

Der Darmstädter Virtuelle Gesamtkatalog

Ausbau und Erweiterungen 1996

Rudi Schmiede, Helmut Wenzel

Seit Jahren arbeitet die Projektgruppe „Dezentrale Systeme im Bibliothekswesen“ an der Verbesserung der Bibliotheksinfrastruktur innerhalb der Technischen Hochschule und der Bibliotheksregion Darmstadt. In unseren Arbeiten zur Gestaltung der EDV-Infrastruktur des Bibliothekswesens auf Landes- und Hochschulebene stand uns als Perspektive stets ein Informationsnetz vor Augen, welches es Wissenschaftlern wie Studenten ermöglichen sollte, direkt vom Arbeitsplatz aus auf Literaturlisten zuzugreifen, ohne erst noch mühsam eine Retrievalsprache oder die Bedienung mehr oder weniger komplexer Programmsysteme erlernen zu müssen und ohne auf proprietäre Netzwerk- und Rechnerstrukturen angewiesen zu sein. Mit der Etablierung des WWW läßt sich diese Vorstellung nun erstmalig in zufriedenstellender Form realisieren.*

Nach der erfolgreichen Implementierung des Prototyps eines integrierten Retrievalsystems für das World Wide Web (WWW) im Jahr 1995 konnte unser Projektteam (*Rudi Schmiede, Helmut Wenzel, Winfried Goß*) in diesem Jahr daran gehen, die mittels EDV erschlossenen bibliographischen Bestände und Informationen der dezentralen TH-Bibliotheken, der Hessischen Landes- und Hochschulbibliothek sowie weiterer Einrichtungen innerhalb eines stabilen Informationssystems im WWW zu integrieren. Das diesem Dienst zugrunde liegende Retrievalsystem wurde von unserer Arbeitsgruppe im Auftrag des Ständigen Ausschusses für das Bibliothekswesen (StA IV) der THD entwickelt und im Berichtsjahr weiter verbessert. Die Finanzierung erfolgte wie bisher durch zentrale Hochschulmittel.

Der Informationsdienst ist unter der folgenden Adresse im Internet zu erreichen: <http://www.ifs.th-darmstadt.de/dvk.html>.

1. Die Ausgangssituation 1996

Nachdem die Finanzierung des Projekts für 1996 gesichert war, wurden die im Arbeitsbericht 1995 bereits ins Auge gefaßten Arbeiten zur Konsolidierung, Stabilisierung und Erweiterung des Systems in Angriff genommen. Hierzu zählten insbesondere:

Vgl. dazu unseren Projektbericht 1995 „Darmstädter Bibliotheken im World Wide Web“ im BIBLIOTHEKSDIENST 30 (1996) 1, S. 75 ff

- Der Ausbau der vorhandenen Datenbasis durch die Einbeziehung aller elektronisch erfaßten Bestände der dezentralen Bibliotheken der THD und die Implementierung einer universalen Schnittstelle zur Datenübernahme aus den verschiedensten Bibliothekssystemen und -formaten.
- Die Konsolidierung des Dienstes und damit verbunden die Transition des Prototyps zu einem integrierten System stabiler Applikationen für den Dauerbetrieb.
- Die Aufrüstung der vorhandenen Low-Cost-Hardware zur Sicherstellung akzeptabler Antwortzeiten bei erheblich gesteigener Nutzungsfrequenz.
- Die konsequente Realisierung des Client-Server-Prinzips innerhalb des Gesamtsystems.
- Der Nachweis der im Lokalsystem PICA des Hessischen Verbundes erfaßten Bestände der Hessischen Landes- und Hochschulbibliothek.
- Die Öffnung des Systems für externe wissenschaftliche und/oder Öffentliche Bibliotheken.

