



Informations- technologien in Bibliotheken

Von
Dr. Harald Jele

R. Oldenbourg Verlag München Wien

Hinweise, Korrekturvorschläge sowie Kommentare sind ausdrücklich erwünscht via Email

an
harald.jele@uni-klu.ac.at

oder *Snailmail* an
Dr. Harald Jele
Seegasse 27
A-9020 Klagenfurt

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Jele, Harald:

Informationstechnologien in Bibliotheken / von Harald Jele. -

München ; Wien : Oldenbourg, 2001

ISBN 3-486-25818-4

© 2001 Oldenbourg Wissenschaftsverlag GmbH

Rosenheimer Straße 145, D-81671 München

Telefon: (089) 45051-0

www.oldenbourg-verlag.de

Das Werk einschließlich aller Abbildungen ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Bearbeitung in elektronischen Systemen.

Gedruckt auf säure- und chlorfreiem Papier

Druck: MB Verlagsdruck, Schrobenuhausen

Bindung: R. Oldenbourg Graphische Betriebe Binderei GmbH

ISBN 3-486-25818-4

Inhaltsverzeichnis

Vorwort	3
1. Einleitung	5
2. Internet	7
2.1 Vorbemerkung	7
2.2 Eine typische, einfache Web-Verbindung	9
2.3 Eine komplexe Web-Verbindung	12
2.4 Anwendungen innerhalb einer Bibliothek: Informations- und Applikationsserver	18
3. Intranet & lokale Netze	20
3.1 Realisierung mit traditionellen <i>LAN</i> -Technologien	23
3.2 Realisierung mit Intranet-Technologien	25
4. Bibliographische Datenbanken	27
4.1 Typische Inhalte	29
4.2 Datenformate	31
4.3 Datenträger	36
4.4 Indexsysteme	38
4.5 Frontends und Retrieval	44
5. Volltextdatenbanken	50
5.1 Erweiterte Retrievalfunktionen	51
6. Bibliothekssysteme	55
6.1 Lokale Bibliothekssysteme	57
6.2 Kooperative Verbundsysteme	63
6.3 Zentralsysteme	68
6.4 Sonderformen	70
7. CD-ROM-Server-Systeme	73
7.1 CD-ROM-Server-Integration in ein klassisches <i>LAN</i>	75
7.2 CD-ROM-Terminal-Server	77
8. Anhang	80
8.1 Abbildungen zu Frontends und Retrieval	80
8.1.1 Die <i>einfache Suche</i> in einer Datenbank	80
8.1.2 Die Expertensuche	82

8.1.3	Indexbrowsing	84
8.1.4	Kombinierte Suchmasken	87
8.1.5	Speicherung des Suchverlaufs	89
8.2	Abbildungen zum Anzeigeformat bibliographischer Daten .	91
8.2.1	Katalogkartenformat	91
8.2.2	Tabellarische Anzeige	93
8.2.3	Anzeige nach dem Kategorienschema	94
8.2.4	Anzeige der Kategorienbezeichner	95
8.2.5	Typische Anzeigeformen in Volltextdatenbanken	96
8.3	Abbildungen zum internen Datenformat bibliographischer Daten	100
8.3.1	MAB2-ähnliche, interne Struktur; hierarchische Speicherung	100
8.3.2	Das <i>allegroC</i> Format	101
8.4	Abbildungen zum Austauschformat bibliographischer Daten	102
8.4.1	MAB2-Diskettenformat für Titeldaten	102
8.4.2	MAB2-Bandformat für Titeldaten (MAB2-Datendienst-Format)	103
8.4.3	MARC-Datenformat	104
8.5	Abbildungen zu den Sonderformen von Bibliothekskatalogen	105
8.5.1	Imagekatalog ohne Index-Suchmöglichkeiten	105
8.5.2	Imagekatalog mit Index-Suchmöglichkeiten	107
8.6	Begriffe zur bibliographischen Datenbankrecherche	110
9.	Abbildungsverzeichnis	118
10.	Abkürzungsverzeichnis	119
11.	Raum für Notizen	122
12.	Begriffsregister	124
13.	Literaturverzeichnis	128
13.1	Angabe der Quellen, die in ihrer gedruckten Form zitiert wurden	128
13.1	Angabe der Quellen, die in ihrer Online-Form zitiert wurden	130

Vorwort

Der rasante Umbruch in der Informationslandschaft allein in den letzten zehn Jahren hat das Arbeiten in und mit Bibliotheken so deutlich und wesentlich verändert, daß all jene, die diesem nicht schritthalten können, teilweise zu den *neuen Analphabeten* zu zählen sind.

