



Navigation, Adaption und Information Retrieval: Informationstechnik im Dienst des Buchkaufs

Stephan Roppel

Die vorliegende Festschrift gab Anlass zu Gedanken, ob und in welchem Maße die Tätigkeit des Autors in der „Industrie“, nämlich als Leiter des Buch- und Mediengeschäftes eines großen amerikanischen Internetbuchhändlers, durch die beim hier Geehrten erlernte wissenschaftliche Methodik und die seinerzeit im Mittelpunkt stehenden Forschungsthemen beeinflusst und gefördert wurde.

Vor der Beantwortung dieser Frage lohnt es sich zum besseren Verständnis, kurz einen Blick auf den Hintergrund der angesprochenen praktischen Tätigkeit zu werfen.

1 Entwicklung und Stellenwert des Internetbuchhandels

Innerhalb weniger Jahre entstand mit dem Internet ein neuer Verbreitungsweg für Bücher, der mit seiner Neuartigkeit und rasanten Wachstumsgeschwindigkeit aus Sicht einer über Jahrhunderte gewachsenen und weitgehend mittelständisch geprägten Branche bedrohlich wirken musste. Dies umso mehr, als sich der Erfolg nicht nur in den USA, einem Flächenstaat mit teilweise lückenhaftem Versorgungsnetz für Bücher, sondern auch in Deutschland mit seiner vorbildlichen Buchhandelsstruktur einstellte, die mittels Barsortimentservice und auf Grund flächendeckender Versorgung durch eine Vielzahl regionaler Buchhandlungen, eine Übernachtslieferung eines Großteils der lieferbaren Titel garantiert.

Dabei haben erfreulicherweise kulturpessimistische Bedrohungsszenarien, die neben dem Verschwinden des Kulturträgers „Buchhändler um die Ecke“ auch gleich das Ende der Lesekultur und des Buchs als Medium selbst an die Wand malten, inzwischen weitgehend einer realistischeren, wenn nicht gar optimistischen Einschätzung Platz gemacht.

Zum einen sehen auch hoch gegriffene Annahmen den Umsatzanteil des Internetbuchhandels, ausgehend von einer Schätzung um 8 % für das Jahr 2003 (vgl. Buch- und Buchhandel in Zahlen 2002), langfristig nicht deutlich jenseits der 15 % des Gesamtmarktes.



Zum anderen zeigt sich, dass Recherchen in den Buchkatalogen der Internetbuchhändler – nicht unbedingt zu deren Freude – nicht selten durch einen stationären Kauf abgeschlossen werden, dass also das neue Onlinemedium in der „Offlinewelt“ Kaufimpulse setzt und oft gar neue Kundenschichten erschließt.

So deuten auch neuere Erkenntnisse (vgl. Amazon.de Benutzerbefragung) hinsichtlich der Bedarfsdeckung bei Internetbuchkäufern auf eine ausgewogene Nutzung verschiedener Vertriebskanäle hin.

Neben dieser wirtschaftlichen und v. a. absatzorientierten Betrachtung sollen auch zwei oft weniger beachtete, die Buchkultur fördernde Effekte des Internetbuchhandels nicht unerwähnt bleiben: so hat sich etwa der Buchkatalog von Amazon.de, der auf einer Kombination von bereits in der Branche vorhandenen Systemen (z.B. Barsortimentskataloge, Verzeichnis lieferbarer Bücher (VLB)) mit eigenen Mehrwertangeboten (z.B. Rezensionen, Verkaufsrang, Sortierverfahren) fußt, zu einer Art „Volksbibliografie“ entwickelt, die von Studenten zur Literatúrauswahl, von Buchhändlern für ihre eigene Einkaufsplanung und von einer breiten Schicht potentieller Leser und Käufer zur Vorabfilterung eigener Leseabsichten genutzt wird. Eine Tatsache, die neben dem primären Ziel des Buchverkaufs eine sekundäre Funktionsebene und damit auch Verantwortung als *Buchinformationsmedium* begründet.