Die Arbeiten an dem unter Allegro implementierten separaten Darmstädter Gesamtkatalog (GK) wurden entgegen unseren ursprünglichen Vorstellungen nicht mehr weitergeführt, vielmehr wurden dessen Funktionen in das WWW-System integriert, so daß in konsequenter Realisierung des Client-Server-Verfahrens nun ein „virtueller“ Gesamtkatalog angeboten wird, der nicht der doppelten Datenhaltung in einem zusätzlichen Stand-Alone-System bedarf.

2. Neue Elemente des Systems

Wegen der Anfang des Jahres zunächst unsicheren Finanzierungssituation und auf Grund von Koordinationsproblemen innerhalb der Hochschule konnten die projektierten Arbeiten erst im März 1996 aufgenommen werden und mußten auf Grund der verlorenen Zeit dann unter dem konzentrierten Einsatz der vorhandenen personellen und materiellen Ressourcen erfolgen.

2.1 Die Datenbasis

Vorrangig wurde zuerst das Angebot der verfügbaren Datenbestände kräftig erweitert, so daß bereits zum Sommersemester 1996 die Kataloge aller dezentralen Bibliotheken der TH, die ihre Bestände mittels EDV verwalten, in den Informationsdienst integriert werden konnten; und zwar unabhängig vom jeweils verwendeten Bibliothekssystem. Obgleich der größte Teil der Fachbereichs-, Instituts- und Lehrstuhlbibliotheken an der TH mittlerweile mit dem Braunschweiger Allegro-Paket arbeitet, existiert eine Reihe von Bibliotheken, welche auf absehbare Zeit mit anderen Programmen arbeiten werden. Um auch deren Bestände nachzuweisen, war in unserem System von Anfang an

ein zunächst einfaches Schnittstellenprogramm enthalten, welches mittlerweile erheblich erweitert wurde, um die Integration verschiedenster Datenformate ohne größeren Aufwand erledigen zu können.

Hierzu wurde ein Konvertierungsmodul entwickelt, das auf dem zum Allegro-Paket gehörenden Tool *<import>* aufsetzt und sowohl die Konvertierung der Daten vornimmt als auch die notwendigen Index-Tabellen erstellt. Dieses im Lauf des Jahres mehrfach modifizierte und optimierte Modul stellt mittlerweile eine standardisierte, universale Schnittstelle zur Übernahme nahezu beliebiger Datenformate in das WWW-Retrievalsystem dar. Über dieses Import-Interface wurden bislang Daten aus den verschiedensten Bibliothekssystemen eingebunden: Allegro, Bibdia, Lars, Lidos, SISIS, Unipals und Urica.

Der Zugriff auf die verschiedenen Datenbestände ist abhängig von der Art der Datenhaltung, die innerhalb der TH auf unterschiedliche Weise erfolgt. Ein erheblicher Teil jener dezentralen Bibliotheken der TH, die mit Allegro unter MS-DOS arbeiten, läßt seine Daten zentral durch das Hochschulrechenzentrum (HRZ) verwalten, das die Datenbestände auf zwei seiner Novell-Server verteilt hat. Da der Zugriff auf diese Daten auf Grund der vorhandenen HRZ-Konfiguration ausschließlich über das Netware-Protokoll (IPX) erfolgen kann, war es notwendig, in unserem WWW-System ein entsprechendes Gateway-Modul zu implementieren, das den unmittelbaren Zugriff auf Novell-Dateisysteme unter UNIX ermöglicht. Wir verwenden dafür das IPX-Modul des amerikanischen Herstellers *NetCon Business Systems*, das sich durch seine bekannte Stabilität wie durch eine hohe Flexibilität auszeichnet.

Soweit einzelne Fachbereiche oder Institute ihre Daten auf eigenen Novell-Servern halten, kann der Zugriff darauf ebenfalls mit Hilfe dieses Moduls erfolgen. Der Zugriff auf die übrigen dezentralen Bibliotheksdaten erfolgt über das Standardprotokoll TCP/IP mittels NFS, rcp oder ftp. Die Gesamtzahl der über das WWW-System erreichbaren Titeldaten liegt zur Zeit bei rund 600.000 Einheiten und wird durch die Beteiligung weiterer Bibliotheken noch steigen.

