

Information Retrieval is for Everybody – Beobachtungen und Thesen

Christian Wolff

Universität Regensburg

D-93040 Regensburg

christian.wolff@sprachlit.uni-regensburg.de

Abstract

In this paper, the idea of *ubiquitous information retrieval* is presented in a storytelling manner. Starting from a rough review of information retrieval system usage, some empirical hints on IR in everyday life are given. Ch. 4 explores the heterogeneity of interaction with IRS for one day in the life of a (common search engine) user. In ch. 5 summarizes these observations and suggests research approaches for modelling information retrieval as an essential component of interaction in the information society.

1 Einleitung

Information Retrieval (IR) has become, mainly as a result of the huge impact of the World Wide Web (WWW) and CD-ROM industry, one of the most important theoretical and practical research topics in Information and Computer Science. (Dominich, 2001, p. xv)

Im Zentrum dieses Aufsatzes steht die These, dass im Kontext der Verfügbarkeit von Information Retrieval-Systemen (IRS) nicht nur im „klassischen“ Bereich der Fachinformation, sondern für buchstäblich alle Bevölkerungs- und Altersschichten und Lebenslagen neue Herausforderungen auf die IR-Forschung zukommen, was sich als Forderung nach einem im Sinne der Fachinformation „deprofessionlasiertem *ubiquitous information retrieval*“ – oder eben dem *Information Retrieval for Everybody* – zu spitzen lässt.

Kap. 2 versucht einen groben Abriss der Entwicklung typischer Nutzungsformen von IRS zu geben. Kap. 2 begründet die Ausgangsthese anhand empirischer Daten und einzelner Beobachtungen. Kap. 3 stellt am Beispiel eines Tages im Leben eines *common search engine user* exemplarisch dar, wie weit Informationssysteme bereits den Alltag durchdringen und welche Probleme sich dabei ergeben. Wesentliche Beobachtungen und Thesen, die sich daraus ableiten lassen, sind in Kap. 4 zusammengefasst.

2 Entwicklungsphasen des Information Retrieval

Information Retrieval hat im Vergleich mit anderen Feldern informationswissenschaftlicher oder informatiknaher Forschung eine durchaus lange Tradition: Bibliothekswis-

senschaftliche Wurzeln reichen weit vor das Zeitalter der digitalen Informationstechnik zurück, frühe Visionen wie Bushs *As we may Think* (Bush, 1991) enthalten bereits wesentliche Konzepte zukünftiger IR-Systeme und die ersten IR-Systeme gehören sicher zu den frühesten Beispielen (professionell) genutzter Informationssysteme (Lilley & Trice, 1989). Auch wenn sich eine zeitliche Einteilung nicht mit letzter Trennschärfe vollziehen lässt, so kann man die (Nutzungs-)Geschichte der Information Retrieval-Systeme (IRS) grob in folgende Phasen einteilen.

2.1 IRS als Instrumente professioneller Dienstleister in vermittelter Kommunikation

In diesem Szenario (idealtypisch: bibliographische Datenbanken auf einem Host, zugänglich über formale Recherchesprache für professionelle *information workers*) existiert der Endnutzer eines Informationssystems noch nicht, Informationsbedürfnisse werden dem professionellen Forscher kommuniziert, der sie dann in eine Retrievalsprache umformuliert und Recherchen durchführt, Feedback des Benutzers erfolgt allenfalls bei Durchsicht der Suchergebnisse.

2.2 IRS als Werkzeug des Fachwissenschaftlers

Mit Verfügbarkeit graphisch-direktmanipulativer Benutzerschnittstellen, spätestens aber mit der Verbreitung formularbasierter Suchzugänge zu professionellen Informationssystemen im WWW, tritt der vermittelten Suche die unmittelbare des IRS durch den Endnutzer an die Seite. Auch wenn die vermittelte Recherche in forschungsnahen unternehmen, Hochschulen und Informationszentren bis heute eine wichtige Rolle spielt, dürfte doch die Mehrzahl von Recherchen mittlerweile direkt durch den primär Suchenden erfolgen. Die immer bessere Kopplung von bibliographischen Nachweissystemen mit den Volltexten digitaler Bibliotheken (insbesondere Journals und Proceedings) macht diesen unmittelbaren Zugriff besonders attraktiv.

2.3 Webbasierte Suche

Mit der Verfügbarkeit großer Informationsmengen im World Wide Web und den Suchmaschinen kommt ab Mitte der 90er Jahre des vergangenen Jahrhunderts ein weiteres Szenario hinzu: Information ganz unterschiedlicher Qualität wird ohne die Barrieren, die durch Lizenzen und eingeschränkte Verfügbarkeit gesetzt sind, unmittelbar zugänglich und deckt dabei auch Bereiche ab, die durch die traditionelle Fachinformation nicht erschlossen werden.

2.4 Post-web Suche: „Jeder sucht nach Allem“

Erst in Ansätzen erkennbar ist ein letztes Szenario, dem sich der nachfolgende Teil dieses Aufsatzes widmen wird: Die weitgehende „Digitalisierung“ (*the digitization of the world picture* (Ceruzzi, 2003, p. 346ff)) immer weiterer Bereiche der alltäglichen Informationsnutzung und die zunehmend nur noch digital vorliegende persönliche und private Information.

