Digitale Bibliotheken

Klaus Tochtermann

In jüngster Zeit findet man in wissenschaftlichen Zeitschriften und in Tagungsbänden zunehmend Artikel zu dem Schlagwort „Digitale Bibliothek". Mit diesem Beitrag werden zugrundeliegende Prinzipien und damit verbundene Aufgabenfelder dargestellt.

1 Einleitung

Das Schlagwort „Digital Library" - im folgenden wird die direkte deutsche Übersetzung „digitale Bibliothek" verwendet - ist in zunehmenden Maße in Zeitschriftartikeln und Tagungsbeiträgen zu finden. Eine erste Fachtagung wurde 1994 an der Texas A&M University veranstaltet [DLT94]. 1995 fand die zweite Fachtagung in Austin, Texas, zu diesem Themenbereich statt [DLT95]. Schließlich befaßt sich die April-Ausgabe 1995 der Communications of the ACM mit diesem Themengebiet [DL95]. Aufgrund dieser Entwicklungen versucht dieser Beitrag das Schlagwort „Digital Library" näher zu beleuchten.

Da keine klare Begriffsdefinition gefunden werden konnte, wird zunächst dargestellt, welche Interpretationsmöglichkeiten für den Begriff „digitale Bibliothek" bislang existieren. Anschließend wird vorgestellt, welche Forschungsbereiche im Zusammenhang mit digitalen Bibliotheken derzeit bearbeitet werden. Weiterhin wird kurz auf die Problematik des Urheberrechts und der elektronischen Bezahlung in digitalen Bibliotheken eingegangen. Der Abschluß dieses Beitrages übt Kritik an dem zugrundeliegenden Special Issue des CACM und gibt einen Ausblick auf noch zu erschließende Aufgabenfelder in diesem Bereich.

2 Begriffsbestimmung

In dem Special Issue zu digitalen Bibliotheken [DL95] ist keine einheitliche Begriffsdefinition zu finden, die das Verständnis des Begriffs „digitale Bibliothek" eindeutig prägt oder gar gegen Begriffe wie „elektronische Bibliothek" abgrenzt. In [FoAk95] wird lediglich das mögliche Interpretationsspektrum für den Begriff „Digital Library" angegeben, das im folgenden aufgelistet wird:

1. Digitale Bibliotheken können einfach als elektronifizierte traditionelle Bibliotheken verstanden werden.

2. Digitale Bibliotheken setzen auf traditionellen Bibliotheken auf und erweitern diese um a) neue Funktionalitäten, wie etwa elektronische Ausleihen, b) neue Interaktionsmöglichkeiten, c) neue Ansätze zur Katalogisierung und Klassifikation sowie der damit verbundenen Suchmöglichkeiten, d) neue Typen von Informationsresourcen, etwa Video.
3. Digitale Bibliotheken sind verteilte und vernetzte Informationssysteme, die geeignete Dienste zum Auffinden, Hinzufügen usw. neuer Informationen anbieten.
4. Digitale Bibliotheken bieten Möglichkeiten zur Kommunikation zwischen ihren Benutzern, die dort neues Wissen einbringen und verteilen können.

Wie diese Zusammenstellung zeigt, ist man derzeit noch weit von einem einheitlichen Begriffsverständnis entfernt. Dennoch haben all diese Auslegungsmöglichkeiten einen kleinsten gemeinsamen Nenner, der in [FoAk95] wie folgt beschrieben wird: "*Much of the power of the digital Library is the flexibility it permits in allowing processing of our collections of tangible objects and their electronic representations. . . . and it is feasible, . . . , to apply it also to the collection of things without direct physical analogs, for example, algorithms, real-time data feeds, computational states, relationships among versions of a physical object showing the historical progression of an idea, multimedia annotations, and tours.*"

3 Forschungsbereiche

Dieser Abschnitt stellt verschiedene Bereiche vor, in denen derzeit Forschungsaktivitäten zum Themenkomplex „digitale Bibliothek" laufen. Aufgrund der vorgegebenen Seitenbegrenzung wird kein Anspruch auf Vollständigkeit erhoben. Vielmehr sei auf die Sektionen „Supporting Technologies" und „Library Projects" in [DL95] verwiesen, die einen sehr breiten Überblick über die aktuelle Forschungslandschaft im Kontext digitaler Bibliotheken geben.