2.2 Die Grundfunktionen des Systems

Obwohl sich der im Vorjahr entwickelte Prototyp des integrierten WWW-Retrievalsystems bereits als recht stabil erwiesen hatte, waren schließlich doch noch tiefgreifende Modifikationen und Neuentwicklungen erforderlich, um daraus eine stabile und zugleich möglichst einfach zu administrierende Applikation für den Dauerbetrieb entstehen zu lassen. Der wichtigste Schritt bestand darin, die Idee des modularen Aufbaus im Gesamtsystem konsequent umzusetzen und die Integration der verschiedenen Module mit Hilfe einheitlicher Schnittstellen- und Interaktionsalgorithmen zu realisieren. Dieses Prinzip erst erlaubt die angemessene Integration der verschiedensten Datenformate

und den standardisierten Durchgriff auf die unterschiedlichen Systemkonfigurationen der beteiligten Bibliotheken.

Das zentrale Ausgabemodul unseres Systems, an das alle Suchanfragen übergeben werden, regelt den Zugriff auf die verschiedenen Datenbestände. Es arbeitet direkt mit *srch* zusammen, einem zum Allegro-Paket gehörenden Datenbankabfrage- und Formatierungsprogramm, und gewährleistet durch seine besondere Funktionalität nicht nur den schnellen Zugriff auf Allegro-Daten, sondern ermöglicht darüber hinaus eine optisch ansprechende Aufbereitung der verfügbaren Recherche-Ergebnisse.

Bevor die Suchanfragen jedoch an das Ausgabemodul übergeben werden können, müssen sie zunächst der entsprechenden Datengrundmenge zugeordnet, zusammengestellt und aufbereitet werden. Dieses *preprocessing* erfolgt jetzt ebenfalls innerhalb eines speziell für diesen Zweck entwickelten Moduls, in dem insbesondere die für den jeweiligen Suchvorgang benötigte Datenbasis vorkonfiguriert wird. Dadurch ist es nunmehr möglich, kombinierte Suchen in verschiedenen Datenbeständen gleichzeitig durchzuführen, sei es die Suche über alle Bibliotheken eines oder mehrerer Fachbereiche oder die simultane Abfrage aller Datenbestände (virtueller Gesamtkatalog). Das Verfahren der Vorverarbeitung gewährleistet kurze Antwortzeiten auch bei komplexen Simultanrecherchen über viele Kataloge. Das Vorverarbeitungsmodul selbst verwendet dabei eine Reihe von Indextabellen, die vom System laufend selbstständig aktualisiert werden. Das Zeitverhalten wird auf diese Weise weitgehend unabhängig von der vorhandenen Datenmenge.

Insgesamt wurden im Lauf des Projektjahres etwa 80 % des Programm-Codes neu erstellt oder entscheidend umgeschrieben, wodurch sich nicht nur die Performance des Systems deutlich verbesserte; gleichzeitig wurde auch die zuvor sehr aufwendige Administration der vorhandenen Informationscluster erheblich erleichtert, und die Einbindungsmöglichkeiten für neue Datenbestände wurden wesentlich verbessert.

2.3 Der Ausbaustand der Hardware

Der Prototyp des Informationssystems war ursprünglich auf einem kleinen UNIX-Rechner der Firma SUN (Sparc 10) installiert worden, da die benötigten Allegro-Programme zunächst nur auf dieser Plattform mit der erforderlichen Stabilität liefen. Es zeigte sich jedoch schon bald, daß dieser Rechner auf Grund des hohen Ressourcenverbrauchs des Betriebssystems (Solaris 2.4), des zu geringen Arbeitsspeichers und wegen der relativ niedrigen Taktrate des Prozessors die rasch steigende Zahl der eingehenden Suchanfragen nicht mehr mit akzeptablen Antwortzeiten bedienen konnte. Da die für die unumgängliche Aufrüstung des WWW-Servers notwendigen Investitionsmittel in