Deutlich wird aus den Szenarien, dass einerseits immer größere Anforderungen an den Endnutzer hinsichtlich seiner Informationskompetenz zu stellen sind (Kenntnis von Retrievalsprachen, Bewertung der Informationsqualität frei verfügbarer Information, Kenntnis geeigneter Informationssysteme (Suchmaschinen etc.)), andererseits eine klare Zielgruppe für Informationssysteme nicht mehr zu erkennen ist: Buchstäblich jeder, in jedem Alter, verwendet Informationssysteme in Beruf und Alltag.

3 Benutzer und Informationskompetenz

Einige empirische Daten und Beobachtungen mögen dies verdeutlichen: Nach der der jährlichen ARD-/ZDF-Online-Studie (hier: Ausgabe 2005) nutzen mittlerweile knapp 60% der Deutschen regelmäßig das Internet, in der Gruppe der 14-19jährigen sind es sogar 90% (van Eimeren & Frees, 2005). Lediglich die höheren Altersgruppen weisen (noch) eine deutlich niedrigere Nutzungsrate für Online-Medien auf. Eine empirische Studie zur Informationskompetenz, die der Autor 2005 an einem Oberpfälzer Gymnasium und an der Universität Regensburg bei ca. 250 Schülern und Studenten durchgeführt hat, ergab erwartungsgemäß eine fast 100%ige private Verfügbarkeit von Computer und Internet und ebenso eine durchgängige Nutzung von Suchmaschinen, insbesondere Google (Hochholzer & Wolff, 2005)¹. Kleinere Vergleichsstudien, die im Sommersemester 2006 an Regensburger Haupt- und Realschulen im Rahmen eines Seminars zur Informationskompetenz durchgeführt wurden, zeigen ein grundsätzlich identisches Bild. Anlässlich eines Vortrags im Rahmen der Regensburger Universität für Kinder konnte der Autor im Juli 2006 bei einem Auditorium von ca. 700 Kindern im Alter zwischen sechs und zehn Jahren feststellen, dass nahezu 100% der Zuhörer nicht nur Internetnutzer sind, sondern auch Suchmaschinen wie Google benutzen (Hammwöhner & Wolff, 2006 (erscheint)). Auch wenn dies sicher keine belastbare empirische Evidenz ist, wird man doch feststellen können, dass Information Retrieval mittlerweile ein Gebiet ist, das auch junge und jüngste Zielgruppen erreicht.

Deutlich wird dabei auch, dass Suchmaschinen traditionelle Mittel der Informationserschließung wie Bibliotheken bereits verdrängen. Gleichzeitig sind Kompetenzdefizite offensichtlich, wie sie bereits frühe Benutzerstudien zu Suchmaschinen nahe legen (Wolff, 2000 m.w.N.): Kaum Nutzung von Suchoperatoren, Verwendung nur weniger Suchbegriffe, vage Vorstellung von den Möglichkeiten, gezielte Recherchestrategien aufzubauen, kein Bewusstsein der sprachlichen Problematik (Flexion, Komposita) der Suche, ein mangelhaftes Verständnis von

geeigneten Kriterien, die Qualität von Information beurteilen.

4 IR im Alltag – *a Day in the Life of the Common (Search Engine) User*

Michael Lesk hat in seinem bekannten Aufsatz *The Seven Ages of Information Retrieval* die Entwicklungsgeschichte des IR mit den Prognosen von Vannevar Bush (Bush, 1991) kontrastiert und sich dabei der Shakespeareschen Metapher von den sieben Lebensaltern bedient (Lesk, 1995). Nachfolgend wird in ähnlicher Weise der hypothetische Tagesablauf eines Internet- bzw. WWW-Nutzers (*the common search engine user*) nachvollzogen. Als Ausgangsmaterial werden dabei einige Studien zur Evaluation von IRS herangezogen, die 2005 und 2006 im Rahmen von Information Retrieval-Projektseminaren an der Universität Regensburg durchgeführt wurden. Bewertet wurden dabei unter anderem: Bilddatenbanken, Bild-, Audio- und Videoverwaltungssysteme, Musikdownload-Plattformen, Partnerbörsen im World Wide Web, Desktopsuchmaschinen und Produktsuchmaschinen.