3.1 Grundlegende Prinzipien

In [LeMa95] wird der Versuch unternommen, zu klären, welche Technologien zugrundeliegen, was für Dokumente in digitalen Bibliotheken zu finden sind und wie mit digitalen Bibliotheken gearbeitet wird.

3.1.1 Technologie

„*Digital Libraries are based on digital technologies.*" Dieser Satz ist nicht sehr überraschend und verdeutlicht, daß digitale Bibliotheken auf digitalen Technologien basieren. Im Gegensatz dazu umfassen traditionelle Bibliotheken einen sehr viel heterogeneren Dokumentbestand, da dort neben Büchern etwa auch Tonbänder, Videokassetten, und Mikrofiches verwaltet werden. Es stellt sich daher die Frage nach der Heterogenität digitaler Bibliotheken. Vor diesem Hintergrund spiegeln die folgenden drei Szenarien mögliche, zukünftige Entwicklungen wider: 1. Digitale Bibliotheken bieten ausschließlich digitale Dokumente, so daß sie eine ergänzende Rolle neben traditionellen Bibliotheken einnehmen werden. 2. Langfristig existieren ausschließlich digitale Bibliotheken. Die existierenden Dokumentbestände traditioneller Bibliotheken werden digitalisiert und neue Dokumente werden nur in digitaler Form aufgenommen. Traditionelle Bibliotheken würden so von digitalen Bibliotheken abgelöst. 3. Es gibt weder digitale noch traditionelle Bibliotheken. Vielmehr werden Bibliotheken, wie wir sie bislang kennen, zusätzlich digitale Dokumente und damit verbundene Informationsdienste anbieten.

Von den Autoren wird das zweite Szenario als unrealistisch eingeschätzt, so daß abzuwarten bleibt, ob Szenario 1 oder Szenario 3 in einigen Jahren zur Realität geworden ist.

3.1.2 Dokumente

„*Digital library collections contain fixed permanent documents.*" Aus der Sicht traditioneller Bibliotheken können mit diesem Satz sehr allgemein die Dokumente charakterisiert werden, die in digitalen Bibliotheken zu finden sind. Aufgrund der zugrundeliegenden Technologie ist es jedoch leicht möglich, in digitalen Bibliotheken auch Dokumente anzubieten, die nicht fixiert und dauerhaft sind. Dabei wird folgende Unterscheidung zwischen „fixiert" und „dauerhaft" gemacht: Ein Dokument ist fixiert, wenn sein Inhalt sich nicht mehr ändert (etwa ein Buch). Ein Dokument ist dauerhaft, wenn sein Inhalt über lange Zeit nützlich ist (etwa die Bibel). Zur Unterscheidung der beiden Begriffe seien ein Fahrplan der Bundesbahn und ein persönliches Kochbuch genannt. Ein Fahrplan ist zwar fixiert, aber nicht dauerhaft. Ein persönliches Kochbuch ist nicht fixiert - da es stets um neue Rezepte ergänzt werden kann -, dafür aber dauerhaft.

Die Autoren fordern jedoch nicht explizit, daß auch nicht fixierte und nicht dauerhafte Dokumente in digitalen Bibliotheken zugänglich sein sollen. So wird lediglich der Anstoß zu derartigen Überlegungen mit der Frage „*To what extent do we want to consider collections of listserv messages, wire service articles, preprints, and other quickly changing and/or ephemeral documents as appropriate material for digital libraries?*" gegeben.