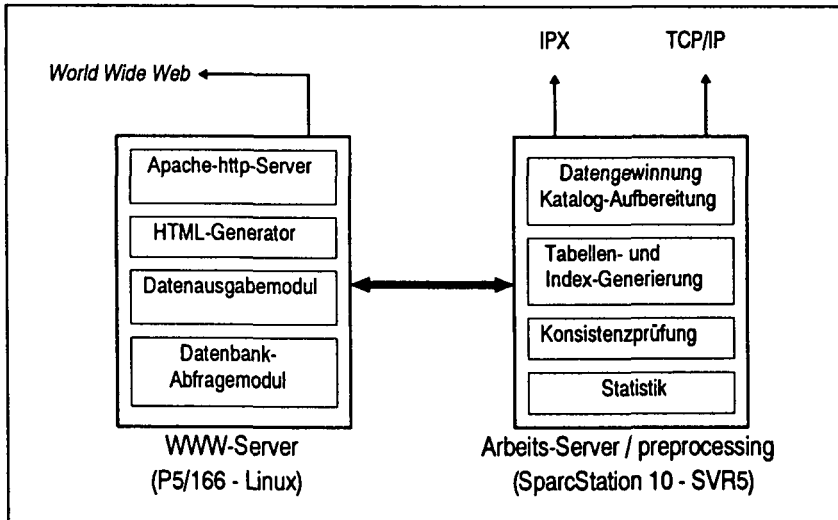
Höhe von DM 12.000,- nicht zur Verfügung standen, mußte ein anderer Weg gefunden werden, um die Performance des Systems anzuheben.

Die Lösung bestand darin, mit Standardkomponenten aus der PC-Welt einen WWW-Server aufzubauen, der mit dem leistungsfähigen, unter GNU-Lizenz verfügbaren Betriebssystem Linux läuft. Dieser neue Server, welcher ausschließlich die WWW-Anfragen bearbeitet, verfügt über einen mit 166 Mhz getakteten Intel-Prozessor (Pentium) sowie 64 MB Hauptspeicher. Nach einigen Anlaufproblemen mit dem Timing des SCSI-Busses konnte der Server, nachdem die notwendigen Modifikationen am Betriebssystemkern (kernel) vorgenommen worden waren, im Herbst dieses Jahres in Betrieb genommen werden.

Es zeigte sich dann sehr rasch, daß seine Leistungsfähigkeit weit über der vergleichbarer Sparc-Rechner liegt. Dies ist vor allem auf die schlanke Architektur des kernels zurückzuführen, hat seine Ursache aber auch in der verwendeten Server-Software für das WWW-Protokoll (httpd). Wir setzen hier einen - von uns leicht modifizierten - Apache-http-Server ein, der sich als äußerst schnell erwiesen hat und nach anfänglichen Stabilitätsproblemen mittlerweile ausgesprochen zuverlässig arbeitet. Im Zusammenspiel mit der eingesetzten Hardware gewährleistet er kurze Antwortzeiten, welche unsere ursprünglichen Erwartungen weit übertroffen haben.

2.4 Das Funktionsprinzip des Verteilten Systems

Mit der Installation des neuen WWW-Servers auf Linux-Basis wurde das Gesamtsystem gleichzeitig auf zwei Rechner verteilt, die im Zusammenspiel die Gesamtfunktionalität des Informationsdienstes gewährleisten. Der Zugriff auf die Bestände der verschiedenen beteiligten Bibliotheken unter IPX und TCP/IP, die Datenkonversion, die automatische Generierung der Indextabellen sowie die Aufbereitung der Nutzungsstatistik erfolgen auf dem Sparc-Rechner. Dieser übergibt die Ergebnisse dann an den WWW-Server, wo sie für einen definierten Zeitraum temporär zwischengespeichert und zur Abfrage genutzt werden. Diese Konfiguration erhöht nicht nur die Performance des Gesamtsystems, sie reduziert darüber hinaus vor allem seine Störanfälligkeit bei gelegentlichen Zugriffsproblemen zu den Daten der beteiligten Bibliotheken bzw. bei sonstigen Netzstörungen, da das WWW-Angebot bis auf weiteres zunächst unbehelligt weiterlaufen kann, während der Arbeitsserver (Sparc) an der Aufbereitung der benötigten Informationen bzw. an der Wiederherstellung der Datenkonsistenz arbeitet.