Die Berufsausbildung angehender Bibliothekare oder die Weiterbildung diesbezüglich interessierter ist diesem Wandel gleichermaßen ausgesetzt.

Dieser Text richtet sich somit an verschiedene Gruppen. Die hier wiedergegebenen Inhalte sind jedoch aus einem Vorlesungsskriptum zur gleichnamigen Vorlesung entstanden und auch weiterhin zu diesem Zweck gedacht und aus diesem Grund stilistisch darauf abgestimmt.

Das Thema „Informationstechnologie in Bibliotheken“ liegt deutlich im Spannungsfeld zwischen *technischer Theorie* und *praktischer Anwendung* durch Bibliothekare und Benutzer.

Diesem Spannungsfeld möchte ich dadurch Rechnung tragen, daß ich einerseits wesentliche Inhalte, die das Berufsverständnis eines Bibliothekars heute mit prägen, in durchaus sehr technischer Art referiere und deren Grundlagen in dieser Weise darstelle. Andererseits versuche ich, an einigen Stellen Hinweise auf praktische Realisierungsmöglichkeiten zu geben bzw. diese beispielhaft zu beschreiben. Die Zahl der praktischen Beispiele, die hier vorgestellt werden, ist bewußt sehr bescheiden und gering gehalten. Dieser Umstand, sollte er stören, kann jedoch sehr einfach durch Rückgriff auf jene Beispiele vermindert werden, die sich in den angegebenen Basiswerken befinden, die im Literaturverzeichnis vermerkt sind.

Zudem relativiert sich die Sinnhaftigkeit des Vermerkens von praktischen Beispielen allein durch den Umstand ihrer technologisch meist kurzen Lebensdauer. Im Rahmen einer Vorlesung müssen diese ohnehin durch das Abhalten von eigenen Übungseinheiten gedeckt werden.

Einige Schwierigkeiten, die sich beim Erstellen des Textes zeigten, waren auf

den Umstand zurückzuführen, daß bestimmte Technologien bzw. deren konkrete Realisierung mit Produktnamen behaftet sind, die ich im eigentlichen Text auszusparen versuche. Leider ist dies nicht immer gelungen bzw. mir möglich gewesen. Trotzdem ist es mir ein Anliegen, mit einem solchen Text keine „Produktschau“ zu entwickeln oder gar zu zeigen, wie bestimmte Firmen bestimmte Technologien einsetzen.

Beim Lesen kann an manchen Stellen durchaus der von mir gewollte oder gar provozierte Eindruck entstehen, daß ich davon ausgehe, daß – durch einen Normenausschuß [!] – *standardisierte* Produkte sogenannten *proprietären* vorzuziehen und daß *Ideen, Modelle, Verfahren, Programme* oder *Technologien* aus der *Open-Source*-Bewegung gegenüber geschlossenen Firmenrealisierungen wahrzunehmen und zu unterstützen sind. Sollte diese, meine Sichtweise, Ihrer persönlichen widersprechen, so muß ich hier um Ihre freundliche Nachsicht bitten.

Begriffe, die in ihrer mitunter gängigen, abgekürzten Form verwendet werden, sind bei ihrer ersten Verwendung (üblicherweise in Fußnoten) erklärt bzw. beschrieben. Die wichtigsten Abkürzungen finden sich zudem im Abkürzungsverzeichnis sowie im Begriffsregister wieder.

Jene (abgekürzten) Begriffe, von denen ich ausgehe, daß sie ohnehin inzwischen zum Allgemeinwissen zählen, wurden hier nicht weiter beschrieben oder erklärt; diese sind aber bereits in den gängigen Fachlexika erklärt zu finden.

Geschlechtsspezifische Bezeichnungen wurden in diesem Text nicht in ihrer männlichen und weiblichen Form, sondern ausschließlich in ihrer männlichen verwendet. Mein Versuch, beide Formen durchgängig einzusetzen, muß leider als gescheitert angesehen werden. Geschlechtsneutrale Bezeichnungen ließen sich leider nicht immer finden.

1. Einleitung

Das Thema *Internet* in einem Skriptum wie diesem überhaupt zu behandeln, scheint mir aus mehreren Gründen als zweifelhaft, zumal in den letzten Jahren weitere Standardwerke zur einschlägigen Literatur hinzugekommen sind und bestehende wesentlich erweitert und auf den letzten Stand gebracht worden sind. Trotzdem ist dieses hier vertreten.

