



Fachhochschule Konstanz

Hochschule für Technik, Wirtschaft und Gestaltung
Fachbereich Informatik

**Die Schlagwortnormdatei als Erschließungsinstrument für
Online-Publikationen**

Björn Eggers

Konstanz, 31. Januar 2001

Diplomarbeit

Diplomarbeit

Zur Erlangung des akademischen Grades

Diplom-Informatiker (FH)

an der

Fachhochschule Konstanz

- Hochschule für Technik, Wirtschaft und Gestaltung -
im Fachbereich Informatik / Wirtschaftsinformatik

Thema: Die Schlagwortnormdatei als Erschließungsinstrument für
Online-Publikationen

Diplomand: Björn Eggers
Freiherr vom Stein Str. 54
78315 Radolfzell

Betreuer: Prof. Dr. Werner Lötzbeyer
Fachhochschule Konstanz

Dipl.-Verw.Wiss. Thomas Frings
Bibliotheksservice-Zentrum Baden-Württemberg

Firma: Bibliotheksservice-Zentrum Baden-Württemberg
Universität Konstanz
78457 Konstanz

Abgabetermin: 31. Januar 2001

Ehrenwörtliche Erklärung

Hiermit erkläre ich, Björn Eggers, geboren am 09. November 1968 in Hildesheim, daß ich meine Diplomarbeit mit dem Titel:

„Die Schlagwortnormdatei als Erschließungsinstrument von Online-Publikationen“

selbständig angefertigt habe. Ich habe keine anderen, als die in der Diplomarbeit angegebenen Hilfsmittel benutzt.

Konstanz, 31. Januar 2001

Björn Eggers

Danksagung

Ich möchte mich bei meinen Eltern bedanken, die mir dieses Studium ermöglichten.

Desweiteren gilt mein Dank Herrn Dipl.-Verw.Wiss. Thomas Frings und Herrn Dipl.Bibl.(FH) Stefan Wolf für die immer bereitwillige Weitergabe von Fachwissen.

Ebenso gilt mein Dank Herrn Prof. Dr. Werner Lötzbeyer für die Betreuung der Diplomarbeit.

Zusammenfassung

- Thema:** Die Schlagwortnormdatei als Erschließungsinstrument für
Online-Publikationen
- Diplomand:** Björn Eggers
Freiherr vom Stein Str. 54
78315 Radolfzell
- Betreuer:** Prof. Dr. Werner Lötzbeyer
Fachhochschule Konstanz

Dipl.-Verw.Wiss. Thomas Frings
Bibliotheksservice-Zentrum Baden-Württemberg
- Firma:** Bibliotheksservice-Zentrum Baden-Württemberg
Universität Konstanz
78457 Konstanz
- Abgabetermin:** 31. Januar 2001

Die Schlagwortnormdatei stellt den gemeinsam gepflegten Thesaurus für die verbale kooperative Sacherschließung der wissenschaftlichen Bibliotheken im deutschsprachigen Raum dar. Für die Metadatenerschließung von Online-Publikationen durch die Verfasser selbst muß sie in geeigneter Weise im Internet zur Verfügung gestellt werden. Diese Lücke schließt dieses Projekt: die bislang nur in bibliothekarischen Eigenformaten angebotene Normdatei wird in einer WWW-fähigen Form aktuell, vollständig und komfortabel recherchierbar angeboten. Die technische Realisierung beruht auf SunOS als Betriebssystem, einer CGI-Schnittstelle, Perl als Programmiersprache und MySQL als unterlegtes relationale Datenbanksystem. Alternativen der Programmierung (JAVA etc.) werden dargestellt.

Inhaltsverzeichnis

INHALTSVERZEICHNIS	1
ABBILDUNGSVERZEICHNIS.....	3
TABELLENVERZEICHNIS.....	3
ABKÜRZUNGSVERZEICHNIS	4
1 EINLEITUNG	5
2 BIBLIOTHEKARISCHE ASPEKTE	7
2.1 LITERATURERSCHLIEßUNG.....	7
2.2 DIE SCHLAGWORTNORMDATEI (SWD)	9
2.3 ONLINE-PUBLIKATIONEN	12
2.4 SÜDWESTDEUTSCHES ONLINE PUBLIKATIONSSYSTEM (SWOP).....	13
2.4.1 Einbringen von Dokumenten.....	13
2.4.2 Übernahme in die SWOP-Datenbank.....	15
2.4.3 Weitere Verwendung der Metadaten.....	15
2.4.4 Recherchieren in SWOP.....	16
2.4.5 Datenfluß in SWOP bzw. OPUS.....	17
3 DIE ENTWICKLUNGSUMGEBUNG	18
3.1 DIE CGI-SCHNITTSTELLE.....	18
3.1.1 Technische Anforderungen an eine CGI-Schnittstelle.....	22
3.2 ALTERNATIVEN ZU CGI.....	26
3.2.1 Java Servlets	26
3.2.2 PHP.....	29
3.2.3 Proprietäre Server-API.....	30
4 DIE ARCHITEKTUR DES PROTOTYPS	31
4.1 DAS DATENMODELL	31
4.2 DIE TABELLENSTRUKTUREN	32
4.3 ADMINISTRATIONSTOOLS.....	34
4.3.1 Das Tool dataextract.....	34
4.3.2 Das Tool dataloader	36

5	DIE SOFTWARETECHNISCHE REALISIERUNG.....	37
5.1	HTML-TEMPLATES	37
5.2	DIE VERSCHIEDENEN SUCHMODIS	37
5.2.1	<i>Exakte Suche</i>	38
5.2.2	<i>Index Suche</i>	39
5.2.3	<i>Teilwortsuche</i>	40
5.3	AUFGABENBESCHREIBUNG DER PROGRAMME.....	42
5.3.1	<i>oswd-suchabfrage.pl</i>	42
5.3.2	<i>oswd-trefferanzeige.pl</i>	43
5.3.3	<i>oswd-einzelanzeige.pl</i>	45
	QUELENNACHWEIS.....	46

Abbildungsverzeichnis

ABB 1.:	DATENFLUß IN SWOP BZW. OPUS	17
ABB 2.:	ABLAUF IN EINER CGI-UMGEBUNG	19
ABB 3.:	ABLAUF IN EINER CGI-UMGEBUNG MIT DATENBANKANBINDUNG	20
ABB 4.:	MARKTANTEILE WEBSERVER, STAND DEZEMBER 2000.....	25
ABB 5.:	VOM JAVA-QUELLCODE ZUR PROGRAMMAUSFÜHRUNG.....	27
ABB 6.:	KOMMUNIKATION MIT SERVLETS	27
ABB 7.:	DAS DATENMODELL	31
ABB 8.:	TABELLENSTRUKTUR BEGRIFFE	32
ABB 9.:	TABELLENSTRUKTUR ANSETZUNGEN.....	32
ABB 10.:	TABELLENSTRUKTUR ISA_SYNONYM	33
ABB 11.:	TABELLENSTRUKTUR ISA_ VERWANDTSCHAFT.....	33
ABB 12.:	TABELLENSTRUKTUR ISA_ OBERBEGRIFF.....	33
ABB 13.:	SCHEMATISCHER ABLAUF VON HTML-TEMPLATES	37
ABB 14.:	BILDSCHIRM-MASKE OSWD-SUCHABFRAGE.PL.....	42
ABB 15.:	WARNMELDUNG ÜBER FEHLGESCHLAGENE SUCHE.....	43
ABB 16.:	BILDSCHIRM-MASKE OSWD-TREFFERANZEIGE.PL.....	43
ABB 17.:	NAVIGATIONSLEISTE EINES BROWSERS	44
ABB 18.:	BILDSCHIRM-MASKE OSWD-EINZELANZEIGE.PL.....	45

Tabellenverzeichnis

TAB. 1:	BEISPIELEINGABEN FÜR FORMULARDATEN IN SWOP	14
TAB. 2:	ERLÄUTERUNG DER UMGEBUNGSVARIABLEN.....	22

Abkürzungsverzeichnis

AK	Alphabetischer Katalog
API	Application Programming Interface
CGI	Common Gateway Interface
DBI	Deutschen Bibliotheksinstitut
HTML	Hyper Text Mark Langague
ISAPI	Internet Server Application Programming Interface
ISO	International Standard Institution
JSDK	Java Servlet Development Kit
MAB2	Maschinelles Austauschformat für Bibliotheken 2
NSAPI	Netscape Server Application Programming Interface
OPUS	Online Publikationssystem der Universität Stuttgart
OS	Operating System
PHP	Personal Home Page Tools
RSWK	Regeln für den Schlagwortkatalog
SWB	SüdWestdeutscher Bibliotheksverbund
SWD	Schlagwortnormdatei
SWK	Schlagwortkatalog
SWOP	SüdWestdeutsches Online Publikationssystem
SyK	Systematische Katalog
WWW	World Wide Web

1 Einleitung

Das rapide Anwachsen der Internetnutzung stellt die Bibliotheken vor neue Herausforderungen: sowohl die Integration dieses Angebots in die vorhandene Informationsinfrastruktur als auch die Verbindung der Beschreibung konventioneller Medien und (wenigstens) der angebotenen Online-Ressourcen des WWW ist zu bewältigen.