Funktionsschema des verteilten Systems

Beide Maschinen bilden so gemeinsam den Kern eines nach dem Client-Server-Prinzip verteilten Informationssystems, an das verschiedene Datenbanken über offene Strukturen relativ problemlos angebunden werden können. Der modulare Aufbau der Software findet auf diese Weise seine Entsprechung in einer modular aufgebauten Hardware-Topologie. Diese Anbindung dürfte sich nach der geplanten Implementierung des Z39.50-Standards noch weiter vereinfachen.

2.5 Die Daten der Hessischen Landes- und Hochschulbibliothek

Die Hessische Landes- und Hochschulbibliothek (LHB) nimmt am Hessischen Verbund (HEBIS) der großen wissenschaftlichen Bibliotheken teil, der seit einiger Zeit mit dem niederländischen PICA-System arbeitet. Zur Zeit sind

rund 180.000 Titel der LHB in PICA erfaßt, was etwa 12 % des Gesamtbestandes entspricht. Das an der LHB unter PICA implementierte „Lokalsystem Darmstadt“ stellt das lokale Endglied des Hessischen Bibliotheksverbundes dar, über das die LHB ihre Titeldaten in den Verbund einstellt und zukünftig ihren eigenen Leihverkehr verwalten soll. (Die dezentralen Bibliotheken der TH nehmen hingegen nicht am Hessischen Verbund teil, da sie zum einen in der Regel nicht die an eine solche Teilnahme geknüpften Voraussetzungen erfüllen und da sie andererseits ihre Unabhängigkeit von zentralisierten Strukturen, in der tatsächlich auch ihre Stärke liegt, bewahren wollen. Sie werden somit auch auf absehbare Zeit nicht aktiv am Hessischen Verbundsystem teilnehmen.)

Bis auf weiteres gestaltet sich der Zugang zu den Daten der LHB unter PICA allerdings recht schwierig, ist er doch nur zeichenorientiert mittels Terminalemulation (telnet) bzw. über eine proprietäre Terminal-Emulationssoftware (IBW) möglich. Zusätzlich erfordert die Recherche die Verwendung einer nicht unkomplizierten Abfragesyntax, so daß der Zugang für Endnutzer nur schwer möglich ist. Zwar ist bislang nur ein geringer Teil des Medienbestandes der Landes- und Hochschulbibliothek unter PICA erfaßt, da sie aber über den Hessischen Zentralkatalog zukünftig auch Zugriff auf die Ressourcen der großen deutschen Bibliotheksverbünde erhalten soll, erwarteten wir nach wie vor, daß sich der Erfassungsgrad durch den Import von Fremddaten mittelfristig deutlich erhöhen und so daß ein maßgeblicher Teil der hier vorhandenen Bestände erschlossen sein wird. Eines unserer wichtigen Ziele war es daher, auch die LHB-Daten möglichst rasch in den WWW-Dienst einzubinden.

Allerdings dauerte es schließlich doch einige Zeit, bis uns die Daten endlich zur Verfügung gestellt wurden, da der Export der LHB-Sätze nicht unmittelbar aus dem PICA-Lokalsystem in Darmstadt erfolgte, sondern im Bibliotheksrechenzentrum des Verbundes in Frankfurt vorgenommen werden mußte. Seit Anfang Oktober sind die erfaßten Bestände der LHB über den WWW-Dienst recherchierbar. Allerdings können wir zur Zeit keine laufende Aktualisierung dieses Bestandes vornehmen, da die Frage der Kostenerstattung für den Datenexport an den Verbund noch nicht geklärt werden konnte und wir deswegen von dort keine regelmäßigen Datenlieferungen erhalten.