Ein weiterer über den aktuellen Verkaufserfolg hinausreichender Effekt des Internetbuchhandels ist sein Beitrag zur langfristigen Sicherung eines breiten Titelangebots – ein Ziel, dessen Erreichung üblicherweise als Hauptargument für den Erhalt der Buchpreisbindung verwendet wird. Die Breite des Katalogs (mehr oder weniger identisch mit dem Angebot lieferbarer Bücher, bzw. durch antiquarisches und Gebrauchtbuchangebot noch darüber hinausgehend) sowie Such- und Empfehlungstechniken bedingen eine überproportionale Präsentation der sog. *Backlist*. Es werden also diejenigen Titel, die durch Verlage noch angeboten werden, ohne dass für sie noch Marketing- und Präsentationsaufwand zur Verfügung steht, gegenüber Neuerscheinungen und Bestsellern relativ gestärkt. Diese die Vielfalt stützende Funktion ist umso bedeutender, als im Gesamtbuchhandel zwar die Zahl der Neuerscheinungen mit 80-90.000 pro Jahr (vgl. Buch- und Buchhandel in Zahlen 2002) konstant bleibt, sich aber die tatsächlichen Absatzzahlen kontinuierlich auf immer weniger Bestseller konzentrieren.

2 Warum sich im Internet so gut nach Büchern suchen lässt

Die auch aus wissenschaftlicher Sicht interessante Frage ist jedoch, warum der Handel mit Büchern im Internet so schnell so erfolgreich werden konnte. Oder anders gefragt: durch welche neuartigen Angebote und Serviceaspekte trägt das Verkaufsmedium *Internet* im allgemeinen und die spezifischen Eigenheiten des Produktsegments *Bücher* dazu bei, dass gerade der Internethandel als neue Hoffnung gilt, für die ansonsten eher als bedroht empfundene Bedeutung und den Fortbestand des Buchs? Eines Mediums, das sich im verschärften Wettbewerb mit konkurrierenden Medien- und Freizeitangeboten bei schrumpfendem Zeitbudget und abnehmender Lesekompetenz immer schlechter behaupten kann (vgl. GFK Consumer Panel 2003).

Hierzu ist es hilfreich, sich die medienspezifischen Vorteile für den potentiellen Buchkäufer und Leser vor Augen zu halten, wobei die Betrachtungen im Folgenden jeweils von einer Beschreibung der Erfolgsfaktoren an der Oberfläche ausgehen und anschließend Rückgriff auf wissenschaftliche Kategorien und Terminologie zu ihrer Einordnung und Bewertung genommen wird.

Aus der Sicht des Kunden gibt es drei wesentliche Erfolgsfaktoren, die den Internethandel gegenüber dem stationären Handel auszeichnen – auf die vice versa natürlich vorhandenen Stärken der klassischen Buchhandlung, die die oben angesprochene Koexistenz garantieren, möchte ich hier im Sinne des Themas nicht weiter eingehen.

2.1 Convenience

Am offenkundigsten und auch in der öffentlichen Diskussion am präsentesten sind die in der Marketingsprache unter *convenience* subsumierten Vorteile des E-Commerce: sie reichen von der direkten Verfügbarkeit des riesigen elektronischen Katalogs und der direkten Bestelleingabe durch den Benutzer, über die Verfügbarkeit von Querverweisen und der virtuell unbegrenzten Lagerkapazität, bis hin zu Zeit- und Kostenersparnis durch Bestellung von zu Hause ohne Ladenschluss und dem stets verfügbaren Kundendienst.

Lässt man einmal die logistischen und im engeren Sinne kundendienstbezogenen Faktoren beiseite, wird deutlich, dass Internetbuchhandel in seiner Domäne ein praktisches Lösungsbündel für grundlegende wissenschaftliche Fragestellungen bezüglich der Nutzerfreundlichkeit von Schnittstellen zu textbasierten Systemen darstellt: Navigation, Visualisierung und

Lesbarkeit von Information in komplexen Systemen stehen dabei im Mittelpunkt des Interesses.

Zwar ist das World Wide Web mit seiner universalen Themenvielfalt selbst Gegenstand der Forschung zum Thema semantische Netze (vgl. Berners-Lee 1998), jedoch kann die Domäne *Bücher* auf eine wesentlich fundiertere und formal durchgebildete Systematik der Strukturierung aufbauen. Die Abbildung von hunderttausenden von Büchern in eine Taxonomie ist der prototypische Anwendungsfall abstrakter (hierarchischer) Begriffssysteme, mit dem zusätzlichen Problem der Mehrfachzugehörigkeit einzelner Elemente zu Oberklassen. Die ebenfalls universale Themenvielfalt der Domäne Buch bedingt größtmögliche Komplexität, die entsprechende Komplexitätsreduktion durch Abbildung in eine hierarchische Datenstruktur hat jahrhundertlange Tradition.