Sicher ist, dass traditionelle Katalogisierung in stärker rationalisierter Form realisiert werden muß; dazu gehört konstitutiv die kooperative Nutzung vorhandener Erschließung, die Entwicklung gemeinsamer Instrumente sowie die Mehrfachverwendung einmal erfaßter Daten. Bei der bibliothekarischen verbalen Sacherschließung bedeutet dies die Übernahme institutionsfremder Erschließungsleistung genau so wie die Verwendung eines gemeinsamen Erschließungsinstruments, der Schlagwortnormdatei (SWD) als Thesaurus. Soll – wie es die verschiedenen Metadateninitiativen postulieren – die initiale Beschreibung von Online-Publikationen durch die Autoren selbst erfolgen, müssen die entsprechenden Werkzeuge zur Verfügung stehen. Für die Inhaltserschließung bedeutet dies, dass die SWD möglichst vollständig, aktuell und einfach im WWW nutzbar angeboten werden.

Diese wird bislang über ein Z39.50-Gateway im Katalog der Deutschen Bibliothek zusammen mit den Titeldaten im bibliothekarischen MAB2-Format angeboten; dieses sind zwar akzeptierte Standard, doch ist die Form des Angebots und die Nutzbarkeit durch verschiedene Eigenheiten nach Meinung der mit dem Sachverhalt befaßten Bibliothekare stark beeinträchtigt.

Ziel des Projekt war es deshalb, im WWW für die Zwecke der Metadatenerschließung eine Datenbank aufzubauen, in die die SWD übertragen werden kann. Zur Bewältigung der Aufgabe war wesentlich das Verständnis der bibliothekarischen Gegebenheiten und Verwendungszwecke, wie es in Kap. 2 dargestellt ist. Zu klären war dann die Struktur der Datenbank, die sachgerechten Suchbedingungen samt Indexaufbereitung und Sortierung, sowie die regelmäßig wiederholbare Transformation der im bibliographischen MAB2-Format vorliegenden Ausgangsdaten (samt Zeichensatz) in die verwendete MySQL-Datenbank.

Kapitel 5 ist gewidmet der Darstellung, der verschiedenen Suchmöglichkeiten, die auf die erzeugte Datenbank aufsetzen und aus bibliothekarischer Sicht notwendig sind.

2 Bibliothekarische Aspekte

2.1 Literaturschließung

Literaturschließung ist eine der Hauptaufgaben von Bibliotheken. Unter Literaturschließung versteht man die Beschreibung des Literaturbestandes nach formalen oder inhaltsbezogenen Kriterien. Diesen Vorgang nennt man auch Katalogisierung.

Ein Katalog ist ein Verzeichnis, das nach alphabetischen oder sachlichen Gesichtspunkten geordnet ist. Als Beispiele für Kataloge können Warenhauskataloge, Briefmarkenkataloge oder Münzkataloge genannt werden. Ein Bibliothekskatalog ist ein geordnetes Verzeichnis über den Medienbestand einer Bibliothek.

Laut [Hack92] gliedert man Bibliothekskataloge nach ihrer inneren Ordnung in vier Hauptarten:

- Alphabetischer Katalog
- Schlagwortkatalog
- Systematischer Katalog
- Standortkatalog

So läßt sich durch einen Bibliothekskatalog feststellen, ob z.B. ein bestimmtes Buch in einer Bibliothek vorhanden ist (Alphabetischer Katalog), welche Werke zu einem Sachgebiet angeboten werden (Schlagwort/Systematischer Katalog) oder an welchem Ort die Medien aufgestellt sind (Standortkatalog).

Die Funktionen dieser verschiedenen Kataloge läßt sich so beschreiben:

a) Der Alphabetische Katalog (AK)

Im Alphabetischen Katalog sind die in einer Bibliothek vorhandenen Medien nach formalen Gesichtspunkten in einer alphabetischer Reihenfolge verzeichnet. Formale Kriterien sind vor allem:

- der Verfassername,
- der Sachtitel
- der Name einer beteiligten Körperschaft.

Im allgemeinen ist der Name „Alphabetischer Katalog“ gebräuchlich, da aber Schlagwortkataloge ebenfalls alphabetisch geordnet sind, liegt in der verwendeten Sprache eine gewisse Mehrdeutigkeit vor. Aus diesem Grund wird dieser Katalog auch oftmals Autorenkatalog, Verfasserkatalog oder auch z.B. in der Schweiz Nominalkatalog genannt.

b) Der Schlagwortkatalog (SWK)

Der Schlagwortkatalog gehört zur Gruppe der inhaltserschließenden Kataloge. Im Gegensatz zu einem Alphabetischen Katalog erschließt ein Sachkatalog nicht nach formalen Gesichtspunkten, sondern nach inhaltlichen Aspekten. Der SWK soll somit die Frage beantworten, welche Bücher in einer Bibliothek über bestimmte Themen oder Gegenstände vorhanden sind. Der Überblick auf bestimmte Autoren oder Sachtitel ist hier nicht angestrebt.

Die einzelnen Werke werden mit Schlagwörtern beschrieben, die den Inhalt eines Buches widerspiegeln. Diese Wiedergabe soll nach Möglichkeit ziemlich kurz, genau und vollständig sein, d.h. den behandelten Gegenstand z. B. des Buches exakt treffen. Die Auflistung der Schlagwörter im SWK ist alphabetisch. Ein Buch kann durch mehrere Schlagwörter in Form einer sogenannten Schlagwortkette beschrieben sein. Durch die alphabetische Ordnung kann sachlich Zusammengehöriges deshalb auseinandergerissen werden. Im Falle eines unterlegten Thesaurus wird dies aber durch die angebotenen Verweisungsapparate gemildert. Der Schlagwortkatalog ist leicht zu handhaben

und kann im Prinzip mit einem Lexikon, das ebenfalls eine alphabetische Aufstellung von Sachbegriffen enthält, verglichen werden.

c) Systematische Katalog (SyK)

Der SyK ist ebenfalls ein Sachkatalog. In ihm sind die Bücher nach inhaltsbezogenen Aspekten nach einem System der Wissenschaften verzeichnet, die nach sachlich-logischen Zusammenhängen geordnet sind. Dies bedeutet, dass Hauptgebiete, wie z.B. Naturwissenschaften oder Sozialwissenschaften immer weiter in Untergebiete aufgeteilt werden. Als Beispiel könnten die Naturwissenschaften in die Untergebiete Chemie und Physik gegliedert werden. Zu beachten ist, dass Untergebiete wiederum unterteilt werden können und so eine fein abgestimmte Hierarchie entsteht, die sachlich zusammengehörende Literatur zusammenführt.

d) Der Standortkatalog

Dieser Kataloge dient als Nachweis der Orte, der angebotenen Medien nach ihrem Standort in der Bibliothek: er spiegelt die Stellung der Werke im Regal wieder. Ausschlaggebend für die innere Ordnung des Standortkataloges ist somit die Art der Reihenfolge von Büchern im Regal. Werden die Werke z.B. alphabetisch in die Regale eingeordnet, so handelt es sich beim Standortkatalog ebenfalls um einen alphabetischen Katalog; werden sie nach der Reihenfolge des Erwerbs aufgestellt, handelt es sich um eine sog. „Numerus currens“-Aufstellung.