Bevor er (oder sie) in den Tag startet, möchte der Nutzer seinen MP3-Player mit geeigneter Musik bestücken. Dazu recherchiert er in eigenen Beständen, die er mit Hilfe online verfügbarer Datendienste wie *Freedb* (<http://freedb.org>) mit Metadaten (ID3-Tags) aufbereitet hat oder er lädt sich Musikstücke von einer Online-Plattform (z. B. *MusicLoad* (<http://www.musicload.de>), *iTunes*, (<http://www.apple.com/de/itunes/>)) herunter. Die Recherche scheitert teilweise schon an orthographischen Schwierigkeiten (*Mozart* vs. *Mozart*, *W.A.* vs. *Wolfgang Amadeus Mozart* ...), an uneinheitlicher Verwendung der Metadaten („Interpret“ vs. „Komponist“ vs. „Teilnehmender Interpret“) oder an unterschiedlichen Genreklassifikationen (vgl. dazu die empirischen Studien (Bainbridge, Cunningham, & Downie, 2003; Cunningham, Jones, & Jones, 2004)). *Query by humming* oder *whistling* kommt morgens nicht in Frage und eine Ähnlichkeitssuche steht nicht zur Verfügung (zum Stand der Music IR vgl. die ISMIR-Proceedings² (<http://ismir.net>) und (Byrd & Crawford, 2002; Downie, 2004)). Neben den Problemen der Recherche treten vergleichbare Schwierigkeiten auch bei der lokalen Verwaltung einer Vielzahl von Musikstücken (als MP3 konvertierte CDs, online erworbene Lizenzen etc.) auf; zur inhaltlichen Erschließungs- und Recherche-problematik tritt die Frage nach der rechtlichen Verfügbarkeit auf verschiedenen Rechnersystemen bzw. Medienplayern (Desktop im Büro; MP3-Player, Laptop zu Hause ...).

Am Vormittag – nehmen wir an, der Benutzer sei wissenschaftlich tätig – sucht er nach Literatur, zunächst in frei verfügbaren Datenbanken wie *Google Scholar* (<http://scholar.google.de/>) oder *Citeseer* (<http://citeseer.ist.psu.edu/>). Trotz ähnlich anmutender Benutzerschnittstellen sind die Ergebnisse aber höchst unterschiedlich. Dem Benutzer ist weder klar, welches Retrievalmodell jeweils zugrunde liegt, noch wie sich die Datenbestände zusammensetzen oder ob er Zugriff auf elektronische Volltexte erwarten kann. Er beschließt daher, seine Suche mit professionellen Werkzeugen fortzusetzen. Mit Hilfe einer „Meta-Meta-Datenbank“ wie dem Datenbank-Infosystem

¹ Die Studie befindet sich derzeit noch in der Auswertungsphase; eine größere Publikation zu diesem Thema ist noch für 2006 geplant.

² International Symposium on Music Information Retrieval.

DBIS (http://www.bibliothek.uni-regensburg.de/dbinfo/?bib_id=ub_r) wählt er zunächst eine geeignete bibliographische Datenbank (z. B. SCOPUS, <http://www.scopus.com>) aus, recherchiert dort und greift, falls die *Elektronische Zeitschriftenbibliothek* (EZB, <http://rzblx1.uni-regensburg.de/ezeit/>) Lizenzen nachweist, über einen *link resolver* auf Volltexte elektronischer Zeitschriftenartikel zu und lädt diese auf seinen Rechner; alternativ sucht er nach Pre- oder Postprints bei *Citeseer* oder auf Publikationsservern von Hochschulen. Nicht immer ist ihm dabei klar, ob er gerade nur in Metadaten oder auch in Volltexten recherchiert; auch die Kriterien für die Verfügbarkeit von Inhalten (Bibliotheken, Verlage, Informationsdienstleister als Betreiber) sind unübersichtlich. Immerhin ist er vom ursprünglichen Informationsbedürfnis über verschiedene Systeme (DBIS, SCOPUS, EZB, Verlagswebsite) bis zu einigen Volltexten vorgedrungen.

Die Menge der für den Nutzer verfügbarer digitaler Information hat mittlerweile ein kaum zu überschauendes Maß angenommen. Um sein Wissensmanagement zu optimieren, bedient er sich daher einer Desktopsuchmaschine (z. B. *Google Desktop* (<http://desktop.google.de>) oder *x-friend* (<http://xfriend.de/>)). Diese lässt ihn gezielt in seinen mittlerweile rund 500.000 Dateien suchen.³ Eine ansprechende Visualisierung seiner Dateiinhalte, die Zusammenfassung zu inhaltlich verwandten Klassen oder das Aussortieren älterer Dateiversionen vermisst er allerdings. Er hofft auf eine grundlegende „Reform“ des zugrunde liegenden Ordnungskonzeptes (hierarchisches Dateisystem), wird darauf aber sicher noch warten müssen (vgl. dazu das Scheitern von Microsoft, ein datenbankorientiertes Dateisystem einzuführen (Clark, 2006)). Er überlegt, ob er seine Dateien durch gezieltes manuelles Tagging vielleicht einfacher ordnen könnte (Macgregor & McCulloch, 2006), kapituliert aber vor der bereits aufgelaufenen Menge an Information. Weitergehende Lösungen, die sich der Technologien und Konzepte des *semantic web* bedienen und dedizierte Software für das *personal information management* (Bruce, 2005; Czerwinski et al., 2006) bereitstellen, befinden sich bestenfalls in einer experimentellen Phase.

