezugszeitraums gekündigt werden.

Redaktionsschluss für

Heft 2/3	4. März 2011
Heft 4	6. Mai 2011
Heft 5	1. Juli 2011

2011

31. Januar bis 2. Februar *Szombathely, Ungarn*	BOBCATSSS 2011	www.bobcatsss2011.com
1. bis 3. Februar *Karlsruhe*	LEARNTEC 2011 19. Internationale Leitmesse und Kongress für professionelle Bildung, Lernen und IT	Karlsruher Messe- und Kongress-GmbH, Festplatz 9, 76137 Karlsruhe, Telefon 0721 3720-0, Fax 0721 3720-2116, info@kmkg.de, www.learntec.de/
16. bis 19. Februar *Hannover*	11. DVMD Fachtagung „Medizinisches Informationsmanagement 3.0“	Sabine Kapsamer, DVMD Geschäftsstelle, Postfach 10 01 29, 68001 Mannheim, Telefon 0621 71761393, Fax 0621 71761395, dvmd@dvmd.de, www.dvmd.de
22. bis 26. Februar *Stuttgart*	Didacta – die Bildungsmesse	Joachim Sauter, Landesmesse Stuttgart GmbH, Messepiazza 1, 70629 Stuttgart, Telefon 0711 18560-2448, Fax 0711 18560-2275 joachim.sauter@messe-stuttgart.de, www.didacta-stuttgart.de
1. bis 5. März *Hannover*	CeBIT 2011 Gastland Türkei	Deutsche Messe, Messegelände, 30521 Hannover, Telefon 0511 89-0, Fax 0511 89-32626, www.cebit.de/cebit2011
9. bis 11. März *Hildesheim*	12. ISI-Konferenz/IuK 2011 Information und Wissen – global, sozial und frei?	Stiftung Universität Hildesheim, Institut für Informationswissenschaft und Sprachtechnologie, Lübecker Straße 3, 31141 Hildesheim, isi2011@uni-hildesheim.de, www.isi2011.de
11. bis 12. März *Hamburg*	4. BibCamp Wir entern die Bibliothek – Das BarCamp für BibliothekarInnen und andere Hacker	www.bibcamp.de
14. bis 15. März *Singapur*	International Information Conference on Search, Data Mining and Visualization (II-SVD)	Infonortics Ltd., Gable House, 46 High Street, Malmesbury SN16 9AT, England, Telefon +44(0)1666 823 252, Fax +44(0)1865 233 981, contact-1@infonortics.com, www.infonortics. com
17. bis 20. März *Leipzig*	Leipziger Buchmesse Ehrengast Serbien	Leipziger Messe GmbH, Projektteam Buchmesse, Messe-Allee 1, 04356 Leipzig, www.leipziger-messe.de
22. März *Potsdam*	2. I-Science Day: vom Web2.0 zum semantischen und visuellen Web3.0	Prof. Dr. Stephan Büttner, Fachhochschule Potsdam, Fachbereich Informationswissenschaften, Friedrich-Ebert-Str. 4, 14467 Potsdam, Telefon 0331 580-1517, Fax 0331 580-1599, st.buettner@fh-potsdam.de
6. bis 7. April *Bad Nauheim*	4. Competitive Intelligence Symposium und 3. ICI Practitioner-Konferenz	Thorsten Bill, GICI Institut für Competitive Intelligence mbH, Korngasse 9, 35510 Butzbach, Telefon 06033 971 377, bill@competitive-intelligence.com, www.competitive-intelligence.com
6. bis 8. April *Karlsruhe*	DGI-Praxis-Tage 2011 Information: gift or poison? Die Kompetenz entscheidet	DGI-Geschäftsstelle, Windmühlstraße 3, 60329 Frankfurt am Main, Telefon 069 430313, Fax 069 4909096, mail@dgi-info.de, www.dgi-info.de
8. April *Karlsruhe*	DGI-Mitgliederversammlung 2011	DGI-Geschäftsstelle, Windmühlstraße 3, 60329 Frankfurt am Main, Telefon 069 430313, Fax 069 4909096, mail@dgi-info.de, www.dgi-info.de
14. bis 15. April *Potsdam*	tekom Frühjahrstagung 2011	Gesellschaft für Technische Kommunikation e.V. (tekom), Dr. Meike Wiehl, Rotebühlstraße 64, 70178 Stuttgart, Telefon 0711 65704-0, Telefax 0711 65704-99, info@tekom.de, www.tekom.de