Mit dieser Übersicht der einzelnen Katalogarten zeigt sich die bibliothekarische Relevanz der Schlagwortkataloge.

2.2 Die Schlagwortnormdatei (SWD)

Um einen möglichst großen Nutzen aus einem Schlagwortkatalog zu ziehen, muß eine einheitliche Vergabe von Schlagwörtern gewährleistet sein. Durch vergleichbare Schlagwortkataloge in den diversen Bibliotheken wird z.B. der Austausch von

Schlagwörtern und Titlerschließung ermöglicht. Dies erreicht man durch die Verwendung von normierten Schlagwörtern, deren Bildung, Verwendung und Einordnung in Verweisungs- und Hierarchieapparat einheitlich geregelt ist. Aus diesem Grunde wurde seit Mitte der 1980er Jahre die Schlagwortnormdatei aufgebaut. Die SWD ist inzwischen eine Gemeinschaftsarbeit von u. a. folgenden Institutionen:

- Die Deutsche Bibliothek
- Bibliotheksverbund Bayern
- Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen
- Südwestdeutscher Bibliotheksverbund
- Gemeinsamer Bibliotheksverbund (Göttingen)
- Verbund der wissenschaftlicher Bibliotheken Österreichs
- Schweizerische Landesbibliothek
- Kooperativer Bibliotheksverbund Berlin-Brandenburg
- Kunstverbund Florenz, München, Rom

Die oben genannten Institutionen arbeiten kooperativ zusammen, d.h. jede teilnehmende Institution hat das Recht und die Pflicht, neue Schlagwörter einzubringen, Korrekturen und Änderungen zu verlangen. Damit verbunden ist eine laufende Änderung und Ergänzung des vorhandenen kontrollierten Schlagwortbestandes. Durch die Vielzahl der SWD-Partner ist die Schlagwortnormdatei für die Sacherschließung im deutschsprachigen Raum bestens geeignet.

Bei der Pflege der Schlagwortnormdatei wird auf die Regeln für den Schlagwortkatalog (RSWK) zurückgegriffen. Die RSWK wurden Anfang der 1980er Jahre von der Kommission für Sacherschließung des Deutschen Bibliotheksinstituts (DBI) erarbeitet. Nach den ersten beiden Auflagen in den Jahren 1986 und 1991 liegt seit 1998 die dritte Auflage vor, die endlich Aufbau und Struktur der SWD explizit regelt; dazu gehört besonders:

- die SWD folgt der DIN 1463 Richtlinien für die Erstellung und Weiterentwicklung von Thesauri. - Berlin : Beuth-Verlag, 1987;
- der verwendete Zeichensatz (ISO 646 bzw. ISO ISO 5426-1983) ist bedeutend umfangreicher als der im WWW übliche ISO 8559-1 ("Latin-1")
- Homonyme und Polyseme sind unterschieden
- die sprachliche Formulierung von Schlagwörtern folgt dem Prinzip, daß die gebräuchlichste Bezeichnung verwendet wird, in der Regel ein Wort der deutschen Sprache;
- eine Schlagwortansetzung oder -verweisung kann aus einem Wort, mehreren Wörtern, oder Gliedern bestehen (mehrgliedrige Schlagwörter);
- Verweisungen strukturieren den Wortschatz der SWD und bieten zusätzliche vielfältige Sucheinstiege. Diese Verweisungen sind hauptsächlich möglich
 - Synonym- bzw. Äquivalenzverweisungen
 - Verweisungen, die chronologische Beziehungen zum Ausdruck bringen (früher-/später-Verweisungen)
 - Verweisungen, die hierarchische Beziehungen zum Ausdruck bringen (übergeordnete / untergeordnete Schlagwörter)
 - Verweisungen, die verwandte Schlagwörter in Beziehung setzen (assoziative Verweisungen)
- die Haltung von Ansetzungen, Verweisungen und weiteren Erläuterungen erfolgt in gemeinsamen Datensätzen.

Derzeit besteht die SWD aus ca 480.000 Schlagwortsätzen mit knapp 1.000.000 Ansetzungen und Verweisungen, die im Projekt ausgewertet wurden, bei einem jährlichen Zuwachs von ca. 35.000 Datensätzen zzgl. umfangreicher Korrekturen und Ergänzungen bestehender Datensätze. Eine Reihe von Verweisungsarten (z.B. chronologische Verweisungen) wurde im Projekt nicht einbezogen, da sie für den geplanten Verwendungszweck nicht als wesentlich erschienen.

2.3 Online-Publikationen

Im ersten Abschnitts wurde die Literaturerfassung im Wesentlichen an einer traditionellen bibliothekarischen Publikation – der Buchform – erläutert. Der darauffolgende zweite Abschnitt behandelte die Schlagwortnormdatei als Instrument der Sacherschließung. In diesem Kapitel möchte ich nun auf neuere Arbeitsgebiete der Bibliotheken eingehen. Es handelt sich hierbei um sog. Online-Publikationen. Die Sacherschließung von Online-Publikationen ist der Grund, dass das dargestellte Projekt begonnen wurde.

Durch die Einführung der neuen Technologien hat sich nach und nach auch eine Erweiterung der Veröffentlichungsarten von Publikationen entwickelt. Das zur Veröffentlichung gedachte Dokument ist nicht mehr länger nur an Printformen wie Bücher, Zeitschriften usw. gebunden. Veröffentlichte Arbeiten können den Interessenten inzwischen auf einfache Weise digital zugänglich gemacht werden. Gleichzeitig entwickelt sich das Bereitstellen und Erschließen elektronischer Objekte im Internet zum zukunftsorientierten Angebot moderner Dienstleistungsbibliotheken.

Die Form der Online-Publikation hängt dabei von der gewählten Erstellungsart ab. Es kann sich um HTML-Seiten im Internet, die wiederum mit anderen HTML-Seiten verlinkt sind, d. h. auf andere Seiten verweisen, handeln. Eine weitere Möglichkeit ist die Speicherung einer softwarespezifischen Dokumentenart, wie z.B. Microsoft Word Dokumente mit der Dateiendung *.doc. Dem entspricht die verbreitete Art der Dokumentenspeicherung im PDF-Format von Adobe. Es sichert die plattformübergreifende einheitliche Darstellung des Dokumentes.

Als Beispiel für ein im Internet veröffentlichtes Dokument möchte ich die Online-Version des Buches „HTML 4 Handbuch“ von Stefan Münz erwähnen. Die Online-Ausgabe ist im deutschsprachigen Internet als „selfHTML“ bekannt und kann unter <http://www.teamone.de/selfaktuell/> angesehen werden. Zu erwähnen ist auch, daß das Bibliotheksservice-Zentrum Baden-Württemberg als erste bibliothekarische Institution in Deutschland seit 1995 einen hypertextfähigen Katalog anbietet, der die

Möglichkeit der Einbindung, der ganzen Ressourcen über Hyperlinks realisiert. Die ältesten so angebotenen Dokumente und Nachweise stammen ebenso aus dem Jahr 1995, so dass das BSZ auf eine breite Expertise in der Erschließung, Präsentation und Archivierung von Volltexten im Internet verweisen kann.

Online-Publikationen bringen eine Reihe von Vorteilen mit sich. Interessenten können schon kurz nach der Veröffentlichung in nie dagewesener Schnelligkeit und Einfachheit auf ein Online-Exemplar zugreifen. Bibliotheken können Kosten einsparen, z.B. in Form von Reduzierung der Lagerkosten oder in Form der Nutzung von Erschließungsdaten, die die Autoren selbst liefern.

Durch Online-Publikationen hat sich ein neues Tätigkeitsfeld für Bibliotheken aufgetan. Mit Fördermitteln des Landes Baden-Württemberg sind in den letzten Jahren virtuelle Forschungsbibliotheken an Universitäts- und Landesbibliotheken entstanden. Eine virtuelle Forschungsbibliothek speichert Dokumente auf Textservern, die an das Internet angeschlossen sind. Sie zeichnen sich darin aus, dass Autoren ihre Publikationen in die Forschungsbibliothek einbringen und zugänglich machen. Im nachfolgenden Abschnitt wird ein bereits implementierte Forschungsbibliothek vorgestellt.