Nachmittags ist etwas Zeit für einen Einkauf, er will ein elektronisches Haushaltsgerät online möglichst günstig bestellen. Die Vielfalt geeigneter Preisberatungsplattformen (z. B. *Froogle* (<http://froogle.google.de/>), *Kelkoo* (<http://www.kelkoo.de/>), *Prosuma* (<http://www.prosuma.de/>) oder *Itsbetter* (<http://www.itsbetter.de/>)) macht die Auswahl schwer. Auch ist ihm nicht klar, nach welchen Kriterien welche Händler auf den Plattformen vertreten sind. Neben inhaltlichen (d. h. hier: produktfunktionsbezogenen) Relevanzkriterien erschweren weitere Aspekte wie Preis, Verfügbarkeit, Lieferzeiten, Versicherungen oder Vertrauen in den Händler die Suche, die zudem medial sehr unterschiedlich unterstützt wird (je nach Plattform u. a. durch Bilder, Datenblätter, Kundenmeinungen und Erfahrungsberichte, Verkaufsratings).

Für eine Geburtstagsfeier will er eine Glückwunschkarte mit digitalem Bildmaterial gestalten, sucht zunächst bei

kommerziellen Agenturen (z. B. Corbis (<http://pro.corbis.com>), Comstock (<http://www.comstock.com/>) oder Photodisc (<http://www.photodisc.com/>)), stellt aber fest, dass die Bildsuche für nicht unmittelbar objektbezogene Suchkriterien keine Ergebnisse bringt („Abendstimmung“), während ihn die *folksonomies* (Marlow, Naaman, Boyd, & Davis, 2006) bei der Bildplattform Flickr (<http://www.flickr.com>) mit rund 1000 Treffern schneller zu einem passenden Ergebnis führen:

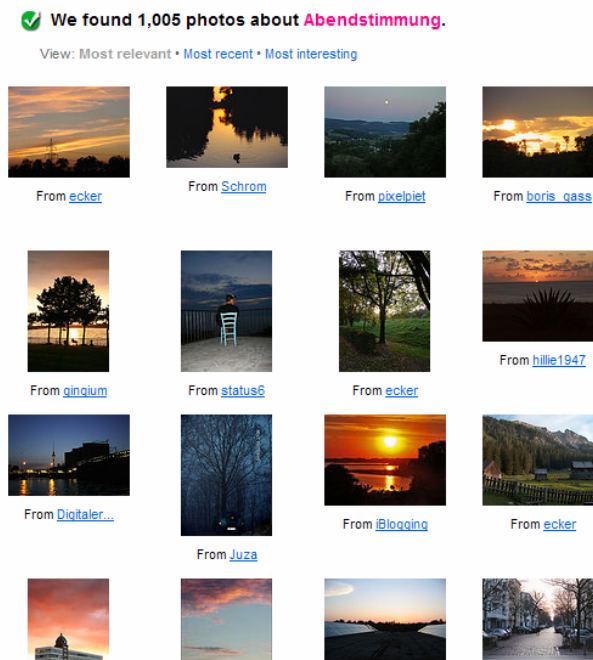


Abb. 1: Beispielrecherche „Abendstimmung“ bei Flickr (<http://www.flickr.com>, Juli 2006)

Am Abend sieht er fern, fremdwortreiche Ausführungen eines bekannten Meteorologen veranlassen ihn, zum Laptop zu greifen und nach diesen Begriffen mit einer Suchmaschine zu recherchieren, die ihn zu entsprechenden Einträgen in der Online-Enzyklopädie Wikipedia (<http://de.wikipedia.org>) leitet. Einfacher wäre ein direkter Zugriff auf das Web über einen MHP-fähigen Digitalfernsehdecoder (*multimedia home platform*, (European Telecommunications Standards Institute, 2003)) oder die Nutzung der Zusatzmaterialien einer interaktiven Fernsehsendung, die geringe Verfügbarkeit insbesondere in Deutschland lässt dies aber als Zukunftsszenario erscheinen (Commission of the European Communities, 2006). Steigende Zahl von Sendeangeboten und ein wenigstens partieller Wechsel vom programmgetriebenen Sendemodus (*broadcasting*) zu bedürfnisgetriebenen *on demand*-Systemen lassen weitere Anwendungsgebiete für IRS (insbesondere intelligente Suchagenten) bereits erkennen.

Später trifft er bei einem Klassentreffen auf viele Bekannte und Freunde, die er zum Teil seit vielen Jahren nicht gesehen hat. Selbstverständlich werden viele Bilder gemacht, neben eigenen lädt er sich wenige Tage später auch Kopien der Bilder aller Anwesenden von einer Online-Plattform und fügt sie seiner bereits auf einige tausend Bilder angewachsenen Sammlung hinzu. Dabei stellt er fest, dass für ihn für viele Personen der Name schon wieder entfallen ist. Die digitalen Bilddaten liefern zwar präzise Informationen über Verschlusszeit, Blende und Kameramodell (und Aufnahmezeitpunkt, vgl. Abb. 2), aber

³ Schätzung nach Angaben von *Google Desktop* für den Datenbestand des Autors; davon allein ca. 100.000 HTML-Dateien und ca. 150.000 Bilddateien.

nichts über Ort und abgebildete Person, individuelles Tagging scheidet schon mangels Wissens aus.

