

Mit dem „Surfbrett“ in die Bibliothek

Der World-Wide-Web-Katalog der Bibliothek der Friedrich-Ebert-Stiftung

Walter Wimmer

1. Die Ausgangslage vor Projektbeginn

Die Bibliothek der Friedrich-Ebert-Stiftung bietet ihren ALLEGRO-Katalog seit Mitte 1995 im Internet an. (Adresse: http://www-fes.gmd.de/library/index_gr.html) Der folgende Artikel soll die Vorgehensweise und die gemachten Erfahrungen verdeutlichen.

Ende 1994 wurde beschlossen, die Tätigkeit der Friedrich-Ebert-Stiftung mit Hilfe eines WWW-Servers im Internet darzustellen. Auf diesem Server sollten auch die Arbeit der Bibliothek und insbesondere der Bibliothekskatalog angemessen präsentiert werden.

Gleichzeitig wurde von der niederländischen Firma DMP eine Retrokonversion der Zettelkataloge durchgeführt. Somit lag der gesamte Bibliothekskatalog in maschinenlesbarer Form vor. Diese Daten werden mit dem Bibliothekssystem ALLEGRO verwaltet, das die Bibliothek der Friedrich-Ebert-Stiftung seit Anfang 1993 einsetzt.

Der WWW-Server der Friedrich-Ebert-Stiftung wird unter Linux, einem UNIX-Betriebssystem für PCs betrieben. Eine Linux-Version des ALLEGRO-Programmpaketes stand ab Mai 1995 zur Verfügung. ALLEGRO ist zur Zeit das einzige Bibliothekssystem, das unter Linux verfügbar ist. Es lag also nahe, den geplanten Internet-Katalog mit Hilfe von ALLEGRO zu verwirklichen.

2. Telnets- oder World-Wide-Web-Katalog ?

Zunächst mußte grundsätzlich entschieden werden, ob der Datenbestand über einen Telnets- oder einen WWW-Katalog verfügbar gemacht werden sollte. Bei einem Katalog, der über Telnets angeboten wird, arbeitet der Benutzer interaktiv mit dem Programm. Er wird also mit einer Programmoberfläche konfrontiert, die ihm zunächst unbekannt ist. Dies zieht einen entsprechend hohen Erklärungsbedarf nach sich.

Die Benutzerschnittstelle eines WWW-Kataloges basiert dagegen auf Formularen, die mit Hilfe der im WWW generell verwendeten Dokumentbeschreibungssprache HTML erstellt werden. Es werden also standardisierte Elemente benutzt, die im Grunde im gesamten World-Wide-Web identisch sind.

Wegen dieser wesentlich höheren Benutzerfreundlichkeit sollte einem allgemeinen Trend folgend ein WWW-Katalog realisiert werden.

3. Konzeption und Realisierung des Projektes

Bei einem WWW-Katalog wird die vom Benutzer in ein Suchformular eingegebene Anfrage an den WWW-Server weitergeleitet, der sie seinerseits an ein sogenanntes CGI-Gateway zur Abarbeitung übergibt. Das „*Common Gateway Interface*“ (CGI) ist ein Standard, der die Struktur der vom WWW-Server übergebenen und an diesen zurückgelieferten Daten festlegt.

Von seiner Grundkonzeption her unterstützt ALLEGRO eine solche Client-Server-Lösung nicht. Es verfügt auch über keine standardisierte Abfragesprache wie SQL, in die die vom WWW-Server weitergeleitete Suchanfrage übersetzt werden könnte.

Anfang 1995 existierten bereits Lösungen für dieses Problem (zum Beispiel an der UB Braunschweig und der Bibliothek der RWTH Aachen), die von uns adaptiert, aber zu einer durchaus eigenständigen Anwendung weiterentwickelt wurden.

ALLEGRO bietet die Möglichkeit, verschiedene Datensatztypen zu verknüpfen. Diese Verknüpfungen werden über Einträge in der Indexdatei der Datenbank realisiert.

Beim Export von Datensätzen aus der Datenbank ist eine Auflösung von Verknüpfungen also zwangsläufig mit Festplattenzugriffen verbunden, die die Geschwindigkeit der Recherche beeinträchtigen. Um die Zugriffszeit zu optimieren, werden bestehende Verknüpfungen bereits bei der Umwandlung der auf einem Novell-Server vorliegenden Originaldaten für die unter Linux betriebene Datenbank aufgelöst. Dies betrifft Band- und Serienstammsätze, aber auch Verweisungsformen aus Personennormsätzen u. ä., die in den Titelsatz integriert werden. So werden nur wenige deutsche Internet-Nutzer Herrn Gorbatschow unter *Gorbacev* suchen. Die dabei entstehende Datenredundanz, die natürlich Festplattenplatz kostet, wird bewußt in Kauf genommen.