2.4 SüdWestdeutsches Online Publikationssystem (SWOP)

SWOP ist ein Publikations- und Archivierungssystem für elektronische Dokumente, das auf der technischen Basis von OPUS (Online Publikationssystem der Universität Stuttgart) Veröffentlichungen des Bibliotheksservice-Zentrums mit der Möglichkeit der strukturierten Volltextrecherche und im Netz bereitstellt.

2.4.1 Einbringen von Dokumenten

Um ein Dokument in SWOP einbringen zu können, muß es inhaltlich und formal beschrieben werden. Dies geschieht mit sogenannten Metadaten (Daten über Daten). Sie enthalten strukturierte Informationen über das Dokument selbst, wie z.B. Angaben

über Autor, Titel, Schlagwort, Abstract oder Zeitpunkt der Veröffentlichung. Durch Metadaten wird die Recherche und Verwaltung von Dokumenten oder Objekten gewährleistet und der Zugriff auf sie optimiert. Sie entsprechen in gewisser Weise dem, was an Erschließungsarbeit in den Bibliotheken seit jeher geleistet wurde. Wie jedes bibliothekarische Regelwerk setzt auch der effektive Einsatz von Metadaten obligatorisch eine strukturierte Standardisierung voraus. Im Bibliothekswesen setzt sich das Elementset der international eingeführten Dublin Core Metadata Initiative durch, das mit 15 in Syntax und Semantik definierten Elementen die vorliegenden Objekte aller Art zu beschreiben versucht. Das BSZ wendet dieses Basisformat seit 1996 an und trägt zu seiner Verbreitung und Spezifizierung im deutschen Sprachgebiet maßgeblich bei. Deshalb wurde gemeinsam in der Bibliotheksregion ein Erfassungsformular entworfen, das die nötigen Daten erhebt, die als Dublin Core Metadaten weiter bearbeitet und in die vorhandenen Bibliothekskataloge mit ihren spezifischen, differenzierteren Formate übertragen werden können. Dieses HTML-Formular steht z.B. unter der URL [http://www.bsz-bw.de/cgi-bin/w3-mysql/opus/uni/neu_allg2.html](http://www.bsz-bw.de/cgi-bin/w3-<u>mysql/opus/uni/neu_allg2.html</u>) zur Verfügung. In Tabelle 1 sind die wichtigsten Eingabefelder aufgelistet.

Feldname	Beispiel einer Eingabe
Dokumentart	Diplomarbeit
Originaltitel des Dokuments	Die Schlagwortnormdatei als Erschliessungs- instrument von Online-Publikationen
Verfasser	Björn Eggers
Kontrollierte Schlagwörter der SWD (Deutsch)	Schlagwortnormdatei <Werk>
Jahr der Fertigstellung	2001

Tab. 1: Beispieleingaben für Formular Daten in SWOP

Ist das obengenannte Formular vollständig ausgefüllt, kann die Dokumentenbeschreibung, über den Anmelde-Button an SWOP transferiert werden. Anschließend wird über ein Perl-Skript der eigentliche File-Upload des Dokuments selbst getätigt. Das übertragene Dokument und die Metadaten sind nun in einer temporären Datenbank gespeichert. Die gültige Veröffentlichung findet nachhaltig durch den Systemadministrator statt, d.h. es ist nicht möglich, Dokumente von außerhalb unautorisiert in das System einzubringen.

2.4.2 Übernahme in die SWOP-Datenbank

Bei einer Neueinbringung wird das Dokument wie bereits erwähnt, zuerst in eine temporäre Datenbank eingetragen. Gleichzeitig wird eine Mail an den für die Pflege der Dokumente zuständigen Sachbearbeiter gesandt. Der über den Neueingang eines Dokumentes informierte Sachbearbeiter überprüft die vom Autor eingegebenen Daten auf sachliche Korrektheit. Dies geschieht durch ein eigenes HTML-Formular. Erst wenn das Dokument von ihm freigegeben wird, erfolgt die Eintragung in die endgültige SWOP-Datenbank.

2.4.3 Weitere Verwendung der Metadaten

Das BSZ Baden-Württemberg betreibt mit der Datenbank des Südwestdeutschen Bibliotheksverbundes (SWB) den maßgeblichen einheitlichen Nachweis aller in der Region vorhandenen wissenschaftlichen Bibliotheksbestände. Er umfaßt die Bundesländer Baden-Württemberg, Sachsen, Saarland und Teile von Rheinland-Pfalz mit derzeit ca. 8.800.000 Titeln und 25.000.000 Bestandsmeldungen. Die einzelnen Dokumentbeschreibungen ("Katalogisate") stehen allen Teilnehmerbibliotheken zur Partizipation zur Verfügung; dies ist möglich, da nach einheitlichen Regelwerken und mit gemeinsamen, unterlegten Normdateien gearbeitet wird. Für den Bereich der verbalen Sacherschließung ist dies Normdatei die SWD.

Erklärtes Ziel der Bibliotheken ist es, die Erschließung von Online-Ressourcen mit der bestehenden Katalogisierung von konventionellen Medien zu verzahnen: der

Wechsel des Publikationsmediums alleine darf nicht die erreichte einheitliche Verzeichnung aller Medienarten konterkarrieren.

Um rationelle Arbeits- und Katalogisierungsformen zu unterstützen, werden deshalb die von den Autoren erhobenen Metadaten in das bibliographische Importformat des SWB umgesetzt, in die SWB-Datenbank eingespielt und dort einer endgültigen Redaktion unterworfen. Dann stehen sie für die Nutzung in lokalen Katalogen (z.B. dem der FH Konstanz unter URL <http://www.bib.fh-konstanz.de/kataloge.shtml>) oder weitere Zwecke zur Verfügung.

Die nahtlose Weiterverwendbarkeit der ursprünglich erhobenen Daten ist für dieses Vorgehen wesentlich. Für die Sacherschließung bedeutet dies: wird die SWD für die konventionelle Beschreibung verwendet, muß die auch für die Metadatenerfassung der Autoren zur Verfügung stehen, um weiter genutzt werden zu können. Diese Lücke schließt dieses Projekt, da die SWD bislang nur beschränkt und nicht aktuell, bzw. in einem proprietären, nicht für die Metadatenerfassung nutzbaren Format angeboten wurde.

2.4.4 Recherchieren in SWOP

Im Gegensatz zum Einbringen von Dokumenten ist das Recherchieren in den Publikationen von außen erlaubt. Das Suchen von Publikationen kann durch eine parametrische Suche erfolgen, d.h. es können bestimmte Metadaten (z.B. Autor, Titel, Schlagwort) auf Übereinstimmung abgefragt werden. Alternativ kann auch eine Volltextsuche durchgeführt werden, in der ein bestimmter Begriff im Dokument gesucht wird.

Sind bei einer Suchanfrage Treffer vorhanden, werden diese in Tabellenform auf einer HTML-Seite angezeigt. Die Treffer bestehen aus einer laufenden Nummer, dem Dokumententitel, dem Autor und das Erscheinungsjahr. Wobei der Titel auf eine sog. Frontdoor verlinkt ist. Auf dieser Frontdoor werden die Metadaten noch einmal angezeigt und das Dokument steht zum Download bzw. zur Ansicht bereit.