2.6 Die Öffnung des Systems für externe Bibliotheken

Im Anschluß an die erfolgreiche Konsolidierungsphase des WWW-Informationsdienstes können wir das System zukünftig auch für die Öffentlichen Bibliotheken der Region zum Nachweis ihrer Bestände öffnen. Die Koordination der Anfragen von interessierten Einrichtungen aus diesem Bereich erfolgt durch die Staatliche Büchereinstelle beim Regierungspräsidium Darmstadt, mit der wir seit zwei Jahren kooperieren. Als Voraussetzung für die Teilnahme

müssen Öffentliche Bibliotheken über ein EDV-System zur Bibliotheksverwaltung verfügen, welches den Export strukturierter Daten erlaubt, wobei das Format weitgehend beliebig sein kann.

Als erste größere Einrichtung wurde im Herbst die Stadtbibliothek Neu-Isenburg aufgenommen, die mehr als 70.000 Titelsätze per EDV erfaßt hat und deren Klientel zu einem nicht unerheblichen Teil aus Studenten der THD besteht. Diese Bibliothek arbeitet mit dem System Bibdia der Firma BiBer, so daß der Datenaustausch über das MAB-Format erfolgen kann. Neu-Isenburg gehörte damit bundesweit zu den ersten Kommunen, welche die Bestände ihrer Bibliothek über das Internet verfügbar machten. Weitere kommunale Einrichtungen aus der Region werden in Kürze folgen.

2.7 Die Auslastung des Systems

Seit der Aufnahme des Realbetriebs hat die Zahl der Suchanfragen stetig zugenommen. Im Zeitraum 1. August bis 20. Oktober 1996 lag die durchschnittliche Zahl der Recherchen pro Tag (inkl. Wochenenden und Feiertage) bei 1.595 Abfragen, was einer Stundenrate von 66,5 entspricht. An Werktagen liegt die tägliche Nutzungsrate zur Zeit bei rund 2.000 Abfragen. Im genannten Zeitraum wurden insgesamt 127.562 Recherchen von 5.497 verschiedenen Rechnern aus durchgeführt. Der mit 78 % größte Nutzungsanteil entfiel dabei naturgemäß auf Einrichtungen der THD, während 14 % der Zugriffe von anderen deutschen Hochschulen aus erfolgten. Die restlichen Anteile entfallen vorwiegend auf internationale wissenschaftliche Organisationen sowie einige kommerzielle Einrichtungen.