2. bis 6. Mai *Mainz*	JAX 2011 - Die Konferenz für Java, Web, Architektur, Cloud & Agile	Software & Support Media GmbH, Geleitstraße 14, 60599 Frankfurt am Main, Telefon 069 630089-0, info@jax.de, www.jax.de
3. bis 5. Mai *Berlin*	Informare! - Der Weg zum Wissen. Digital.	Arnoud de Kemp, digiprimo GmbH & Co. KG, Postfach 103305, 69023 Heidelberg, info@infomare-wissen-und-koennen.com, www.informare-wissen-und-koennen.com
11. Mai *Magdeburg*	Frühjahrstagung der Fachgruppe Dokumentation im Deutschen Museumsbund	Prof. Monika Hagedorn-Sauppe, Institut für Museumsforschung, In der Halde 1, 14195 Berlin, Telefon 030 8301-460, Fax 030 8301-504, m.hagedorn@smb.spk-berlin.de, www.museumsbund.de
23. Mai *Pullach bei München*	11. Deutsches Single Source Forum	Comet Communication GmbH, Goethestraße 17, 80336 München, Telefon 089 59989260, Fax 089 59989269, info@comet-comm.de, www.single-source-forum.de
26. bis 27. Mai *Malaga, Spanien*	19. EBLIDA-Konferenz	EBLIDA, Grote Marktstraat 43, 2511 BH Den Haag, Niederlande Telefon +31 70 30 90 551, Fax +31 70 30 90 558, eblida@eblida.org, www.eblida.org
7. bis 10. Juni *Berlin*	100. Deutscher Bibliothekartag Bibliotheken für die Zukunft – Zukunft für die Bibliotheken	K.I.T. GmbH, Association & Conference, Management Group & Co. KG, Kurfürstendamm 71, 10709 Berlin, Telefon 030 24603-0, Fax 030 24603-200, bibliothekartag2011@kit-group.org, www.kit-group.org
17. Juni *Köln*	3. Nacht der Technik Köln	http://nacht-der-technik.de
17. bis 19. Juni *Paris, Frankreich*	i-expo 2011	Groupement Français de l'Industrie de l'Information, 25, rue Claude Tillier - 75 012 Paris Telefon +33 1 43729652, Fax +33 1 43720456, gfii@gfii.asso.fr, www.i-expo.net
29. Juni bis 2. Juli *Barcelona, Spanien*	40. LIBER-Konferenz	IFEMA - Feria de Madrid Parque Ferial Juan Carlos I, Madrid, , Spanien, Telefon +34 902 22 15 15, Fax +34 91 722 58 01
13. bis 18. August *San Juan, Puerto Rico*	Weltkongress Bibliothek und Information: 77. IFLA Generalkonferenz und Generalversammlung	Hella Klauser, Sekretariat des IFLA-Nationalkomitee, c/o Deutscher Bibliotheksverband e.V., Straße des 17. Juni 114, 10623 Berlin, Telefon 030 64498 9916, Fax 030 64498 9929, klauser@bibliotheksverband.de, www.bibliotheksverband.de

Informare!

3. - 5. Mai 2011
im Moskau
in Berlin

www.informare-wissen-und-koennen.com

Mehr WISSEN. Besser Entscheiden.

- Die Konferenz
- Die Poster Sessions
- Die Workshops
- Die Ausstellung „Die Kunst der Information“
- Das Informare!Camp
- Die lange Nacht der Suchmaschinen

INFORMARE!COMMUNITY

HOT NEWS

Wir sehen uns in Berlin vom 3. - 5. Mai 2011. Alles unter einem Dach im berühmten „Moskau“

Wissen und Können verpflichtet!

Wir starten eine neue, jährlich stattfindenden Konferenz zum hochaktuellen Thema Informationskompetenz im „Moskau“ in Berlin unter der Schirmherrschaft der Frankfurter Buchmesse.

Gleichzeitig starten wir eine neue und sichere „Community“ für mehr Informationskompetenz.

Machen Sie mit, seien Sie dabei!

Aktuelle Infos zu allen Angeboten:

für **Konferenzteilnehmer**

für **Aussteller** zur Ausstellung „Die Kunst der Information“

für **Sponsoren**

Informare!

www.informare-wissen-und-koennen.com