Interessanterweise existieren im Markt parallel und relativ wenig sich gegenseitig beeinflussend zwei Systeme: im Buchhandel die Gruppierung nach thematischen Warengruppen (vgl. Rautenberg 2003:530), in den Barsortimentskatalogen auch angereichert durch eine eher flache Schlagwortsystematik, in der Bibliothekswissenschaft und -praxis bibliothekarische Kataloge (vgl. Rautenberg 2003:65) mit differenzierterer und kontrollierter Sachgruppenanalyse und Verschlagwortung.

Beide Systeme werden durch den Buchhandel im Internet eingesetzt: eine dem klassischen Handel entlehnte Warengruppensystematik tritt dem Kunden an der Oberfläche entgegen und steuert die Navigation (vgl. Abb. 1), die andere, bibliothekarisch fundierte Klassifikation und Verschlagwortung, bleibt dem Nutzer verborgen, steuert aber Suchfunktion und Personalisierungsangebote.

Die für den „Weg zum Buch“ notwendige Darstellung eines komplexen hierarchischen Systems in einer graphischen Benutzeroberfläche und das *Browsing* in einem solchen Onlinekatalog stellen Interaktions- und Navigationsherausforderungen dar, die Anfang der neunziger Jahre einen relativ neuen Zweig der Forschung bildeten (vgl. Roppel 1998). In der Rückschau erhielt die Beschäftigung mit den interaktiven Aspekten thematisch komplexer Hierarchien und semantischer Netze, also Hypertextstrukturen und Navigation, mit der schnellen Verbreitung des WWW und des Internetbuchhandels erst einen breiten praktischen Anwendungsbezug.



Stöbern

Unser Angebot nach Sachgebieten geordnet -- mit den neuesten Empfehlungen der Redaktion.

- **Antiquarische Bücher**
Längst vergriffene Bücher und wertvolle Raritäten: Hier finden Sie Bücher, die zwischen 1945 und 1971 erschienen sind.
- **Belletristik**
Romane, Erzählungen, Essays & Reportagen, Lyrik, Dramatik, Humor, Märchen ...
- **Börse & Geld**
Börse, Aktien, Geldanlage, Immobilien, Steuern, Versicherungen, Sparen ...
- **Business & Karriere**
Management, Marketing, E-Business, Wirtschaft, Job, Karriere ...
- **Computer & Internet**
Betriebssysteme, Programmierung, Cyberkultur, Hardware, Internet, Software ...
- **Erotik**
Sex-Ratgeber, Romane & Erzählungen, Fotobände, Comics & Mangas, Kalender ...
- **Fachbücher**
Biowissenschaften, Chemie, Geowissenschaften, Ingenieurwissenschaften, Mathematik, Medizin, Physik & Astronomie, Psychologie, Recht ...
- **Film, Kultur & Comics**
Film & Fernsehen, Musik, Malerei & Skulptur, Comics & Mangas, Fotografie, Architektur, Design, Medien, Theater, Popkultur ...
- **Gebrauchte Bücher**
Ob aktuelle Bestseller oder vergriffene Bücher: in unserem riesigen Angebot finden Sie gebrauchte Bücher günstiger!
- **Hörbücher**
Hörspiele, Lesungen, seltene Aufnahmen und Live-Mitschnitte aus allen ...
- **Kochen & Lifestyle**
Essen, Trinken, Wohnen, Garten, Mode, Freizeit, Haustiere, Wellness ...
- **Krimis & Thriller**
Historische Krimis, Noir, Psycho-, Polit- und Justizthriller, Regionalkrimis, Landhauskrimis, Britische Detektive ...
- **Lernen & Nachschlagen**
Allgemeinbildung, Einführungen, Kataloge, Lernhilfen, Lexika, Schule & Abitur, Sprachen, Wissenskanon ...
- **Musiknoten**
Songbooks, Partituren, Unterrichtswerke, Pop, Rock, Jazz, Klassik ...
- **Naturwissenschaften & Technik**
Astronomie, Medizin, Physik, Biologie, Mathematik, Ökologie, Chemie, Ingenieurwissenschaften ...
- **Politik, Biografien & Geschichte**
Porträts & Erinnerungen, Epochen, Historie & Zeitgeschichte, Gesellschaft & aktuelle Debatten, Umwelt & Verkehr ...
- **Preis-Hits**
Viel Buch für wenig Geld: Belletristik, Lifestyle, Sachbücher und mehr.
- **Ratgeber**
Psychologie, Lebenshilfe, Gesundheit, Partnerschaft, Erotik, Familie, Recht ...
- **Reise & Sport**
Abenteurer & Lesereisen, Bildbände, Camping, Gourmetguides, Karten, Reise- & Sprachführer, Sport & Fitness, Survival ...
- **Religion & Esoterik**
Christentum, Buddhismus, Hinduismus, Islam, Judentum, Konfuzianismus, Taoismus, Götter, Esoterik, Astrologie ...
- **Science Fiction, Fantasy & Horror**