3.1.3 Arbeiten mit digitalen Bibliotheken

„*Digital Libraries are to be used by individuals working alone.*" Aufgrund der zugrundeliegenden Technology liegt die Annahme nahe, daß Benutzer digitaler

Bibliotheken allein an ihrem Arbeitsplatzrechner sitzen und durch Dokumentbestände browsen, Dokumente suchen, lesen oder gar selbst schreiben. Diese Annahme steht jedoch im Kontrast zu traditionellen Bibliotheken, in denen Bibliothekare Benutzern bei der Dokumentsuche beratend zur Seite stehen. Weiterhin dienen Bibliotheken häufig als Treffpunkt für gemeinsame Recherchen.

Die Autoren verweisen auf das Internet und die damit gemachten Erfahrungen, die gezeigt haben, daß Benutzer des Internets nicht nur Informationen sondern mindestens genauso viel Kommunikation mit anderen Benutzern wünschen. Überträgt man diese Erfahrung als Anforderung an digitale Bibliotheken, so müssen diese auch Kommunikations- und Beratungsdienste anbieten. Beratungsdienste sind insbesondere dann erforderlich, wenn riesige Informationsbestände vernetzt über verschiedene Rechner an verschiedenen Orten verteilt sind.

3.2 Benutzungsschnittstellen und Retrievaltechniken

Der Erfolg digitaler Bibliotheken hängt in besonderem Maße von den Möglichkeiten zum Auffinden gesuchter Information ab. In diesem Zusammenhang spielen zum einen intuitive und komfortable Benutzungsschnittstellen und zum anderen effektive Retrievaltechniken eine entscheidende Rolle. In [RaPe95] werden zu beiden Bereichen existierende Ansätze aus dem Information Retrieval vorgestellt bzw. an Bedürfnisse digitaler Bibliotheken angepaßt. Im folgenden werden lediglich schlagwortartig die wesentlichen Aspekte aus [RaPe95] zusammengefaßt.

3.2.1 Retrievaltechniken

Laut Aussage der Autoren besteht klassisches Information Retrieval aus einer isolierten Aufgabe, in der Anfragen gegenüber homogenen Dokumentbeständen gemacht werden. Diese Annahme trifft jedoch für den Umgang mit digitalen und traditionellen Bibliotheken nicht zu. Vielmehr trifft man hier die folgenden weiteren Anforderungen an, die durch geeignete Retrievaltechniken zu unterstützen sind.

Iterative Anfrageverfeinerung: Häufig sind Benutzer nicht in der Lage, auf Anhieb präzise zu formulieren welche Information sie suchen. Hier sind Anfragetechniken einzusetzen, die Benutzer entsprechend unterstützen.

Heterogene Dokumentbestände: Benutzer suchen Dokumente unterschiedlichen Typs, etwa Videokassette und Buch in traditionellen Bibliotheken. Unterschiedliche Dokumenttypen haben jedoch unterschiedliche Charakteristiken und damit verbunden unterschiedliche Zugriffsmethoden. So werden digitalisierte Textdokumente häufig indiziert, so daß Suchanfragen auf den einmal erstellten Indexen operieren. Im Gegensatz dazu ist es denkbar, daß Videos in digitalen Bibliotheken über inhalts-orientierte Anfragen gesucht werden können, etwa „Suche das Video mit einem blauen Auto, das durch die Wüste fährt.".

Paralleler Zugriff: Da die Ausführung von Anfragen zum Teil recht zeitaufwendig sein kann, etwa weil über ein LAN oder WAN auf Datenbestände zugegriffen wird, ist es wünschenswert Benutzern Möglichkeiten einzuräumen, die mehrere parallele Tätigkeiten zulassen. So können etwa mehrere Anfrage gleichzeitig gestellt werden, oder es kann mit der Analyse bereits erhaltener Resultate begonnen werden, bevor die letzte Anfrage abgeschlossen wurde.

Umfassender Arbeitsprozeß: Der Informationszugriff ist in der Regel in weitere Teile eines umfassenden Arbeitsprozesses eingeflochten. Somit sind neben den eigentlichen Mechansimen zum Auffinden von Informationen auch Aspekte für die Weiterbearbeitung dieser Informationen in einem Arbeitsprozess zu integrieren. Beispielsweise sind Techniken anzubieten, die Benutzer etwa bei der Interpretation und der Analyse der erhaltenen Ergebnisse unterstützen.