Bezugnehmend auf den erstgenannten Umstand werden in den betreffenden Kapiteln jedoch nur die wesentlichsten technologischen Aspekte zu diesem Thema beschrieben, indem zuerst *einfache*, klassische, auch für kleine bibliothekarische Einrichtungen notwendige *Internettechnologien* behandelt und anschließend deren *komplexe* Varianten besprochen werden.

Zu diesem Schritt hat letztlich der Umstand geführt, daß mit modernen, zeitgemäßen Informationstechnologien zunehmend das Konzept von eigenständigen *Applikationsservern* auf der Basis gängiger und standardisierter *Internettechnologien* verfolgt wird, das besser beschrieben werden kann, wenn die betroffenen Technologien in diesem Text auch genannt und in einen sinnstiftenden Kontext gestellt sind.

Ähnliches gilt zum Thema *Intranet*, das hier natürlich nur sehr verkürzt dargestellt und in jenem Kontext referiert wird, indem es in weiterer Folge bei der Beschreibung von *CD-ROM-Server-Systemen* zitiert und fortgeführt wird.

Bei der Behandlung der Themen *Bibliographische Datenbanken*, *Volltextdatenbanken* und *Bibliothekssysteme* wurde hingegen versucht, stellenweise durchaus umfassender zu referieren und die wesentlichsten Merkmale, inhaltlichen Schwerpunkte und Kennzeichen einer technischen Realisierung zusammenzufassen; zumal diese Themen – auch einem Informatiker – nicht ohne weiteres soweit zugänglich sind, daß sie technologisch immer einer einfachen Lösung zugeführt werden können.

Die Realisierung des Kapitels zu den *Bibliographischen Datenbanken* ist dabei

im speziellen als eine Sammlung von Kennzeichen zu verstehen, ohne daß ich diese Aufzählung hier in einer fertigen Sichtweise verstehen möchte.

2. Internet

2.1 Vorbemerkung

Technologien, die heute („landläufig“) unter dem Begriff *Internet* verstanden werden, sind häufig jene, die mit dem Entstehen des *WWW*¹ entwickelt wurden und damit auch weite Verbreitung fanden:

Dabei bietet ein Web-Server Zugriff auf verschiedene (textuelle oder graphische) Inhalte, Daten², akustische oder visuelle Informationen³ durch einen einzigen, „plattformunabhängigen“⁴ Client, meist über den sog. *Web-Browser*⁵.

Daneben existieren im Internet weitere Technologien, die sich voneinander relativ unabhängig entwickelt haben und eigener Clients bedürfen wie z.B.: *FTP*

1 *WWW*: World Wide Web, oft auch „*das Web*“ genannt

2 „*Daten*“ im Sinne von „Einträgen“ in Datenbanken

3 wie gespeicherte *Tondokumente*, *Filme* oder durch sog. *Streaming-Formate* Live-Töne (z.B. *Live-Radio* oder *Realaudio*) oder Live-(*Realtime/Echtzeit*-)Bilder von „*Web-Cameras*“

4 unter einer *Plattform* versteht man in diesem Kontext das spezifische Betriebssystem, mit dem ein Computer betrieben wird: die bekanntesten in der Web-Welt sind Windows (in den Versionen 3.11, 95, 98, 2000 und NT), MacOS (für Apple MacIntosh), LINUX (für Intel- und PowerPCs sowie für Alpha-Workstation und -Server) sowie div. kommerzielle UNIX-Varianten (Sun-Solaris, True64-Unix (DigitalUNIX/ULtrix), IBM-AIX).

Die heute sich am Markt befindenden Web-Clients sind leider nicht von ihrer Plattform unabhängig. Die wenigsten existieren überhaupt für mehr als zwei Plattformen und zeigen mitunter auf unterschiedlichen Plattformen ein (leicht) unterschiedliches Verhalten.

Plattformunabhängigkeit existiert aber in diesem Sinn: ein Windows-Browser kann mit einem UNIX-Web-Server und umgekehrt üblicherweise problemlos kommunizieren, da die verwendeten Grundlagen (Protokolle) für alle Betriebssysteme standardisiert sind

5 als *Browser* werden diese Clients deshalb bezeichnet, weil sie selbst eigentlich keine Anwendungen ausführen, sondern Inhalte anzeigen (die auch Ergebnisse von sog. Skripts (=Programmen) sein können). Anwendungen wie z.B. für das Abspielen eines Videos werden durch *Browser-Plugins* realisiert (=externe, dem Browser fremde Programme, die dieser mitverwaltet und bei Bedarf automatisch startet).