Zusätzlich wird noch eine Umsetzung vom ASCII-Zeichensatz in den unter Linux verwendeten ANSI-Zeichensatz durchgeführt, obwohl dies bei einem nicht interaktiv genutzten System eigentlich nicht erforderlich wäre, da Umlaute und Diakritika bei der Ausgabe in das HTML-Format durch Ersatzfrequenzen dargestellt werden müssen.

Für die Recherchedatenbank wurde ein Kreuzindex erstellt, in dem die Indexeinträge durch spezielle Präfixe einzelnen Suchaspekten zugewiesen werden (beispielsweise au=ebert, friedrich; st=stiftung; ej=1994). Dies erleichtert die

Programmierung derjenigen Teile des CGI-Gateways, die den Datenbankindex auswerten.

Diese Modifikationen konnten mit Hilfe der in ALLEGRO implementierten Datenmanipulationssprache durchgeführt werden, deren Befehle in sogenannten Parameterdateien gesammelt und von den eigentlichen ALLEGRO-Programmen umgesetzt werden.

Das CGI-Gateway-Programm basiert auf einer Funktion des ALLEGRO-Freitext-Suchprogramms SRCH, das durch sogenannte Nachladebefehle dazu veranlaßt werden kann, den Index einer beim Programmaufruf spezifizierten Datenbank auszuwerten. Zur Erstellung des Gateways muß der Aufruf des ALLEGRO-Programms SRCH in ein spezielles Programm eingebunden werden, das eine Umgebung generiert, in der SRCH sinnvoll arbeiten kann.

Dieses Programm wurde in der im Internet vielfach benutzten Skriptsprache PERL geschrieben. Es könnten jedoch auch andere Programmiersprachen verwendet werden. PERL-Skripten werden erst zur Laufzeit interpretiert und teilweise kompiliert. Dies kostet Rechenzeit. Auf der anderen Seite unterstützt PERL die Verarbeitung von Textdaten, die im ALLEGRO-Format vorliegen, sehr gut. Es ist sowohl unter UNIX als auch unter MS-DOS verfügbar. Dadurch können PERL-Skripten auf beiden Betriebssystemen genutzt werden. Nicht zuletzt ist PERL leichter erlernbar als etwa C.

4. Eine Benutzeranfrage wird bearbeitet

Eine Benutzeranfrage an den WWW-Katalog wird vom WWW-Server an das CGI-Gateway weitergeleitet. Das PERL-Programm extrahiert zunächst die einzelnen Suchbegriffe und eventuell vorhandene andere Festlegungen (z. B. die maximale Treffermenge) aus der übergebenen Zeichenkette.

Nun werden zwei temporäre Dateien erzeugt, die vom Programm SRCH ausgewertet werden und dessen Verhalten steuern. Eine Datei im ALLEGRO-Datenformat enthält in einem einzigen Datensatz Informationen darüber, welcher Suchbegriff zur Auswertung des Datenbankindexes herangezogen werden soll und wieviele Suchbegriffe zu welchen Suchaspekten überhaupt existieren. Diese Datei wird von SRCH pro forma durchsucht, da es als Freitext-Suchprogramm eine Datei benötigt, die durchsucht werden kann. De facto werden jedoch über den festgelegten Suchbegriff Datensätze aus der Datenbank nachgeladen. Zum Nachladen werden zuerst Suchaspekte mit einer durchschnittlich geringeren Treffermenge benutzt, also z. B. Autoren vor Titelstichwörtern.

In einer zweiten Datei werden die variablen Teile einer Parameterdatei generiert, die vom Programm SRCH für jeden gefundenen Datensatz abgearbeitet

wird. Dabei werden diese Datensätze daraufhin überprüft, ob alle weiteren Suchbegriffe enthalten oder - bei einer Negation von Suchbegriffen - nicht enthalten sind. Nur wenn alle Suchbedingungen erfüllt sind, wird der Datensatz in einer weiteren temporären Datei abgespeichert. Die Datensätze werden sortiert und vom PERL-Programm zu Titelaufnahmen aufbereitet, die mit HTML-Befehlen formatiert sind.

Die aufbereiteten Daten werden an den WWW-Server und von diesem an den Benutzer weitergeleitet. Alle temporären Dateien werden abschließend wieder gelöscht.