3.2.2 Benutzungsschnittstellen

Sehr häufig werden sogenannte Fokus+Kontext-Techniken an den Benutzungsschnittstellen eingesetzt. Die ihnen zugrundeliegende Idee besteht darin, die gewünschte Information im Detail darzustellen und zudem einen Eindruck über den Kontext zu geben, indem sich die Information befindet. Diese Technik hat sich aufgrund der Tatsache bewährt, daß Informationen kaum isoliert sondern vielmehr in einem konkreten Kontext zu sehen und zu verstehen sind.

Häufig werden auch farbige 3D-Effekte mit geeigneten Metaphern, wie etwa Spiralkalender, Dokumentenlinsen, Blickpunktmauern oder Kegelbäumen, zur besseren Visualisierung eingesetzt. Entsprechende Abbildungen sind in [RaPe95] auf Seite 36 zu finden.

3.3 Lehren und Lernen

In [MaMa95] wird die Rolle digitaler Bibliotheken im Kontext von Lehren und Lernen beleuchtet. Zunächst unterscheiden die Autoren zwischen formalem, informalem und professionellem Lernen und zeigen, wie digitale Bibliotheken Einfluß darauf nehmen kann.

Unter formalem Lernen verstehen die Autoren systematisches und gelenktes Lernen, wie man es beispielsweise in Schulen antrifft. Demgegenüber wird informelles Lernen nur vom Lerner selbst kontrolliert. Als Beispiel sei selbstmotiviertes Weiterbilden in verschiedenen Bereichen des alltäglichen

Lebens genannt. Professionelles Lernen ist schließlich rein beruflich motiviert und dient dazu seine beruflichen Fähigkeiten auszubauen.

Aus der Sicht traditioneller Bibliotheken liegen all diesen Lernformen dieselben Technologien, etwa Buchdruck oder Fotografie, zugrunde. Die jeweiligen Informationsresourcen sind jedoch physikalisch voneinander getrennt, so daß für jede Lernform ein eigener Bestand an Informationsresourcen vonnöten ist, beispielsweise eine Schulbuchbibliothek für formales Lernen und persönliche Fachzeitschriften für professionelles Lernen. Daraus resultieren die Nachteil, daß die Integration unterschiedlicher Lernformen kaum möglich sind.

Digitale Bibliotheken können diesen Nachteil ausräumen, da hier die Grenzen zwischen den verschiedenen Informationsresourcen aufgehoben werden. Bei einem digitalisierten und evtl. vernetzten Datenbestand, der die Informationsresourcen für die verschiedenen Lerntypen enthält, ist es ohne weiteres möglich, unterschiedliche Lerntypen miteinander zu verbinden.

Diese Aussage der Autoren wird anhand zahlreicher durchgeführter Projekte und damit den dabei gemachten Erfahrungen untermauert. Diese Projekte befaßten sich u.a. mit den Bereichen virtueller Klassenraum, netzwerk-basierte elektronische Konferenz, PC-Bibliothek oder Einsatz elektronischer Zeitschriften.

4 Urheberrecht und Gebühren

Bei Büchern, Zeitschriften etc. garantiert das Urheberrecht Autoren bzw. Verlagen eine kontrollierte Vervielfältigung und einen kontrollierten Verkauf. Traditionelle Bibliotheken haben hier z.T. Sonderrechte, die das Urheberrecht einschränken. Damit soll es ihnen erleichtert werden, Informationen der Öffentlichkeit zugänglich zu machen, ohne selbst ein Urheberrecht zu verletzen. Beispielsweise ist es in wissenschaftlichen Bibliotheken erlaubt, Teile einer Zeitschrift oder eines Tagungsbandes für wissenschaftliche Zwecke zu kopieren.