3. Perspektiven

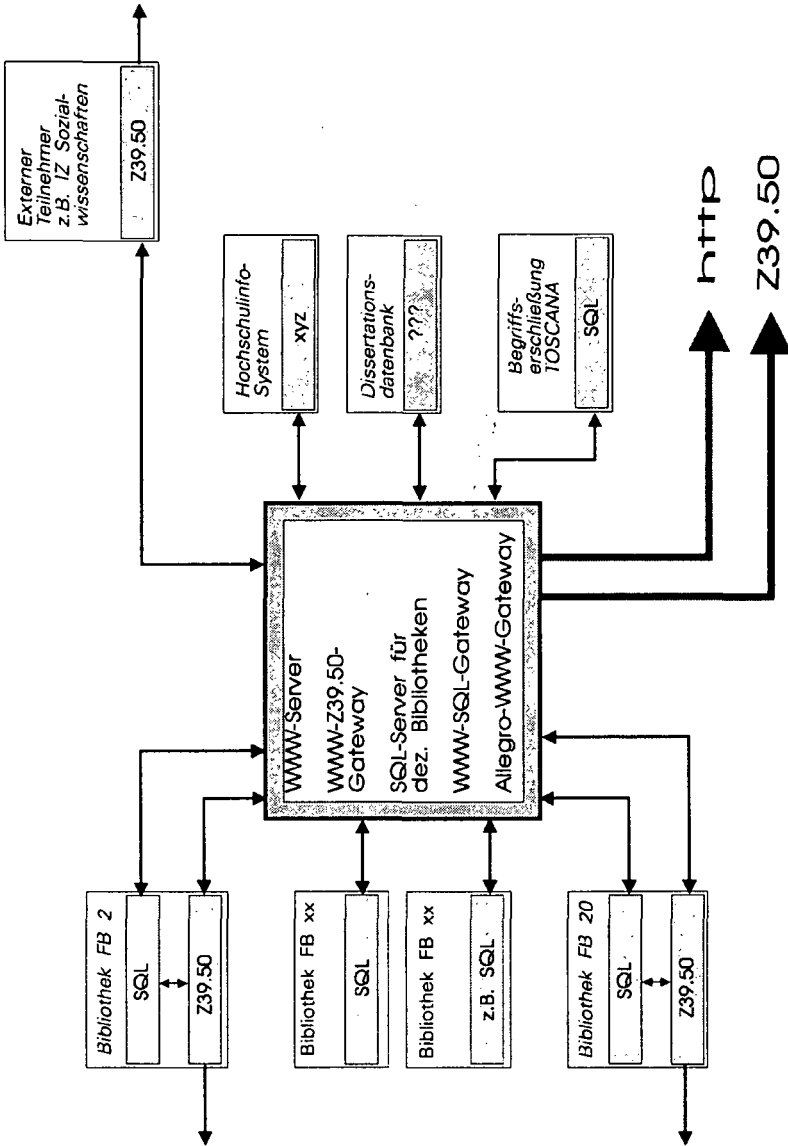
Der bestehende Dienst soll im nächsten Jahr weiter ausgebaut und optimiert werden. Neben der Integration weiterer Datenbestände ist vor allem die regelmäßige Aktualisierung der PICA-Daten aus der LHB sicherzustellen. Dazu sind die Modalitäten des Datenaustauschs sowie die anhängigen Kostenfragen baldmöglichst zu klären, und der Datentransfer ist entsprechend zu institutionalisieren.

Nachdem sich der Dienst mittlerweile innerhalb der TH wie auch im Internet stabil etabliert hat, muß schließlich die Systembasis weiter verbessert werden. Die wichtigsten Voraussetzungen dafür liegen in der Verwendung eines leistungsfähigen Datenbanksystems, das in seiner Performance wie in seinem Funktionsumfang über die bislang verwendete Allegro-Datenbank hinausgeht und insbesondere SQL-Abfragen erlaubt. Diese Umstellung könnte mit überschaubarem Kostenaufwand im Lauf des nächsten Jahres erfolgen.

Darüber hinaus soll mittelfristig eine Schnittstelle nach Z39.50 implementiert werden, um die Kommunikation mit anderen Informationsdiensten im internationalen Forschungs- und Kooperationsraum zu verbessern. Dies ist besonders auch im Hinblick auf die absehbaren Entwicklungstendenzen dokumentenbezogener Informationssysteme notwendig, welche zukünftig sich nicht mehr bloß mit dem Nachweis bibliographischer Daten werden begnügen können, sondern zunehmend vor die Anforderung gestellt werden dürften, die sachliche Erschließung und Bereitstellung multimedialer Dokumente zu realisieren.

Diese Fragen, die immer stärker in den Vordergrund rücken, erfordern jedoch zunächst umfangreiche konzeptionelle Vorarbeiten, deren Ergebnisse zum Ausbau des bestehenden Dienstes zu einem universalen Informationserschließungs- und Retrievalsystem führen sollen (vgl. dazu den in Anlage 1 dargestellten Konzeptansatz). Die Basis dafür ist mit dem bestehenden Informationsangebot geschaffen.

Anlage 1



Wenzel 07/96