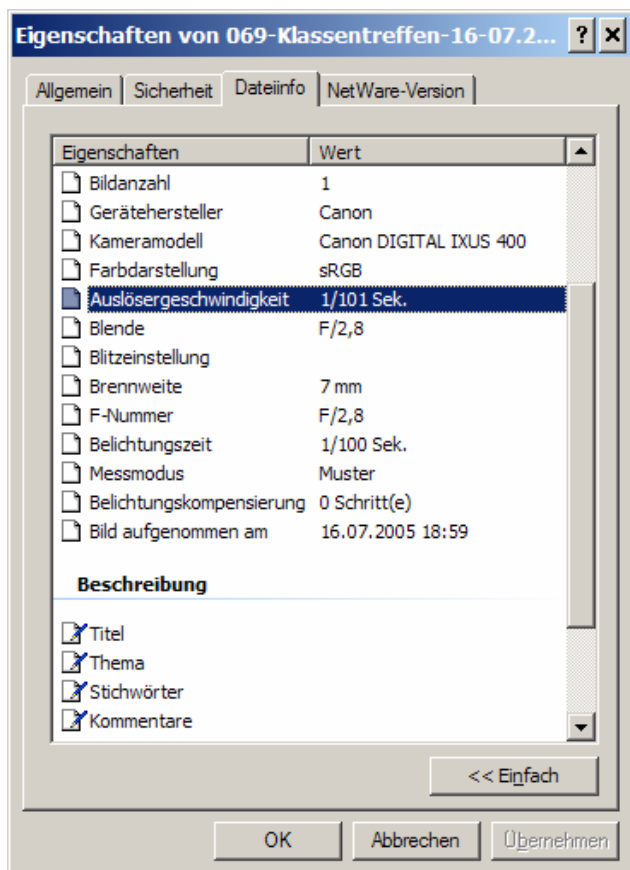


Abb. 2: Technische und (fehlende) inhaltliche Metadaten eines Digitalphotos

Nachts erneuert die Desktopsuchmaschine ihren Index, Suchagenten durchforschen das Netz nach weiteren relevanten Informationen (z. B. *alerting services* digitaler Bibliotheken, Einkaufsprofile etc.).

Die genannten Szenarien lassen sich für weitere Lebenslagen beinahe beliebig erweitern (z. B. Suche nach Erfahrungsberichten von Leidensgenossen bei Krankheit, Suche nach Kontakten und Partnern (Partnersuchbörsen, *social networking*-Plattformen wie *openBC* (<http://www.openbc.com>), sie dienten hier primär dem Zweck die Vielfältigkeit der Interaktion mit IRS zu illustrieren.

5 Beobachtungen und Thesen

Mit Blick auf die Information Retrieval Forschung ist zunächst zu fragen, ob die beobachteten Phänomene überhaupt Gegenstand der IR-Forschung sind oder sein sollen.

5.1 IR und Alltag als Gegenstand der Wissenschaft

Betrachtet man traditionelle Lehrbuchdefinitionen, so zeigt sich hier kein Widerspruch, da dort generisch alle relevanten Prozesse der Informationserschließung ohne Einschränkung auf bestimmte Informationsarten oder Medien angesprochen werden:

Information Retrieval (IR) is concerned with the representation, storage, organization, and accessing of information items. In principle, no restriction is placed on the type of item handled in information retrieval. (Salton & McGill, 1983, p. 1)

Information Retrieval (IR) deals with the representation, storage, organization of, and access to information items. The representation and organization of the information items should provide the user with easy access to the information in which he is interested. (Baeza-Yates & Ribeiro-Neto, 1999, p. 1)

Es erscheint legitim, die Phänomene des Information Retrieval *im Alltag* als Erweiterung der Forschungsperspektive des IR zu begreifen. Im historischen Vergleich wird man feststellen, dass die Frage, wie etwa jemand seine Bibliothek, Diasammlung oder private Korrespondenz organisiert und nutzt, von einem Nischenthema der Volkskunde, Kulturwissenschaft oder Literatursoziologie, zu einer auch technologisch interessanten Frage der Informationswissenschaft und Informatik geworden ist. Im Sinne der Generalhypothese von Ben Shneiderman („the new computing is about what users can do“, Shneiderman, 2002, p. 2) kann man hierin die Verwissenschaftlichung des Alltags erkennen – in der Informations- und Wissensgesellschaft sind IR-Systeme zum zentralen und alltäglichen Instrument geworden:

- Jeder ist Benutzer von sehr heterogenen IRS.
- Zunehmend werden *alle Medien* zum Gegenstand des IR: Nicht nur die Recherche an sich (Suchmaschinen-Paradigma), sondern auch die Aufbereitung, insbesondere die Erschließung von Information durch textuelle Metadaten wird für jedermann/-frau relevant.
- Nicht nur das WWW als globaler Datenspeicher ist dramatisch angewachsen, auch die Menge individuell verfügbarer digitaler Information wächst und über traditionelle Mediengrenzen hinweg (Text, Bild, Musik, Video). Extrapoliert man die schon heute erreichte Situation für einige Jahre oder gar Jahrzehnte, so ist leicht zuerkennen, dass neue Verfahren der Informationserschließung, -aufbereitung und -organisation und dringend erforderlich sind..
- Zeitliche, ökonomische, organisatorische Kriterien gehen in die Effektivitätsbewertung mit ein, in vielen Fällen wird man bei der Systembewertung kaum abstrahieren können („lieber ein kostenloses Bild bei Flickr als ein besseres, aber teures bei Corbis“).
- Phänomene der Medienkonvergenz in den digitalen Medien gewinnen an Bedeutung; die klassischen Rundfunkmedien werden in Kürze über die „traditionellen“ digitalen Datennetze (Internet, WWW) verfügbar sein, umgekehrt zeichnet sich (siehe oben) auch die Kopplung von WWW und digitalem Fernsehen ab (Theunert, 2002).⁴
- Die Potentiale der Social Software (Bächle, 2006; Möller, 2005; Sixtus, 2005) für die Informationerschließung sind zunächst nur phänomenologisch erkannt worden: Eine Reihe von Portalen erleichtern die gemeinsame Erschließung und Distribution unterschiedlicher Medien für sehr heterogene Nutzerkreise

⁴ Medienkonvergenz der Inhalte könnte dabei mit Medien-divergenz der Nutzungsgeräte und -szenarien einhergehen, wenn im Sinne des *ubiquitous computing* ganz unterschiedliche Endgeräte verfügbar werden.

(z. B. CiteULike (<http://www.citeulike.org>) im Bereich wissenschaftlicher Literatur, Flickr im Bereich Bildmedien).

- Als Rechercheparadigma ist dabei die begriffsbasierte, textuelle Suche immer noch das primäre Mittel der Anfrageformulierung, da z. B. automatische Bilderkennung (Santini, 2001) oder Videoanalyse (Feng, Siu, & Zhang, 2003) noch nicht flächendeckend verfügbar sind und innovative Visualisierungsformen als Interaktions- und Präsentationsparadigmen sich bisher nicht durchsetzen konnten (Arnold & Wolff, 2005).

Vor dem Hintergrund der Heterogenität von Nutzern und Systemen dürften sich die bekannten Probleme der IR-Forschung verschärfen, insbesondere im Bereich der Informationskompetenz – der viel diskutierte *digital divide* (Kizza, 2003; Shneiderman, 2002) bezieht sich mittlerweile (in den Industrienationen) weniger auf verfügbare *Hardware* oder den *Zugang* zu den Datennetzen als auf die Kompetenz im Umgang mit Informationssystemen (für 1998/99 noch anders: Wagner, Pischner, & Haisken-DeNew, 2002). Dazu gehören:

- Aspekte der sprachlichen Aufbereitung von Suchbegriffen (Vollformen, Eigennamen, Komposita).
- Kenntnisse von den Möglichkeiten der Anfrageformulierung (fehlendes oder falsches mentales Modell von der Retrievalfunktion).
- Die besonderen Probleme nicht-textueller Medien (Musik, Bild, Film etc.), insbesondere die Beschreibung durch textuelle Metadaten.

Über den „common user“ ist generell noch recht wenig bekannt, für die Frühphase der Suchmaschinennutzung ab 1995 konstatiert (Ceruzzi, 2003, p. 329) zwar: “Computer-savvy Internet users did not need a portal. They preferred brute-force search engines and were not afraid to construct complex searches using Boolean algebra to find what they wanted.” Diese Beobachtung steht in deutlichem Gegensatz zu empirischen Studien zum Rechercheverhalten und der Informationskompetenz von WWW-Nutzern (siehe oben Kap. 3).

Der alltäglichen und nicht berufsbedingten Nutzung des Internet und seiner Informationsdienste sind mittlerweile einige Studien gewidmet worden (Bakardjieva, 2005; Silverstone, 2005; Wellman & Haythornthwaite, 2002), der besonderen Problematik der Interaktion mit IRS schenken diese aber kaum Beachtung. Die stärkere Nutzerorientierung hat im IR-Bereich zwar schon seit längerem im *cognitive viewpoint* des IR ihren Niederschlag gefunden (Ellis, 1998; Ford, 2004; Ingwersen, 1999), bezieht sich aber ebenfalls auf Fachinformation und den akademischen Nutzer, also auf wissenschaftsnahe Kontexte. Auch ungeachtet dieser theoretischen und methodischen Perspektive liegt der Schwerpunkt der publizierten IR-Forschung sicher eher im Bereich der Weiterentwicklung von IR-Algorithmen bzw. der Technologieentwicklung.⁵

⁵ Als Indiz kann man hier die thematischen und methodischen Schwerpunkte der *Annual ACM Conference on Research and Development in Information Retrieval* heranziehen, die als internationale IR-Leittagung gelten kann, vgl. http://portal.acm.org/browse_dl.cfm?linked=1&part=series&idx=SERIES278&coll=ACM&dl=ACM&CFID=626045&CFTOKEN=83337271.