Für digitale Bibliotheken müssen nun ebenfalls Rechte und Vorschriften entwickelt werden, die ein unkontrolliertes Vervielfältigen der zugrundeliegenden Informationsbestände einschränken. Ein Vorschlag besteht, ähnlich wie bei video-on-demand, darin, Gebühren entweder auf der Basis „per-unit-of-time-online" oder auf der Basis „per-byte-delivered" zu erheben [Sam95]. Zudem sind rechtliche Grundlagen zu formulieren, die digitale Bibliotheken davor schützen, daß digitale Dokumente aus der Bibliothek geholt werden, um dann auf einem anderen Server gegen Bezahlung weiter angeboten zu werden.

Ein konkreter Ansatz [ACM95], wie er etwa für die geplante elektronische Ausgabe der CACM vorgesehen ist, besteht darin, den kostenlosen Zugang nur bezahlenden Mitgliedern zu erlauben. Nicht-Mitglieder müßten für jeden Zugriff eine Urheberrecht-Gebühr bezahlen. Ob sich dieser Ansatz durchsetzen kann, hängt jedoch in hohem Maße von der Qualität der Zugriffsdienste ab, da dort z.B. die Gebühren der Nicht-Mitglieder zu entrichten sein werden. Sollen die Beiträge eines elektronischen CACM zudem von anderen Distributoren verteilt werden, müssen diese Urheberrecht-Gebühren bezahlen, falls sie einen Personenkreis ansprechen, der größer ist als 1% der ACM-Mitglieder (derzeit ca. 800).

Zur Entrichtung von Gebühren bietet sich die Verwendung des elektronischen Zahlungsverkehrs an. Von der Möglichkeit, Nummern von Kreditkarten und deren Ablaufdatum im Internet zu verschicken, wird aufgrund des damit verbundenen hohen Risikos von Mißbrauch abgeraten. Aus diesem Grund gibt es mittlerweile verschiedene Ansätze [Cha92], [StSt94], [CoKe95], die einen sicheren elektronischen Zahlungsverkehr versprechen. All diese Ansätze basieren auf Grundlagen aus der Kryptografie. Im folgenden werden exemplarisch die zugrundeliegenden Ideen der holländischen Firma DigiCash transparent gemacht:

Wenn ein Kunde elektronisch bei einem Betreiber einer digitalen Bibliothek bezahlen möchte, erzeugt er zunächst eine Zufallsnummer für eine Geldnote (analog den Seriennummern bei Geldscheinen). Diese Nummer erweitert um seine private Kennung schickt der Kunde an seine Bank. Diese kennt die öffentliche Kennung des Kunden, mit der sie die eingehende Nachricht eindeutig dem Kunden zuordnen kann. Die Bank ordnet der Nummer für die Geldnote den gewünschten Betrag zu und bucht das Geld vom Konto des Kunden ab. Ergänzt um die private Kennung der Bank wird der jetzt mit einem Wert versehene „digitale" Geldschein an den Kunden zurückgeschickt. Bei der digitalen Bibliothek kann der Kunde mit diesem Geldschein bezahlen. Diese schickt den „digitalen" Geldschein an die Bank, die zunächst verifiziert, ob von dieser Geldschein von ihr vergeben wurde. Ist dies der Fall, wird der Betrag auf dem Konto der Bibliothek verbucht.

Da bei dieser Methode die Banken grundsätzlich festhalten können, welche Geldscheine an welche Kunden ausgeliefert wurden, können sie auch nachvollziehen, wann und wo ein Kunde wieviel Geld ausgegeben hat. Um diesen Nachteil auszuräumen werden sogenannte „blinde" Signaturen verwendet, die derartige Überwachungen verhindern können.

5 Kritik und Ausblick

Wie die kurze Einführung zeigte, ist der Begriff „digitale Bibliothek” noch nicht eindeutig geprägt. So konnten Beiträge aus dem Special Issue der CACM diesen Begriff nicht weiter eingrenzen. Als kritisch ist hier insbesondere zu bemerken, daß die Gasteditoren [FoAk95] bezüglich der Begriffsklärung nur Anregungen für das Interpretationsspektrum gaben[1]. Der Beitrag über zugrundeliegende Annahmen digitaler Bibliotheken [LeMa95] formuliert zwar Anforderungen, stellt diese aber bald wieder in Frage und macht nur Vorschläge, was noch alles berücksichtigt werden könnte, um den Begriff „digitale Bibliothek” einzugrenzen. Aufgrund dieser Unschärfe im Begriff sollten in nächster Zeit Anstrengungen unternommen werden, digitale Bibliotheken von Forschungsbereichen wie Hypermedia, Benutzermodellierung, Benutzungs-schnittstellen, Information Retrieval, Hypermedia oder WWW und Internet abzusetzen. Andernfalls besteht die Gefahr, daß lediglich bereits vorhandenes unter neuem Namen verkauft wird.