Abb. 1: Warengruppen als Basis der Navigation

Betrachtet man das konkrete „Lösungsbündel“ bei Amazon.de hinsichtlich Darstellung, Navigation und Browsing zeigt sich, dass die Teillösungen nicht über den aktuellen, und bereits etablierten *state of the art* der Benutzerschnittstellentechniken und Hierarchiedarstellungen hinausgehen. So fehlen z.B. Visualisierungen von Beziehungen zwischen einzelnen Büchern oder Buchsegmenten, die über die grafischen Elemente von Standard Windows-Elementen hinausgehen. Ehrgeizigeren Visualisierungs- und Navigationslösungen sind natürlich auch durch die Notwendigkeit, gleichzeitig den Bedienungsstandards des Internets und den technischen Voraussetzungen vieler Millionen von Kunden gerecht zu werden, Grenzen gesetzt. Die angebotene Kombination der Teillösungen ist jedoch in hohem Maße zweckmäßig und durch Auswertung einer Vielzahl von empirischen Befunden kontinuierlich optimiert.

2.2 Personalisierung

Weniger deutlich an der Oberfläche erkennbar sind die oft unter dem Schlagwort *One to One Marketing* oder *Personalisierung* zusammengefassten Techniken zur individuellen Ansprache des Kunden: diese ermöglichen, gestützt auf eine breite und gleichzeitig differenzierte Datenbasis, grundsätzlich neue und in noch nicht absehbarem Maß verfeinerbare Anpassung von Angeboten an einzelne Kunden („...personalisation allows the relationship with customers on the Internet to migrate from anonymous mass marketing and sales to

„one to one’ marketing”, Kobsa 2001:54). Diese Techniken stellen eines der faszinierendsten und sicherlich noch in den Kinderschuhen steckenden Innovationspotentiale für die Anwendungsdomäne *Buchhandel* dar: sie basieren auf der Auswertung und Assoziation verschiedener Datenquellen: individuelle Kauf-, Such- und Navigationsentscheidungen, die Assoziierung von Produkten mit der entsprechenden Kaufhistorie anderer Kunden (sog. *collaborative filtering*), sowie der Fähigkeit zur Ansprache nach flexiblen Kriterien selektierbarer Kundensegmente. (Für einen Überblick über die Bandbreite im E-Commerce verwendeter Empfehlungsalgorithmen und die konkret bei Amazon genutzten Techniken vgl. Linden et al. 2003).

Stephans Shop
Name bearbeiten

Hallo, Dr. Stephan Roppel. Klicken Sie [hier](#) für persönliche Empfehlungen in [Bücher](#), [Musik](#), [DVD](#), [Video](#), [Software](#) und [Videogames](#).

Ihre persönliche Seite

Wandern auf Sardinien
von Andreas Stieglitz
Preis: **EUR 12,00**

Kurzbeschreibung
Wildes Bergland, herrliche Küsten - das Wanderparadies Sardinien ist vor allem im Frühjahr und Herbst ein lohnendes Ziel für alle, die... [Mehr dazu](#) | [\(Warum wurde mir das empfohlen?\)](#)

► [Mehr dazu auf Ihrer persönlichen Seite](#)

Ihre Empfehlungen

Blitzeis
von Peter Stamm
Preis: **EUR 7,00**

Kurzbeschreibung
Schauplätze dieser Erzählungen sind Spanish Harlem in New York, ein Dorf im schweizerischen Thurgau und eine Lungenklinik nahe einer... [Mehr dazu](#) | [\(Warum wurde mir das empfohlen?\)](#)

- 📖 [Ungefähre Landschaft](#), von Peter Stamm [\(Warum?\)](#)
- 📖 [Wie interpretiert man einen Roman](#), von Hans-Dieter Gelfert [\(Warum?\)](#)
- 📖 [In fremden Gärten](#), von Peter Stamm [\(Warum?\)](#)

► [Mehr dazu in Ihren persönlichen Empfehlungen](#)

NEU FÜR SIE

Hallo, Dr. Stephan Roppel.
Das gibt es heute neu für Sie (Wenn Sie nicht Dr. Stephan Roppel sind, [klicken Sie hier](#).)