5.2 Fazit: Konsequenzen für die Forschung

Viele der in Kap. 4 angedeuteten Probleme der Informationssuche im Alltag lassen sich durch *technologische Weiterentwicklungen* abmildern und beheben. Dies kann man wiederum anhand eines praktischen Beispiels – hier: Die Metadatenerfassung in der Digitalphotographie leicht illustrieren:

- Die Kopplung von Photographie mit Spracherkennung, könnte die Erfassung von Metadaten (*tagging*) an der Quelle, d. h. in unmittelbarem Zusammenhang mit der Bildentstehung erlauben – eine Kamera oder ein Kamerahandy hat eher ein Mikrofon als eine Tastatur).
- Maschinelles Lernen und Mustererkennung können helfen, bekannte Personen auf Bildern – im Sinne von Shneidermans Schichtenmodell des persönlichen Umfelds sicher eine überschaubare Zahl (Shneiderman, 2002, p. 80ff) – automatisch zu erschließen.
- Mit Hilfe von RFID-Tags⁶ und einer Modellierung der Informationsfreigabe nach Bekanntheitsgrad ließen sich Metadaten von einer fotografierten Person leicht auf eine Kamera übertragen.
- Die Integration eines GPS-Empfängers⁷ in die Kamera, gekoppelt mit einem geographischen Informationssystem, würde die automatische Übernahme von Ortsinformation ermöglichen.

Weniger offensichtlich sind technologische Lösungen für das komplexere Problem der persönlichen Wissensorganisation (*personal information management*). Voraussetzung dafür wäre – und das ergibt eine zusätzliche Forschungsperspektive – ein geeignetes Modell unseres informationellen Alltags und unserer informationellen Umwelt. Ansatzpunkte kann man hierfür in der Informationskompetenz-Forschung finden, allerdings ist diese stark bibliotheks- und damit auch wissenschaftsorientiert (Eisenberg, Lowe, & Spitzer, 2004; Neely, 2002; Pickering Thomas, 1999). Neben der notwendigen theoretischen Modellbildung für das *ubiquitous information retrieval* im Alltag sind vor allem alters- und gruppenspezifische empirische Studien zur Informationskompetenz, typischen „alltäglichen“ Informationsbedürfnissen und zur Praxis des Umgangs mit IRS ein dringliches Desiderat. Das etwas plakative Fazit kann daher lauten: *Information Retrieval is for everybody, but we don't know much about anybody.*

Literatur

- Arnold, C., & Wolff, C. (2005). Evaluierung von Visualisierungsformaten bei der webbasierten Suche. In Forschungszentrum Jülich (Ed.), *Knowledge eXtended. Die Zusammenarbeit von Wissenschaftlern, Bibliothekaren und IT-Spezialisten* (pp. 275-286). Jülich: Forschungszentrum Jülich GmbH.
- Bächle, M. (2006). Social software. *Informatik-Spektrum*, 29(2), 121-124.
- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Harlow et al. / New York: Addison-Wesley / ACM Press.
- Bainbridge, D., Cunningham, S. J., & Downie, J. S. (2003). *Analysis of queries to a Wizard-of-Oz MIR system: Challenging assumptions*.

⁶ Radio Frequency Identification.

⁷ Global Positioning System.