2.4.5 Datenfluß in SWOP bzw. OPUS

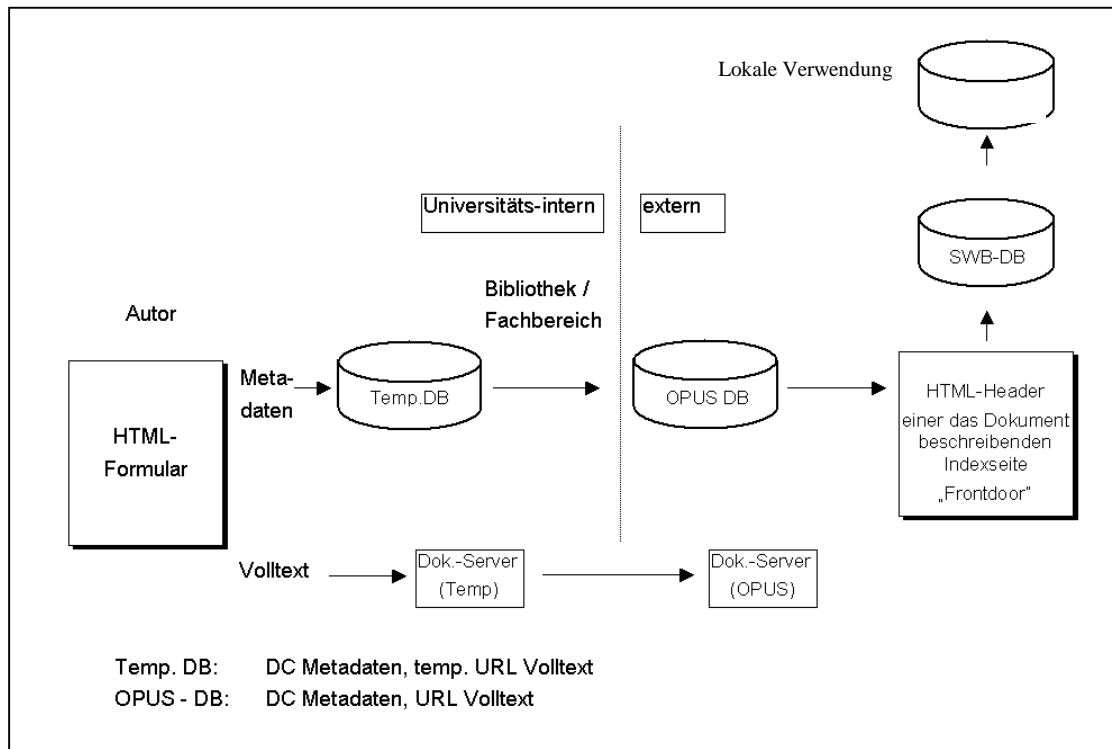


Abb 1.: Datenfluß in SWOP bzw. Opus

3 Die Entwicklungsumgebung

Im ersten Teil dieses Kapitels wird die benutzte Entwicklungsumgebung in Form einer theoretischen Abhandlung vorgestellt. Im zweiten Teil werden Alternativen zur gewählten Entwicklungsumgebung aufgezeigt.

3.1 Die CGI-Schnittstelle

In den Anfängen des Internets waren statische HTML-Seiten der Standard. Im Laufe der Zeit wurde erkannt, dass diese Form der Webpräsenz für größere Internet-Projekte ungeeignet ist. Aus diesem Grunde ist man zur Erstellung von dynamischen Webseiten übergegangen. Hierbei wird die entsprechende HTML-Seite nicht starr nach dem gleichen Prinzip aufgebaut, sondern sie wird aufgrund von Interaktionen mit dem Anwender erstellt. Ein gutes Beispiel für diese dynamischen HTML-Seiten sind die im Internet weitverbreiteten Suchdienste, wie z.B. Yahoo oder Lycos.

Um diese Dynamik realisieren zu können, wurde die CGI-Schnittstelle eingeführt. CGI steht für Common Gateway Interface. Sie war einer der ersten Standards zur Erstellung von interaktiven HTML-Seiten. Zur Zeit ist die Version CGI/1.1. gültig. In der Zwischenzeit sind einige Alternativen zu CGI¹ entworfen worden. Einige dieser neuen Technologien bieten erhöhte Performance Vorteile gegenüber CGI. Dennoch ist das Common Gateway Interface auch heute noch eine der meist verbreitesten technischen Möglichkeiten um Dynamik im Internet zu gewährleisten. So schreibt Münz [Muen98]: „Ein entscheidender Vorteil der CGI-Schnittstelle bleibt jedoch die Tatsache, daß es sich – ähnlich wie bei HTML – um einen kommerziell unabhängigen, kostenlosen, produktübergreifenden Standard handelt.“

CGI ist eine Schnittstelle für Client/Server Applikationen. Der Benutzer einer solchen Anwendung ruft als erstes eine HTML-Seite auf. Danach kann er auf der Client Seite

¹ Auf einige von ihnen wird in Kapitel 3.2 näher eingegangen.

in seinem WWW-Browser Daten eingeben. Diese benutzerspezifischen Eingaben werden– evtl. zusammen mit vordefinierten Daten – an einen HTTP-Server übertragen.. In der Umgangssprache werden die HTTP-Server auch oftmals als Webserver bezeichnet. Aus Gründen der Lesbarkeit wird in diesem Dokument ab dieser stelle nur noch von Webservern gesprochen. Der Webserver ist für die Ablaufsteuerung in einer Client/Server-Umgebung verantwortlich. Durch seine CGI-Schnittstelle gibt er die empfangenen Daten an ein bestimmtes Computerprogramm weiter. In diesem Programm werden die Daten weiterverarbeitet. Die Art der Datenmanipulation hängt natürlich von der Internetapplikation ab. Bei dieser Datenbearbeitung kann eine neue HTML-Seite generiert werden, die wiederum vom Webserver an den Client gesendet wird. Dieser Vorgang ist in Abbildung 2 festgehalten.

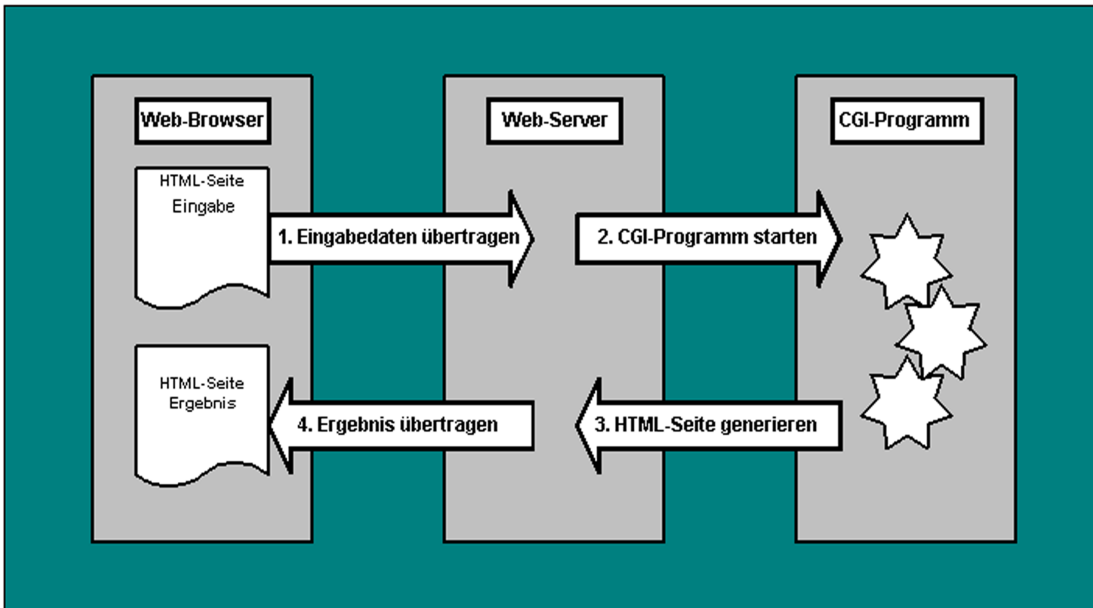


Abb 2.: Ablauf in einer CGI-Umgebung

Die Verarbeitung der übertragenen Daten ist nicht die einzige Aufgabe, die ein CGI-Programm leisten kann. Ein weiterer Vorteil ist, daß es auch mit externen Quellen kommunizieren kann. Meistens handelt es sich – wie auch in diesem Projekt – um die Anbindung eines Datenbankmanagementsystem (DBMS). In der Abbildung 3 ist dieser Vorgang skizziert.

Hiermit können benutzerspezifische Daten oder angebotene Informationen gespeichert, abgerufen, geändert, usw. werden.