Ihre Neuerscheinungen


[Mein Leben](#)

Weitere Kategorien

- 📖 [Politik, Biografien & Geschichte](#)
- 📖 [Kinder- & Jugendbücher](#)
- 📖 [Lernen & Nachschlagen](#)
- 📖 [Kinder & Familie](#)
- 📖 [Film, Kultur & Comics](#)

► [Weitere Neuerscheinungen](#)

Abb. 2: Seite mit persönlichen Empfehlungen auf Basis von historischen Daten der gesamten Kundenbasis

Als illustrierendes Beispiel soll die bei amazon.de verfügbare Funktion dienen, die eine für jeden Besucher der Website individuell zusammengestellte Buch und Produktliste erstellt (vgl. Abb. 2): diese enthält z.B. als „verwandt“ definierte Bücher, wobei Verwandtschaftsgrad bzw. „Ähnlichkeit“ über die Analyse gleichzeitig bestellter unterschiedlicher Produkte durch die gesamte Kundenbasis definiert ist. Die Qualität solcher Angebotsformen, gemessen durch sog. *click-through* und *conversion* Analysen, die die Annahmequote der

Systemvorschläge durch Kunden bewerten, ist deutlich höher als diejenige menschlicher, „redaktioneller“ Vorschläge oder Bestsellerlisten: “The ... effectiveness vastly exceeds those of untargeted content such as banner advertisements and top-seller lists” (Linden et al. 2003:76). Interessanterweise sind diese Ergebnisse stark abhängig vom jeweiligen Produktsegment bzw. dessen Auswahlbreite: Buchempfehlungen funktionieren z.B. deutlich besser als solche für Filme auf DVD. Die Empfehlungsqualität kann durch Kundenfeedback mit Bewertung einzelner Vorschläge weiter verbessert werden, ebenfalls können Vorschläge, je nach individuellen und kulturellen Präferenzen, automatisch akzeptiert oder aber nur auf expliziten Kundenwunsch ein- und ausgeschaltet werden.

Ein verwandter, unter dem Schlagwort „Entdecken“ gefasster Vorteil besteht darin, dass durch die vielfältigen Vorschlags-, Such-, und Vernetzungsverfahren auch solche Bücher sozusagen auf Augenhöhe des Kunden gelangen, deren Präsentation oder Lagerhaltung beim traditionellen Buchhandel aus Platz und Aufwandsgründen schon längst nicht mehr möglich war.

Auch hier kann man aus wissenschaftlicher Sicht wieder von einem Lösungsbündel für eine zentrale Fragestellung der Forschung zur Benutzerfreundlichkeit interaktiver Systeme sprechen: die Frage nach der Machbarkeit und Zweckmäßigkeit systemseitiger *Adaption* an bestimmte Charakteristika des Anwenders, die im Rahmen der Benutzermodellforschung gestellt wird (für einen Überblick über den aktuellen Stand der Forschung vgl. Kobsa 2001).

Die Basis der Adaption im Internetbuchhandel bildet ein vielschichtiges Modell über Benutzerinteressen, das sich aus einer Vielzahl unterschiedlicher Quellen speist (explizite Interessensbekundung von Kunden durch Kauf, Ableitung von Annahmen durch Auswertung von Käufen anderer Kunden etc.) und eine Vielzahl von nahe liegenden Regeln heranzieht, um Interessen zu prognostizieren („Kunden, die identische Produkte kaufen, interessieren sich für gleiches“, „Kunden die Produkte mit ähnlichen Merkmalen kaufen, interessieren sich für gleiches“, „ein Kunde der Bücher eines Autors liest, interessiert sich auch für andere Bücher des Autors“ etc.).

Eine der kritischen Fragen des Benutzermodellbaus, nämlich wie aufwändig und zuverlässig die Erhebung solcher Daten ist, wird durch die Domäne *Einkaufen* (Kunden bringen durch Kauf eines Buches zuverlässig und ohne Zeitverlust ihr Interesse zum Ausdruck) praktisch „umsonst“ erledigt.

Auch andere kritische Aspekte, nämlich die der Qualität der auf Modellinhalten basierenden Anpassungsleistungen, werden pragmatisch gelöst

und durch die Integration in eine ohnehin auf Vorschlägen basierende Umgebung entschärft: Persönliche Vorschläge verdrängen nicht sonst standardisiertes oder erwartetes, sondern ergänzen von Redakteuren ausgewählte Angebote. Die Listen personalisierter „adaptiv generierter“ Vorschläge können, müssen aber nicht angesehen werden.