Als positiv ist zu bemerken, daß dem Leser des Special Issues der CACM eine Vielzahl an Initiativen im Zusammenhang mit digitalen Bibliotheken an die Hand gegeben wird. Die jeweils genannten URLs gestatten es, im World-Wide-Web Zusatzinformationen zu den einzelnen Projekten einzusehen.

Bereiche, die in nächsten Zeit zu bearbeiten sind, betreffen etwa das Urheberrecht-Problem und den elektronischen Zahlungsverkehr, die Gestaltung neuer und die Integration vorhandener Dienstleistungsformen von seiten der Bibliotheks–betreiber, die Entwicklung und den Einsatz von Werkzeugen zur Kooperation und Kommunikation, die Versionierung von Dokumenten, die Entwicklung neuer Lehr- und Lernstrategien, die Benutzermodellierung sowie die Modellierung angemessener Benutzungsschnittstellen. Die wesentliche Herausforderung, die aus Sicht der Informatik an digitale Bibliotheken zu sellen ist, besteht darin, Ergebnisse aus den oben genannten Gebieten auf digitale Bibliotheken abzustimmen und in einen einheitlichen Rahmen zu integrieren.

6 Literatur

[ACM95] ACM Interim Copyright Policies; in: [DL95], S. 104-107.

[Cha92] D. Chaum; Achieving Electronic Privacy; in: Scientific American, August 1992, S. 96-101. (elektronische Version: URL: http://www.digicash.com/publish).

[1]Es sei jedoch erwähnt, daß neueste Artikel, etwa [NüFu95], Beiträge hierzu leisten.

[CoKe95] S.B. Cousins, S.P. Ketchpel et al.; Managing Multiple Payment Mechanisms in Digital Libraries; in: [DLT95].

[DL95] Special Issue: Digital Libraries; in: Communications of the ACM Bd. 38, Nr. 4, 1995.

[DLT94] J.L. Schnase, J.J. Leggett, R.K. Furuta, T. Metcalfe; Proceedings of the 1st Annual Conference on the Theory and Practice of Digital Libraries: Digital Libraries 1994, Texas A&M University, 1994, (elektronische Version: URL: http://atg1.wustl.edu/DL94).

[DLT95] D.L. Levy; Proceedings of the Second International Conference on the Theory and Practice of Digital Libraries, Austin, Texas, 1995, (elektronische Version: URL: http://csdl.tamu.edu/DL95).

[FoAk95] E. Fox, R. Akscyn, R. Furuta, J. Leggett; Digital Libraries; in [DL95], S. 23-28.

[RaPe95] Ramana Rao, Jan.O.Pedersen et al.; Rich Interaction in the Digital Library; in: [DL95], S. 29-39.

[LeMa95] D. Levy, C. Marshall; Going Digital: A Look at Assumptions Underlying Digital Libraries; in: [DL95], S. 77-75.

[MaMa95] G. Marchionini, H. Maurer; The Roles of Digital Libraries in Teaching and Learning; in: [DL95], S. 67-75.

[NüFu95] P. Nürnberg, R. Furuta et al.; Digital Libraries: Issues and Architectures; in: [DLT95].

[Sam95] Pamela Samuelson; Copyright and Digital Libraries; in: [DL95], S. 15-21 + S. 110.

[StSt94] L.H. Stein, E.A. Stefferud et al.; The Green Commerce Model; in: Technical Report, First Virtual Holding Incorporated, (elektronische Version: URL: http://www.fv.com/tech/green-model.html).