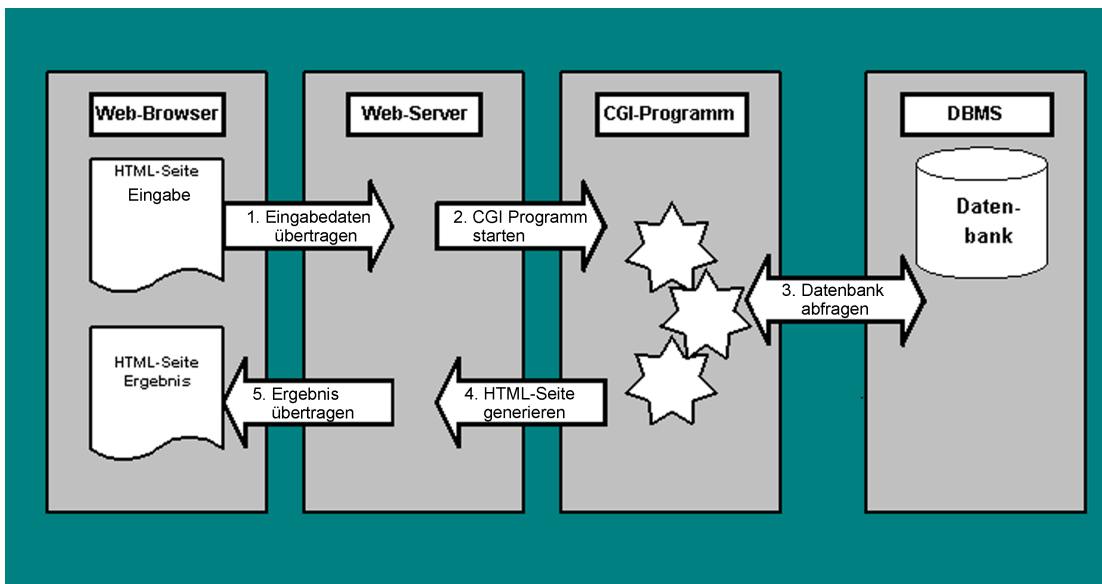


Abb 3.: Ablauf in einer CGI-Umgebung mit Datenbankbindung

In den meisten Webservern ist eine CGI-Schnittstelle fest integriert und ist durch folgende zwei Punkte spezifiziert:

- Ein dem Webserver bekanntes Verzeichnis in dem die CGI-Programme abgelegt sind. Nur Programme, die sich in diesem Verzeichnis befinden sind ausführbar. Der Name des Verzeichnisses hängt von der Konfiguration des Webserver ab. In den meisten Fällen wird dieser Ordner cgi-bin benannt.

- Die Fähigkeit sog. Umgebungsvariablen abzurufen. In den Umgebungsvariablen sind die übertragenen Daten gespeichert.

Der in der Literatur oft zu findende Ausdruck CGI-Programmierung suggeriert dem Leser, daß das benötigte Programm in CGI programmiert wurde. Dies ist jedoch nicht der Fall, denn CGI ist ein standardisiertes Protokoll und keine Programmiersprache. Mit welcher Programmiersprache wird nun aber ein sog. CGI-Programm geschrieben? Hierzu können verschiedene Möglichkeiten in Betracht gezogen werden.

Im einfachsten Fall kann es sich um eine vom Betriebssystem unterstützte Skriptsprache, wie z.B. die ksh-Shell handeln. Es können aber auch höhere Programmiersprachen, wie z.B. C/C++ verwendet werden. Der Nachteil dieser Sprachen ist, das sie nicht portierbar sind. Sie müssen speziell für das jeweilige Betriebssystem, auf dem das Programm eingesetzt werden soll, kompiliert werden. Als dritte Möglichkeit können aber auch Interpretersprachen eingesetzt werden. Voraussetzung hierfür ist, daß ein entsprechender Interpreter für das System installiert ist. Die verbreitetste Programmiersprache in CGI-Umgebungen ist wohl die Interpretersprache Perl¹.

CGI-Programme sind somit eigenständig ausführbare Programme oder Skripts – wofür sie allerdings in den wenigsten Fällen geschrieben wurden. Üblicherweise werden CGI-Programme erst mit einer eingehenden Anfrage eines Benutzers (via Formular) gestartet. Wenn sich das Ergebnis der Anfrage auf dem Client-Computer darstellt, ist das Programm meistens bereits abgelaufen. Es handelt sich also nicht um permanent laufende Programme. Sie sind lediglich Prozesse auf der Serverseite. Dies ist ein wesentlicher Nachteil bei der CGI-Programmierung.

In den vorangegangenen Erläuterungen wurde der Datenfluß vom Benutzer über dem Webserver mit seiner CGI-Schnittstelle zum CGI-Programm skizziert. Es stellt sich daher die Frage, wie das CGI-Programm nun an die übertragenen Daten gelangen kann. Dies geschieht durch folgenden Mechanismus: Der Webserver erhält die eingehenden Daten und speichert sie in sog. Umgebungsvariablen. Diese

¹ Sie wird auch in unserem Projekt eingesetzt. In den nachfolgenden Abschnitten wird deshalb näher auf sie eingegangen.

Umgebungsvariablen kann das CGI-Programm dann wiederum auslesen und intern zur Weiterverarbeitung speichern. Nach [Maur96] können die Umgebungsvariablen nach folgenden Kriterien gegliedert werden:

- A) Informationen über den Server
- B) Informationen über den Client
- C) Allgemeine Informationen über die Anfrage
- D) Eigentliche Parameterübergabe

Daraus ergibt sich folgende Auflistung der Umgebungsvariablen mit ihrer Bedeutung:

Kriterium	Name der Variable	Bedeutung der Variable
A	GATEWAY_INTERFACE	CGI-Version
A	SERVER_NAME	Name des Servers
A	SERVER_SOFTWARE	Server-Software
B	REMOTE_ADDR	IP-Adresse des Clients
B	REMOTE_HOST	Rechnername des Clients
B	SERVER_PROTOCOL	Protokoll-Version
C	SERVER_PORT	Port
C	AUTH_TYPE	Authentifizierung
C	REMOTE_USER	User-ID
C	REMOTE_IDENT	Benutzername
C	SCRIPT_NAME	Pfad auf CGI-Programm
C	HTTP_*	HTTP-Header
D	REQUEST_METHOD	Methode
D	QUERY_STRING	Query-String
D	CONTENT_LENGTH	Länge der Eingabe
D	CONTENT_TYPE	Art der Eingabe
D	PATH_INFO	Pfadinformationen
D	PATH_TRANSLATED	Pfadangaben

Tab. 2: Erläuterung der Umgebungsvariablen

3.1.1 Technische Anforderungen an eine CGI-Schnittstelle

Neben den Hardwarevoraussetzungen – die hier außer Acht gelassen werden – sind vier Softwarehauptkomponenten von Nöten. Im einzelnen handelt es sich um:

- Betriebssystem
- Webserver
- Datenbank
- Programmiersprache

In der Praxis wird oft das sog. LAMP-Konzept angewendet. Wobei LAMP¹ für vier Vertreter der oben aufgeführten Softwarekomponenten steht: **L**inux, **A**pache, **M**ySQL und **P**erl.

In den folgenden Abschnitten werden die von uns eingesetzten Komponenten kurz und einführend beschrieben.

3.1.1.1 Betriebssystem: Solaris

An das Operating System (OS) werden Anforderungen, wie die Kommunikation mit den Prozessoren (Prozessverwaltung), Verwaltung des Arbeitsspeichers und die Speicherung der Daten (Filesystem) gestellt. Auf diese Hauptaufgaben soll im einzelnen hier nicht weiter eingegangen werden.

An dieser Stelle möchte ich lediglich erwähnen, dass für dieses Projekt das Betriebssystem SunOS 5.6 der Firma SUN eingesetzt wird. SunOS 5.6 basiert auf dem bekannten Betriebssystem UNIX.

Durch den Einsatz eines UNIX-Betriebssystems kann also nicht mehr von den oben erwähnten LAMP-Konzept gesprochen werden².

¹ Neuerdings wird auch beim Einsatz von PHP von einem LAMP-Konzept gesprochen.

² Da es sich bei Linux um eine Freewareversion von Unix handelt, fällt dieser Unterschied in der Praxis nicht bzw. kaum ins Gewicht, denn alle anderen Komponenten entsprechen exakt diesem Konzept.

3.1.1.2 Webserver: Apache

Der Webserver fungiert als Schaltzentrale, die externe HTTP-Anfragen (Requests) beantwortet. Weiterhin soll er Aufgaben, die er nicht selber bewältigen kann, an externe Programme delegieren. Im HTTP-Standard 1.0 bzw. 1.1 ist er als eigenständiger Prozeß, der auf dem Server läuft, definiert.