Bekannt und weit verbreitet sind *Netscape*, der *Microsoft Internet Explorer* oder auch *Opera*, sowie *Lynx* als eine Sonderform eines Browsers, der Inhalte nur textuell anzeigt und sich deshalb auch für blinde Personen zum Lesen von Textinhalten mit einer Brailletastatur eignet

(das File Transfer Protocol) zum Übertragen von Daten in Dateiform, *Email*⁶ zum Verschicken „elektronischer Post“, *News*⁷ (oder auch: *Newsgroups* im sog. *Use-Net*) als Diskussionsforen.⁸

Im Gegensatz zur ehemals *konventionellen* Technologie brachte dieser Ansatz eine deutlichere Trennung von Client- und Server-Funktionalitäten mit sich. Die Aufgabe des Clients ist dabei häufig, Daten anzuzeigen, diese (graphisch, tabellarisch etc.) aufzubereiten oder zu sortieren. Der Server hingegen ist als Anbieter und Verteiler der Daten zu sehen. Zudem regelt der Server den Zugang zur Information entsprechend einer bestimmten, gespeicherten Rechtevergabe.⁹

Daß dieser Ansatz nicht immer konsequent weiterverfolgt wurde, sieht man anhand der nachfolgenden Graphik: Typische Serveraufgaben (wie die Aufbereitung eines Datenretrievals) werden oft durch *Skriptsprachen* (wie *Java-Script*) im Browser erledigt. Der Server bekommt einen schon fertigen Suchstring übermittelt und führt das eigentliche Retrieval anhand dessen in der Datenbank

6 mit den wichtigsten Protokollen: *POP* (Point of Presence) für eingehende Mails, *SMTP* (Simple Message Transfer Protocol) für ausgehende Mails, *IMAP* (Internet Message Access Protocol) für einen Abgleich mehrerer lokaler Mailfolder mit einem zentralen

7 das für das Übertragen von *Newsgroups* von einem Newsserver zum -client verwendete, standardisierte Protokoll ist das *NNTP* (*Network News Transfer Protocol*)

8 das Installieren, Verwalten und Benutzen so verschiedener, spezialisierter Clients ist für den Benutzer wie den Administrator dieser Programme eher unkomfortabel: zumindest auch aus diesen Gründen wurden bei den meisten Web-Browsern solche Clients integriert und die eigentlich verschiedenen Funktionalitäten durch die Anpassung sich sehr ähnlicher graphischer Benutzerschnittstellen (*GUI* = Graphical User Interface / Oberflächen) leichter handhabbar gemacht. Die meisten Browser ermöglichen dies dadurch, daß sie in einem Eingabefeld nicht nur die Angabe von Locations (*URLs* = Unified Ressource Locator, Ort der Information bzw. Ressource) zulassen, sondern daneben auch Angaben zum Übertragungsprotokoll akzeptieren. So können durch Eingabe z.B. von `ftp://ftp.anywhere.org` neben den üblichen `http(s)://` - Verbindungen auch *FTP*-Verbindungen hergestellt werden

9 bei der Beschreibung der Kommunikation zwischen dem Client und dem Server spricht man in diesem Zusammenhang auch von der Realisierung einer spezifischen *Client-Server-Kommunikation* bzw. der Realisierung eines *Client/Server-Modells*

durch.¹⁰

2.2 Eine typische, einfache Web-Verbindung

Abbildung 1 zeigt das Prinzip einer einfachen Web-Verbindung, die heute noch typische Verbreitung findet:

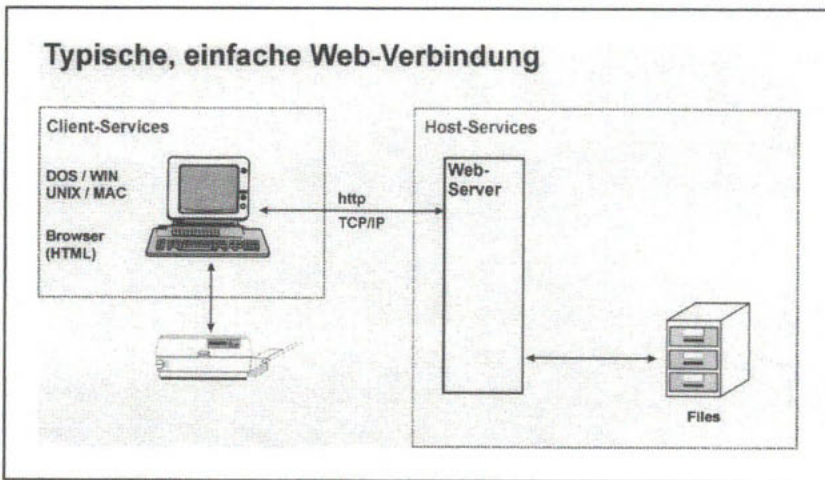


Abbildung1: Typische, einfache Web-Verbindung

Im Zentrum dieses Ansatzes befindet sich ein *Web-Server*¹¹, dessen Zugriff auf unterschiedliche Weise beschränkt oder der öffentlich zugänglich sein kann. Einerseits ist das Einführen einer Zugriffsbeschränkung durch das Vorschalten einer sog. *Firewall* möglich. Diese regelt den Datenverkehr zwischen der Außenwelt und dem Server, indem alle Datenpakete, die den Server erreichen

¹⁰ Vor- und Nachteile dieser voneinander verschiedenen Ansätze werden im nachfolgenden Textteil kurz besprochen

¹¹ bekannte Software-Produkte für die Realisierung von Web-Servern sind der frei verfügbare, kostenlose *Apache-Server* (der auf allen gängigen Plattformen eingesetzt werden kann) oder der kostenpflichtige *Microsoft Internet Information Server (IIS)* für die Windows-Plattform

oder verlassen dürfen durch sie hindurch geschleust („getunnelt“) und dabei auf Gültigkeit und „Plausibilität“ geprüft werden. Die Gültigkeit wird dabei in den meisten Fällen durch das Zulassen oder Sperren von *IP-Adressen* geregelt.¹² Anderenfalls werden nicht öffentlich zugängliche Web-Server durch eine Prüfung der Benutzerberechtigung (mit Benutzer-Namen und Passwort) gesichert bzw. in Kombination mit einer *Firewall* installiert.

Das Absichern eines Web-Servers gegenüber unberechtigten Zugriffen stellt im typischen Fall jedoch eher die Ausnahme dar. Web-Server bieten heute zumeist öffentlich zugängliche Inhalte.

Im Fall des sog. *Teleworking*, wenn also Mitarbeiter einer Firma außerhalb dieser auf nicht öffentlich zugängliche Inhalte *aktiv* (schreibend) und *passiv* (lesend) Zugriff benötigen, sind sowohl der Einsatz von Firewalls (nach innen und nach außen) in Kombination mit dem Schutz durch *Benutzernamen* und *Passwörter* (=login) notwendig.¹³

Mit dem Web-Server können weitere, sichere *Datenbank*- und *Fileserver*-Verbindungen aus dem lokalen Netz nach außen hergestellt werden:

Ein *Web-Browser* sollte aus Sicherheitsgründen nie direkt auf eine Datenbank

12 IP-Adressen (*IP* = *Internet Protocol*) regeln den Datenverkehr innerhalb des Internet: weltweit können öffentliche IP-Adressen nur einmal verwendet werden und sind aus diesem Grund eindeutig. Internetadressen sind sowohl in einer lesbaren Form (dem Namen) als auch in Zahlenform vorhanden. Die Zuordnung zwischen den IP-Adressen in Zahlenform und der entsprechenden Namensform regeln sog. *Name-Server*. Als Bsp. entspricht die IP-Adresse 143.205.64.64 dem Namen alpha1-ub.sci.uni-klu.ac.at.

Ein Datenpaket kann im Internet aus diesem Grund auch leicht an seinen Adressaten zugestellt werden, da jedes solche die Angaben zum Absender und zum Adressaten trägt. Mit geeigneten Filtern ist es jedoch relativ leicht möglich, Internet-Pakete bzw. deren Angaben und Inhalte zu fälschen und damit durch eine solche Firewall zu gelangen (zu „hacken“). Aus diesem Grund prüfen Firewalls zudem die Plausibilität, wie (=in welcher Reihenfolge) Datenpakete ankommen (z.B. dupliziert) und weitergegeben werden, um auf eventuelle *Hacking*-Versuche Hinweise zu bekommen.

Eine gut verständliche Übersicht zur (Daten-)Kommunikation im Internet findet sich in Hosenfeld & Brauer (1995) oder auch in Holtschneider (1996)

13 Web-Server unter dem Betriebssystem LINUX sind aus dem Grund auch deshalb weit verbreitet, da LINUX Firewall-Mechanismen schon auf der Ebene des *Programm-Kernels* (ohne weitere Zusatzsoftware auf Anwendungsebene) bietet. Da der *Kernel* alle weiteren System-Prozesse überwacht, gilt LINUX auch als besonders einbruchssicher gegenüber Hacking-Versuchen