- tions about what people really want. Paper presented at the Proceedings of the 4th International Conference on Music Information Retrieval, Baltimore, MD.
- Bakardjieva, M. (2005). *Internet Society. The Internet in Everyday Life*. London / Thousand Oaks, CA / New Delhi: Sage Publications.
- Bruce, H. (2005). Personal, anticipated information need. *Information Research*, 10(5).
- Bush, V. (1991). As We May Think (1945) In J. M. Nyce & P. Kahn (Eds.), *From Memex to Hypertext: Vannevar Bush and the Mind's Machine*. Boston: Academic Press.
- Byrd, D., & Crawford, T. (2002). Problems of Music Information Retrieval in the Real World. *Information Processing & Management*, 38.(2), 249-272.
- Ceruzzi, P. E. (2003). *A History of Modern Computing* (2nd ed.). Cambridge, MA / London: The MIT Press.
- Clark, Q. (2006). What's in Store. Update to the Update [Electronic Version]. *Microsoft Developer Network (MSDN). WinFS Team Blog*, 2006. Retrieved July 31, 2006 from <http://blogs.msdn.com/winfs/archive/2006/06/26/648075.aspx>.
- Commission of the European Communities. (2006). *Communication From The Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions on Reviewing the Interoperability of Digital Interactive Television Services Pursuant to Communication Com(2004) 541 Of 30 July 2004*. Brüssel: Commission of The European Communities.
- Cunningham, S. J., Jones, M., & Jones, S. (2004). *Organizing Digital Music for Use: An Examination of Personal Music Collections*. Paper presented at the Proceedings of the 5th International Conference on Music Information Retrieval, Barcelona.
- Czerwinski, M., Gage, D., W., Gemmell, J., Marshall, C., C., Pérez-Quñonesis, M. A., Skeels, M., M., et al. (2006). Digital memories in an era of ubiquitous computing and abundant storage. *Communications of the ACM*, 49(1), 44-50.
- Dominich, S. (2001). *Mathematical Foundations of Information Retrieval*. Dordrecht et al.: Kluwer Academic Publishing.
- Downie, J. S. (2004). A Sample of Music Information Retrieval Approaches. *Journal of the American Society for Information Science and Technology* 55(12), 1033-1036.
- Eisenberg, M. B., Lowe, C. A., & Spitzer, K. L. (2004). *Information Literacy. Essential Skills for the Information Age*. New York et al.: Macmillan Publishers.
- Ellis, D. (1998). Paradigms and research traditions in information retrieval research. *Information Services and Use*, 18(4), 225-241.
- European Telecommunications Standards Institute. (2003). *Digital Video Broadcasting (DVB). Globally Executable MHP (GEM) Specification 1.0.0*. (No. Dokument ETSI TS 102 819).
- Feng, D. D., Siu, W.-C., & Zhang, H.-J. (Eds.). (2003). *Multimedia Information Retrieval and Management - Technological Fundamentals and Applications*: Berlin: Springer Verlag.
- Ford, N. (2004). Modeling cognitive processes in information seeking: From popper to pask. *Journal of the American Society for Information Science and Technology*, 55(9), 769-782.
- Hammwöhner, R., & Wolff, C. (2006 (erscheint)). Wie funktioniert das Internet? In M. Fölling-Albers (Ed.), *Regensburger Universität für Kinder 2006*. Regensburg: pro Regensburg e.V.
- Hochholzer, R., & Wolff, C. (2005). *Informationskompetenz – status quo und Desiderate für die Forschung*. Regensburg: Universität Regensburg, Institut für Germanistik und Institut für Medien-, Informations- und Kulturwissenschaft.
- Ingwersen, P. (1999). Cognitive information retrieval. *Annual Review of Information Science and Technology*, 34, 3-52.
- Kizza, J. M. (2003). *Ethical and Social Issues in the Information Age* (2nd ed.). New York et al.: Springer.
- Lesk, M. (1995). *The Seven Ages of Information Retrieval*: International Federation of Library Associations and Institutions, Universal Dataflow and Telecommunications Core Activity (UDT).
- Lilley, D. B., & Trice, R. W. (1989). *A History of Information Science. 1945 - 1985*. San Diego, CA: Academic Press.
- Macgregor, G., & McCulloch, E. (2006). Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review*, 55(5), 291-300.
- Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). *Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead*. Paper presented at the WWW2006 Tagging Workshop Edinburgh.
- Möller, E. (2005). *Die heimliche Medienrevolution. Wie Weblogs, Wikis und freie Software die Welt verändern*. Hannover: Heise.
- Neely, T. Y. (2002). *Sociological and Psychological Aspects of Information Literacy in Higher Education*. Lanham/MD: Scarecrow Press.
- Pickering Thomas, N. (1999). *Information Literacy and Information Skills Instruction, Applying Research to Practice in the School Library Media Center*. Westport/CT: Libraries Unlimited.
- Salton, G., & McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. New York et al.: McGraw-Hill.
- Santini, S. (2001). *Exploratory Image Databases. Content-based Retrieval*. San Diego, CA: Academic Press.
- Shneiderman, B. (2002). *Leonardo's Laptop: Human Needs and the New Computing Technologies*. Cambridge, MA / London: The MIT Press.
- Silverstone, R. (Ed.). (2005). *Media, Technology and Everyday Life in Europe. From Information to Communication*. Aldershot / Burlington, VT: Ashgate.
- Sixtus, M. (2005). Das Web sind wir. *Technology Review*(Juli 2005), 44-52.
- Theunert, H. (Ed.). (2002). *Medienkonvergenz: Angebot und Nutzung. Eine Fachdiskussion veranstaltet von BLM und ZDF* (Vol. 70). München: Fischer.
- van Eimeren, B., & Frees, B. (2005). ARD/ZDF-Online-Studie 2005. Nach dem Boom: Größter Zuwachs in internetfernen Gruppen. *Media Perspektiven*(8/2005), 362-379.
- Wagner, G. G., Pischner, R., & Haisken-DeNew, J. P. (2002). The Changing Digital Divide in Germany. In B. Wellman & C. Haythornthwaite (Eds.), *The Internet in Everyday Life*. (pp. 164-185). Malden, MA / Oxford / Carlton: Blackwell Publishing.
- Wellman, B., & Haythornthwaite, C. (Eds.). (2002). *The Internet in Everyday Life*. Malden, MA / Oxford / Carlton: Blackwell Publishing.
- Wolff, C. (2000). Vergleichende Evaluierung von Such- und Metasuchmaschinen im World Wide Web. In G. Knorz & R. Kuhlen (Eds.), *Informationskompetenz – Basiskompetenz in der Informationsgesellschaft. Proc. 7. Intern. Symposium f. Informationswissenschaft, ISI 2000* (pp. 31-48). Konstanz: UVK.

Anmerkung

Alle WWW-Ressourcen, die in diesem Artikel erwähnt sind, wurden im Juli 2006 verifiziert.