In unserem Projekt setzen wir den Webserver Apache 1.3.14 ein. Er ist eine Weiterentwicklung des bekannten NCSA httpd 1.3 Webservers. In [Laur99] wird dazu folgendes angemerkt: „Apache verdankt seinen Namen der Tatsache, dass er aus vorhandenem Code und einigen Patches entwickelt wurde“. Der Apache zeichnet sich unter anderem dadurch aus, dass er auf fast allen Plattformen verfügbar ist und sein Quelltext offen liegt (Open Source). Aufgrund dieser Aspekte ist er einer der verbreitetsten Webserver im Internet. Dies belegt auch eine Studie der Firma Netcraft Ltd. (<http://www.netcraft.com/survey/>). In ihr wurde ein Marktanteil des Apache Webservers von über 60% ermittelt.

Der Apache ist kostenlos im Internet unter <http://www.apache.org> verfügbar.

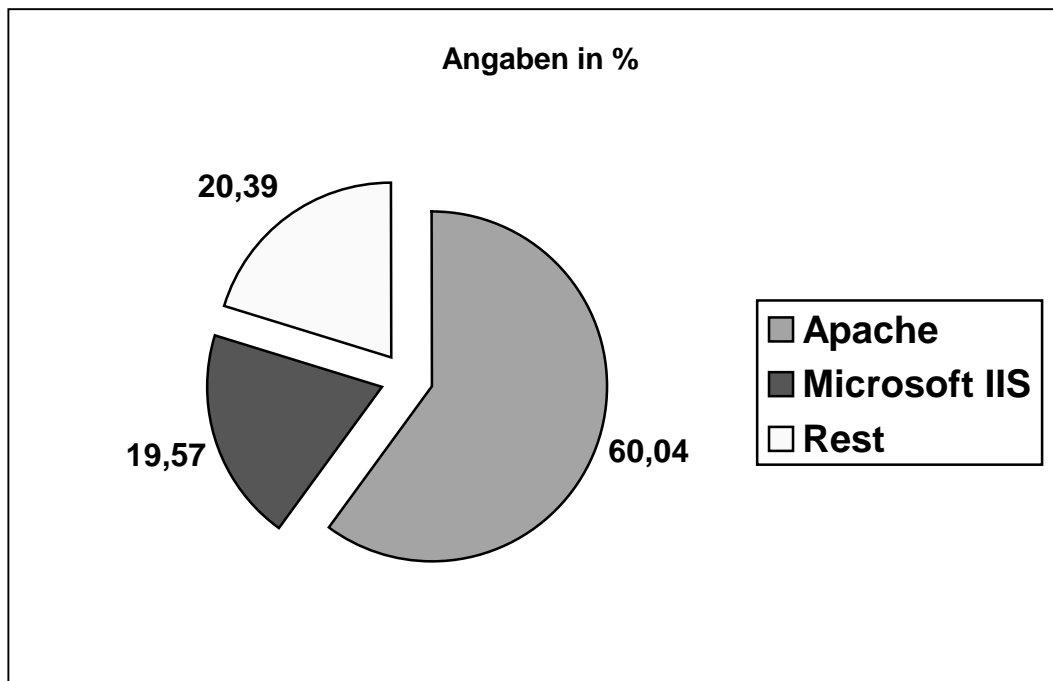


Abb 4.: Marktanteile Webserver, Stand Dezember 2000

3.1.1.3 Datenbankmanagementsystem: MySQL

Die Nutzung einer Datenbank bietet sich besonders bei Anwendungen mit einem umfangreichen Datenstamm an. Durch sie wird eine strukturierte Darstellung von Daten möglich. Die Datenbank ist somit eine effiziente Art, einen großen Datenbestand zu pflegen.

Man kann zwischen hierarchischen, relationalen und objektorientierten Datenbanken unterscheiden. Im World Wide Web (WWW) werden überwiegend relationale Datenbanken eingesetzt. Hierbei werden die Daten in Tabellen abgelegt, die untereinander verknüpft werden können.

Um die Daten manipulieren zu können, wird die standardisierte Sprache SQL (Standard Query Language) verwendet. Sie ist datenbankunabhängig und bietet eine abstrakte Möglichkeit, auf unterschiedliche Datenbanken zuzugreifen.

Wir setzen in unserem Projekt die Datenbank MySQL in der Version 3.23.28 ein.

Bei MySQL handelt es sich um ein Open Source Produkt. Die relationale Datenbank kann im Internet unter <http://www.mysql.com> bezogen werden.

3.1.1.4 Programmiersprache: Perl

Eine Programmiersprache ist eine Kommunikationsschnittstelle zwischen Mensch und Rechner. Im Gegensatz zur menschlichen Sprache (z.B. Deutsch), die als Kommunikationsmittel zwischen Menschen eingesetzt wird, ist sie eindeutig. In einer Programmiersprache gibt es keine mehrdeutige Begriffe, wie z.B. „Kiefer“ als Teil des Skeletts oder alternativ als Nadelbaumgattung. Programmiersprachen haben eine streng definierte Grammatik, die sog. Syntax. In einem Programm müssen Anweisungen syntaktisch korrekt sein, d.h. eine Anweisung kann nur eine einzige Bedeutung haben. Es muß definitiv festgelegt werden, ob „Kiefer“ nun ein Teil des menschlichen Gesichtsschädels ist oder ob es sich um einen Nadelbaum im Wald handelt.

Ziel einer Programmiersprache ist es, Probleme, die von menschlicher Seite definiert worden sind, in eine dem Rechner verständliche Form zu übersetzen. Durch diese „Übersetzung“ ist dann der Computer in der Lage, das Problem zu lösen.

Die von uns eingesetzte Programmiersprache ist Perl. Der Name Perl steht für Practical Extraction and Reporting Language und ist eine Entwicklung von Larry Wall. Sie ist eine interpretierte Sprache und erblickte bereits 1987 das Licht der Welt. Nach [Hajj98] ist Perl eine *Sprache*, die *praktisch* zu verwenden ist und deren wesentlicher Zweck in der Auswertung von Dateien, insbesondere dem *Extrahieren* von Informationen und dem Präsentieren dieser Informationen als gut leserliche *Reports* liegt. Dieser Eigenschaft und dem Umstand, dass Perl ein Kind der Open Source Philosophie ist, verdankt sie ihre Popularität. Obwohl diese Programmiersprache hauptsächlich im CGI-Bereich eingesetzt wird, kann sie für die Erstellung fast jeder Applikation herangezogen werden. Seit der Version 5 kann Perl sogar objektorientiert angewendet werden.

3.2 Alternativen zu CGI

Wie bereits erwähnt gibt es in der Zwischenzeit Alternativen zu CGI. In den folgenden Kapiteln werden einige neue Technologien beleuchtet. Den größten Teil bei dieser Betrachtung nehmen Java Servlets ein. Daneben werden einige Skript-basierte Lösungen kurz vorgestellt.

3.2.1 Java Servlets

Der Servlet-Ansatz basiert auf der Java-Technologie von Sun Microsystems. Java ist eine objektorientierte, plattformunabhängige Programmiersprache.

Um ein Java-Programm, das im Quellcode vorliegt, ausführen zu können, muß als erstes der Java-Code auf einem beliebigen System in sog. Bytecode kompiliert werden. Dieser Bytecode ist verwandt mit Maschinencode, jedoch ist er nicht an einen spezifischen Prozessor gebunden. Der Bytecode kann dann auf jedem System mit Hilfe eines Interpreters ausgeführt werden. Dieser Bytecode-Interpreter wird auch „Java-Laufzeit-Interpreter“ oder „virtuelle Java-Maschine“ genannt. Hierdurch wird die Plattformunabhängigkeit erreicht.