Zu dieser in der einschlägigen Literatur geforderten Unaufdringlichkeit von adaptiven Vorschlägen kommt die in der Forschung ebenfalls als wichtig akzeptierte Möglichkeit der Anpassbarkeit der Datengrundlage durch den Benutzer selbst, indem er Vorschläge differenziert bewerten und damit das eigene Benutzermodell steuern kann. Typische regelkonforme „Fehlvorschläge“, die z.B. auf Basis der Auswertung von Geschenkkäufen entstehen und mit eigenen Interessen nicht übereinstimmen, können so vermieden werden.

Vergleicht man den Grad der Innovation der eben diskutierten Adaptionsysteme im Internetbuchhandel mit den oben angesprochenen Navigationslösungen, so wird deutlich, dass hier wirklich Meilensteine der Entwicklung gesetzt wurden und werden. Auswertungs- und Vorschlagsalgorithmen betreten mit jeder Weiterentwicklung Neuland, gleichzeitig lässt die Breite ihrer Anwendung Testsysteme der Forschung weit hinter sich und führt zu jeweils empirisch validierten Verbesserungen, um die theoretisch bekannten Klippen adaptiver Systeme praktisch zu umgehen.

2.3 Die Suchfunktion

Als letztes soll diejenige Zugangsmöglichkeit zum Buch im Internet angesprochen werden, die für viele Aufgaben die leistungsfähigere Alternative zum Navigieren und Browsen darstellt: die Suchfunktion.

Die für Information Retrieval-Systeme und die sie betreffende Forschung (vgl. für einen Überblick Baeza-Yates 1999) grundlegende Unterscheidung, ob der Volltext oder lediglich sog. Metadaten (vgl. Rautenberg 2003:355f.) die Basis des Suchindex bilden, war bisher für die Anwendungsdomäne „Buchkataloge“ eindeutig: natürlich konnten nur die bibliografischen Angaben (Autor, Verlag, Titel, etc.) und evtl. vorhanden Schlagworte verwendet werden. Volltexte einer nennenswerten Anzahl bzw. repräsentativen Auswahl von Büchern waren weder im Markt noch in Bibliothekssystemen elektronisch als Datenbank vorhanden. Schon die Nutzung der begrenzten Metadaten-Suchbasis und ihre Kombination mit Verkaufsrankinformationen zur automatischen Filterung und Sortierung von Suchergebnissen brachte gegenüber herkömmlichen Buchdatenbanken einen deutlichen Funktionsvorsprung, zumindest für die große Mehrzahl der Kunden. Ihre Ergänzung durch benutzerseitige Sortier- (nach Preis, nach Verkaufsrang etc.) und Erweiterungsfunkti-

onen („more like this“) bildet einen weitergehenden Nutzensvorteil in Richtung eines sog. *Information Self-Service*.

2.4 Search Inside the Book

Die im Jahr 2003 bei amazon.com eingeführte Funktion „Search Inside the Book“, welche bereits zum Start für mehr als 120.000 Bücher den gesamten Text und die vorhandenen Metadaten als Suchbasis nutzt und gleichzeitig die – mit wirksamen Copyright-Schutzmassnahmen eingeschränkte – Faksimileanzeige der die Suchbegriffe enthaltenden Buchseiten ermöglicht, greift dagegen im Anspruch und auch in der ersten Umsetzung weit über bisher bekannte Anwendungsbeispiele von Information Retrieval-Technologie hinaus.

Nichts weniger als die Schaffung einer universellen digitalen Buch-Suchmaschine kann hier als Vision gelten, wie es z.B. Wolf 2003 mit der Überschrift seines Aufsatzes „The Great Library of Amazonia“ zum Ausdruck bringt. Wenn man, ausgehend von der inzwischen immensen Verbreitung und den Recherchemöglichkeiten von Internetsuchmaschinen, sich vergegenwärtigt, dass das eigentliche Wissen dieser Welt v. a. in Büchern enthalten ist, wird die Faszination dieses Projektes, jenseits seines unmittelbaren kommerziellen Reizes, deutlich.

Sicherlich stellt die Verwendung neuester Technologie inhaltserschliessender Algorithmen und die Integration weiterer Informationsquellen zur Relevanzbestimmung gefundener Bücher einen eindrucksvollen Schritt in der Anwendung von Information Retrieval-Technologie dar. Jedoch liegt das eigentlich Revolutionäre von „Search Inside the Book“ weniger in seinen Algorithmen, als in dem Mut und der Entschlossenheit, ein solches System – mit seinem Anspruch, die bestehenden Grenzen für die Suche in der Gesamtheit des verfügbaren Buchangebots grundsätzlich zu erweitern – überhaupt anzupacken.