Bei der Formulierung von Suchanfragen ist innerhalb eines Suchaspektes die Verknüpfung beliebig vieler Suchbegriffe mit den drei Booleschen Operatoren möglich. Falls der Benutzer keinen speziellen Verknüpfungoperator (/ für oder, ! für nicht) angibt, wird die Schnittmenge der Suchbegriffe gebildet. So wird beispielsweise die Suche nach einem Sachtitel als „Und-Verknüpfung“ der einzelnen Titelstichwörter realisiert. Bei der Suche nach Jahreszahlen können darüber hinaus Vergleichsoperatoren (größer, größer gleich usw.) eingesetzt werden.

Leider ist bisher noch keine Klammerung von Suchbegriffen möglich. Dies führt zu Fehlinterpretationen bei der „Oder-Verknüpfung“ idiomatischer Wendungen, beispielsweise „Weimarer Republik“ / (oder) „Deutsches Reich“.

5. Erfahrungen und Desiderate

Die PERL-Skripten und Parameterdateien konnten im skizzierten Leistungsumfang bereits unter MS-DOS/Windows 3.11 realisiert werden, bevor die ALLEGRO-Version für Linux vorlag. Bei der Umsetzung auf das UNIX-Betriebssystem Linux war der Kampf mit „Kinderkrankheiten“ der portierten ALLEGRO-Programme recht zeitraubend. Diese wurden aber in der Zwischenzeit von den Programm-Entwicklern beseitigt.

Die Bibliothek der Friedrich-Ebert-Stiftung hat ein elektronisches Gästebuch eingerichtet, um Reaktionen und Anregungen der Benutzer des WWW-Kataloges zu sammeln. Diese Benutzerreaktionen sind insgesamt sehr positiv. Zahlenmaterial über die Benutzungsfrequenz existiert jedoch noch nicht.

Wegen der Retro-Konversion der Zettelkataloge, bei der alle Ungereimtheiten der Kataloghistorie übernommen wurden, differiert die Qualität der in der Datenbank erfaßten Daten sehr stark. Insbesondere konnten bei der Konversion keine Normsätze verwendet werden. Dies beeinträchtigt die Erstellung von abrufbaren Suchlisten für normierte Inhalte, wie beispielsweise Schlagworte, sehr stark. Zur Zeit kann der Benutzer zwar nach normierten Inhalten

suchen, erhält aber keinen Überblick über das Wortmaterial, das ihm zur Verfügung steht.

6. Mögliche zusätzliche Funktionen eines WWW-Kataloges

Wenn die beschriebenen Abfrageroutinen vorliegen, ist es relativ einfach, das zurückgelieferte Suchergebnis mit weiteren Funktionalitäten anzureichern, indem man zusätzliche HTML-Befehle einbindet.

Gefundene Datensätze können als Teil eines weiteren Suchformulars formatiert werden. Dadurch wird es möglich einzelne Titel zu markieren und in ein Bestellformular zu übernehmen. Auf diese Art konnte mit Hilfe einer ALLEGRO-Datenbank ein Online-Bestellsystem für die von der Friedrich-Ebert-Stiftung herausgegebenen Publikationen verwirklicht werden.

In einer ALLEGRO-Datenbank können die Dateinamen von Dokumenten erfasst werden, die mit der jeweiligen Titelaufnahme verknüpft werden sollen. Hierbei kann es sich um Texte (Abstracts, Volltexte) aber auch um Bild- und Tondokumente handeln. Bei der Ausgabe von Rechercheergebnissen können aus diesen Informationen Hyperlinks auf diese Quellen erzeugt werden, denen der Benutzer folgen kann. Dadurch kann ein ALLEGRO-Katalog sehr einfach zu einem Multimedia-Katalog weiterentwickelt werden. Bei uns wird dieses Verfahren bisher nur zur Einbindung von Abstracts genutzt.

7. Wie soll es weitergehen ?

Um die gefundene Literatur zu erhalten, wird der Benutzer bisher auf die Fernleihe verwiesen. Angesichts des schnellen Literaturnachweises liegt es nahe den Internet-Katalog um ein Bestellsystem für das Online-Document-Delivery von Zeitschriftenaufsätzen zu erweitern. Hierzu wurden erste Vorüberlegungen angestellt. Zunächst ist daran gedacht, bestellte Aufsätze zu kopieren und mittels Fax oder Brief zu übermitteln. Später könnten die Aufsätze gescannt und über E-Mail versandt werden. Alternativ wäre die Bereitstellung auf einem FTP-Server denkbar.

Der WWW-Katalog der Bibliothek der Friedrich-Ebert-Stiftung verdeutlicht einmal mehr die hohe Flexibilität von ALLEGRO. Dadurch konnte in Eigenarbeit - ohne die kostspielige Unterstützung durch ein Softwarehaus - ein Internet-Katalog entwickelt werden, der den Vergleich mit ähnlichen Angeboten im World-Wide-Web nicht zu scheuen braucht.