All 63941 results for information retrieval :

Sort by:

1.  **Modern Information Retrieval**
by Ricardo Baeza-Yates (Author), et al (Paperback)
Avg. Customer Rating: ★★★★★



Usually ships in 24 hours
List Price: \$50.00
[Buy new:](#) **\$50.00**

[Used & new](#) from **\$39.40**

2.  **Information Storage and Retrieval**
by Robert R. Korfhage (Author) (Hardcover)
Avg. Customer Rating: ★★★★★
Other Editions: [Hardcover](#)



Usually ships in 24 hours
List Price: \$75.00
[Buy new:](#) **\$75.00**

[Used & new](#) from **\$58.67**

NEW! • Excerpt from [page 1](#) "... the design and use of a system to handle complex **information** tasks efficiently is the work of a specialist. This book addresses **information** storage and **information retrieval**-two sides of a coin: If a person is to be ..."
▶ See more [reference to information retrieval](#) in this book.

3.  **Text Information Retrieval Systems**
by Charles Meadow (Author), et al (Hardcover)
Other Editions: [Hardcover](#) - March 1992 | [See all \(2\)](#)



Usually ships in 11 days
List Price: \$73.95
[Buy new:](#) **\$73.95**

NEW! • Excerpt from [page 1](#) "... Introduction 1.1 What Is **Information**? To know what **information retrieval** is, we must first know what infor- mation is. There ..."
▶ See more [reference to information retrieval](#) in this book.

Abb. 3: Ergebnis von Search Inside the Book mit Direktzugriff auf Fundstellen

Zum Hintergrund: In der Verlagsbranche ist die mangelnde Verfügbarkeit und mehr als heterogene Datenlage der elektronischen Manuskripte für die meisten Werke seit Jahren als Problem erkannt. Eine Vielzahl von Projekten zur medienunabhängigen Inhaltsspeicherung in sog. Verlagsdatenbanken, auf Basis von Dokumentbeschreibungssprachen wie SGML und XML (vgl. Rautenberg 2003:471) mit dem Ziel, die Fähigkeit zur Publikation neuer inhaltlicher Zusammenstellungen und in unterschiedlichen elektronischen und papiergebundenen Medien zu erhöhen, haben in der Regel nicht zu einer grundlegenden Verbesserung der Datenqualität und -verfügbarkeit geführt.

Statt nun dieses Datenproblem als unüberwindliche praktische Hürde zu verstehen, wurde von Seiten Amazon eine riesige Buchdigitalisierungslogistik in Gang gesetzt. Diese nimmt als Input physische Bücher und produziert als Output Millionen von bildschirmlesbaren PDF-Seiten sowie einen Suchindex von bisher ungekannter Größe. Nach dem Scannen aller Seiten einzelner Bücher werden diese durch die Technik des *Optical Character Recognition* (OCR) in maschinenlesbaren Code übertragen und im nächsten Schritt einem Indexierungsprozess unterworfen, der Textinhalte als mögliche Suchbegriffe speichert und zugänglich macht (vgl. Wolf 2003).

Natürlich entsteht damit keine Lösung für die oben angesprochenen verlegerischen Ziele elektronischer Datenverfügbarkeit, für die elektronische Durchsuchbarkeit von Büchern war jedoch die ausreichende Basis geschaffen. Einer weiteren grundsätzlichen Herausforderung der Buchwirtschaft, dem Schutz, der Klärung und der Vergütung des Urheberrechts an Texten, die mit deren zunehmender elektronischer Verfügbarkeit eine neue Dimension erhält, wurde mit einem pragmatischen Sicherheitskonzept begegnet. Bedenken von Verlagen und Autoren hinsichtlich des evtl. möglichen Copyright-Missbrauchs auf Basis jetzt am Bildschirm lesbarer Buchseiten, wurde so begegnet, dass der Aufwand für missbräuchliche Verwendung (also Kopieren, Drucken, Downloaden) so wirkungsvoll erhöht wurde, dass er über demjenigen für das Kopieren des physischen Buches liegt und damit praktisch gesehen kein Problem mehr darstellt.

Stellt man abschließend wiederum die Frage nach dem Innovationsgrad von *Search Inside the Book*, kann man, sowohl hinsichtlich der Breite des Anwendungskontextes als auch hinsichtlich des für die Leser entstandenen Mehrwertes von einem Meilenstein sprechen.

Sollte es darüber hinaus gelingen, neben aktuell lieferbaren Verlagsprogrammen auch nennenswerte Teile von Bibliotheksbeständen in ein digitales Suchprogramm zu integrieren (und hier bestünde die Herausforderung nicht zuletzt in der Klärung, ob und durch wen überhaupt noch Urheberrechte gehalten werden) könnte eine noch viel weitergehende Umsetzung in Richtung einer digitalen Bibliothek Wirklichkeit werden (vgl. Gutenberg 2003 zum in diese Richtung weisenden Gutenberg-Projekt für den Deutschen Sprachraum).

In jedem Fall darf man z.B. hinsichtlich in Zukunft zu erschließender Buchsegmente und Sprachen sowie für neu sich ergebende Anwendungsfelder äußerst spannende Betätigungsfelder für die Forschung in Informations-, Medien-, Buch- und nicht zuletzt Rechtswissenschaft erwarten.

3 Fazit

Auf den ersten Blick ergibt sich, zurückkommend auf die Ausgangsfrage, inwiefern zwischen ursprünglichen wissenschaftlichen Fragestellungen des Autors und der aktuellen Tätigkeit Überschneidungen und wechselseitige Beeinflussung bestehen, ein klar positiver Befund. Nimmt man etwa Titel und Inhalt von Roppel 1998 „Visualisierung und Adaption. Techniken zur Verbesserung der Interaktion mit hierarchisch strukturierter Information“, als Ausgangspunkt und ergänzt eine Reihe weiterer Themen aus dem angrenzenden Forschungskontext, ergibt sich eine erstaunlich hohe Entsprechung zu den Kernfragen der Produktentwicklung des eigenen Arbeitsumfeldes Internetbuchhandel.

Neben den inhaltlichen Übereinstimmungen ist nicht zuletzt eine Entsprechung im Methodischen augenscheinlich: für eines der Kernpostulate von Jürgen Krause bei der kontextabhängigen Systementwicklung – „Verbesserung durch Rapid Prototyping“ (vgl. Marx & Schudnagis 1997:43-58) – gibt es in einem so komplexen und ständig im „Live-Betrieb“ befindlichen System wie Amazon gar keine Alternative. Fortschritte etwa im Navigationsdesign, in der Suchergebnispräsentation und in den Adaptionssystemen werden ausnahmslos durch die Messung von die Kundenzufriedenheit betreffenden Kennzahlen und ihre Auswertung in statistischen Testverfahren abgesichert.

Dass die weitere Entwicklung der in diesem Beitrag diskutierten Aspekte des Internetbuchhandels – Navigation, Adaption und Information Retrieval – ein lohnendes Thema zukünftiger wissenschaftlicher Betätigung darstellt, bedarf dabei hoffentlich keiner besonderen Erwähnung mehr!

4 Literaturverzeichnis

- Amazon.de Benutzerbefragung. Interne Studie zur Bedarfsdeckung (unveröffentlicht). Amazon.de, München, 2001.
- Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier (1999). Modern Information Retrieval. Harlow et al: Addison-Wesley; New York: ACM Press.
- Berners-Lee, Tim. Semantic Web Roadmap. 1998. <http://www.w3.org/DesignIssues/Semantic.html> [Zugriff September 2004].
- Buch und Buchhandel in Zahlen. Hrsg. Vom Börsenverein des Deutschen Buchhandels. Frankfurt a.M. 2002.
- GfK Consumer Panel „Medien“. Nürnberg, 2003.
- Gutenberg-DE, Edition 7 CD-ROM Projekt. Hille und Partner, Hamburg, 2003.

- Kobsa, Alfred (2001). "Generic User Modeling Systems." In: *User Modeling and User Adapted Interaction* (11) (2001), 49-63.
- Linden, Greg, Smith, Brent, York, Jeremy (2003). "Amazon.com Recommendations. Item to Item collaborative Filtering." In: *IEEE Internet Computing*, (1)(2003), 76-80.
- Marx, Jutta, Schudnagis Monika (1997). „Überblick über die Wing-IIR Benutzertests und methodisches Vorgehen.“ In: Krause, J, Womser-Hacker Christa (ed.) (1997). *Vages Information Retrieval und graphische Benutzungsoberflächen*. Konstanz, 43-58.
- Rautenberg, Ursula (ed.) (2003). *Reclams Sachlexikon des Buches*. Stuttgart: Reclam.
- Roppel, Stephan (1998). *Visualisierung und Adaption. Techniken zur Verbesserung der Interaktion mit hierarchisch strukturierter Information*. Konstanz.
- Wolf, Gary (2003). "The great library of Amazonia." In: *Wired Magazine* (12)(2003).