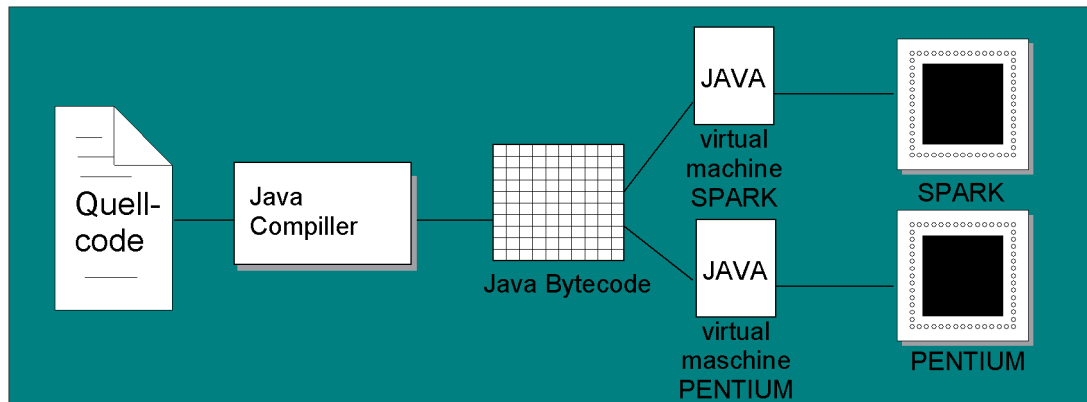


Abb 5.: Vom Java-Quellcode zur Programmausführung

Dem Programmierer wird durch die Servlet API eine Schnittstelle für serverseitige Java-Programme an die Hand gegeben. Servlets sind in Java geschriebene Programme, die auf dem Server ausgeführt werden.

Ähnlich wie bei CGI gibt es die Möglichkeit, auf Umgebungsvariablen zuzugreifen und es existieren Datenströme in Form von Anfragen und Antworten. Die Servlet API unterscheidet sich allerdings von der CGI-Schnittstelle darin, dass bei einer Anfrage nicht immer ein neuer Prozeß kreiert wird, sondern nur ein Thread. Hierdurch entsteht bei konkurrierenden Anfragen ein Geschwindigkeitsvorteil.

Die folgende Abbildung verdeutlicht den Weg vom Client zum Servlet.

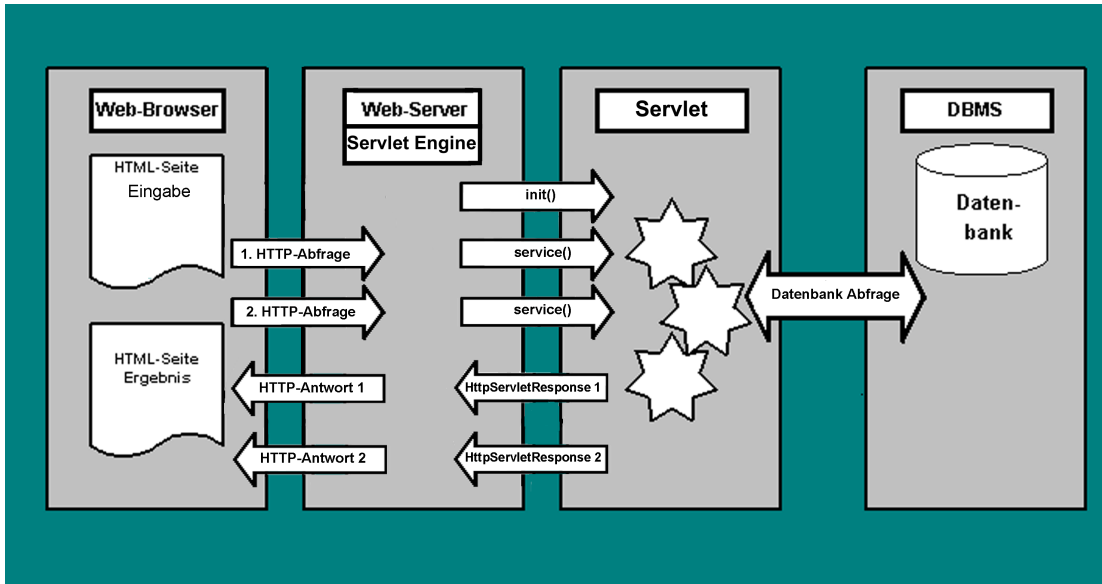


Abb 6.: Kommunikation mit Servlets

Im Vergleich zu CGI kann man auf dieser Abbildung erkennen, dass zusätzlich zum Webserver eine korrespondierende Servlet-Engine notwendig ist. Die Servlet-Engine ist neben dem Webserver und dem Java Servlet Development Kit (JSDK) eine wichtige Voraussetzung, um Servlets einsetzen zu können. Eine ganze Reihe freier und kommerzieller Servlets-Engines kann über die Homepage von Sun (<http://java.sun.com/products/servlet/download.html>) bezogen werden. Für den bekannten Webserver Apache kann man die Servlet-Engine „Jserv“ direkt unter <http://java.apache.org/> erhalten.

Um ein Servlet zusätzlich an eine Datenbank anzubinden, muß ein entsprechender JDBC-Treiber in den Speicher geladen werden. Dieser registriert sich beim DriverManager und wird dadurch als gültiger Datenbanktreiber erkannt. Der

DriverManager ist ein Objekt, das in der Bibliothek java.sql.* implementiert ist. Durch die Verwendung des Driver Managers wird eine gleichzeitige Anbindung an mehreren Datenbanken ermöglicht. Uunvereinbare Datenbanksysteme können nun miteinander verknüpft werden.

Abschließend seien hier noch einmal die Vor- und Nachteile der Servlet-Technologie zusammengefaßt. Der Programmiersprache Java wird derzeit große Aufmerksamkeit gewidmet. Man kann sogar behaupten, dass sie *hype*¹ ist. Laut [Krue2000] haben die Designer von Java ihre Bemühungen bei der Entwicklung dieser neuen Sprache wie folgt zusammengefaßt:

„Java soll eine einfache, objektorientierte, verteilte, interpretierte, robuste, sichere, architekturneutrale, portable, performante, nebenläufige, dynamische Programmiersprache sein.“

Tatsächlich sind diese Forderungen an eine Programmiersprache in Java relativ gut umgesetzt worden. Dieser Umstand schlägt natürlich auch bei den Servlets durch. Neben einem guten Sicherheitskonzept sticht besonders der Geschwindigkeitsvorteil hervor.

Servlets sind trotz all dieser Vorteile zur Zeit noch sehr geringfügig verbreitet. Dies ist vor allem mit der Gegebenheit zu erklären, dass die populären Webserver, wie Apache oder IIS, nur mangelhaft direkt unterstützt werden.

¹ Hype engl. Medienrummel

3.2.2 PHP

Eine weitere Möglichkeit WWW-Seiten dynamisch zu gestalten ist PHP. Diese Abkürzung stand ursprünglich für „Personal Home Page Tools“. PHP wurde von Rasmus Lerdorf geschrieben. Die erste Vorversion wurde im Herbst 1994 veröffentlicht. Zur Zeit trägt die aktuelle Version die Bezeichnung PHP 4.04. Mit zunehmender Verbreitung hat sich auch die Erklärung „PHP Hypertext Preprocessor“ für PHP eingebürgert.

PHP ist eine Erweiterung für Webserver. Für den Apache ist zum Beispiel das Modul `mod_php3` erhältlich. Bei diesen Erweiterungen bzw. Modulen handelt es sich um eingebaute Skript-Engines. Technisch gesehen sind diese Skript-Engines Interpreter.

Im Gegensatz zu Skriptsprachen, wie Perl steht der Code aber nicht in einem eigenständigen Programm, sondern die PHP-Anweisungen befinden sich direkt im HTML-Quelltext. Der Unterschied zu puren HTML-Seiten besteht darin, dass Seiten, die PHP-Code enthalten, mit einer speziellen Dateiendung versehen werden. Übliche Dateiendungen, in Abhängigkeit der Konfiguration des Webserver, sind z.B. `.php` oder `.phtml`. Durch diese abweichende Dateiendung weiß der Webserver, dass sich in dieser Seite PHP-Code befindet. Der Webserver erkennt PHP-Anweisungen an einer Folge spezieller Zeichen. Eine PHP-Anweisung wird „escaped“, also vom sonstigen HTML-Text hervorgehoben. Eine solche Zeichenfolge könnte z.B. `<?php ...?>` sein.

Findet der Webserver eine solche Anweisung innerhalb einer Webseite, gibt er diese an den PHP-Interpreter weiter. Der Code wird von ihm weiterverarbeitet. Als Ergebnis sendet er dem Webserver HTML-Code zurück. Dieser setzt diesen HTML-Code genau dort ein, wo ursprünglich die PHP-Anweisung stand.

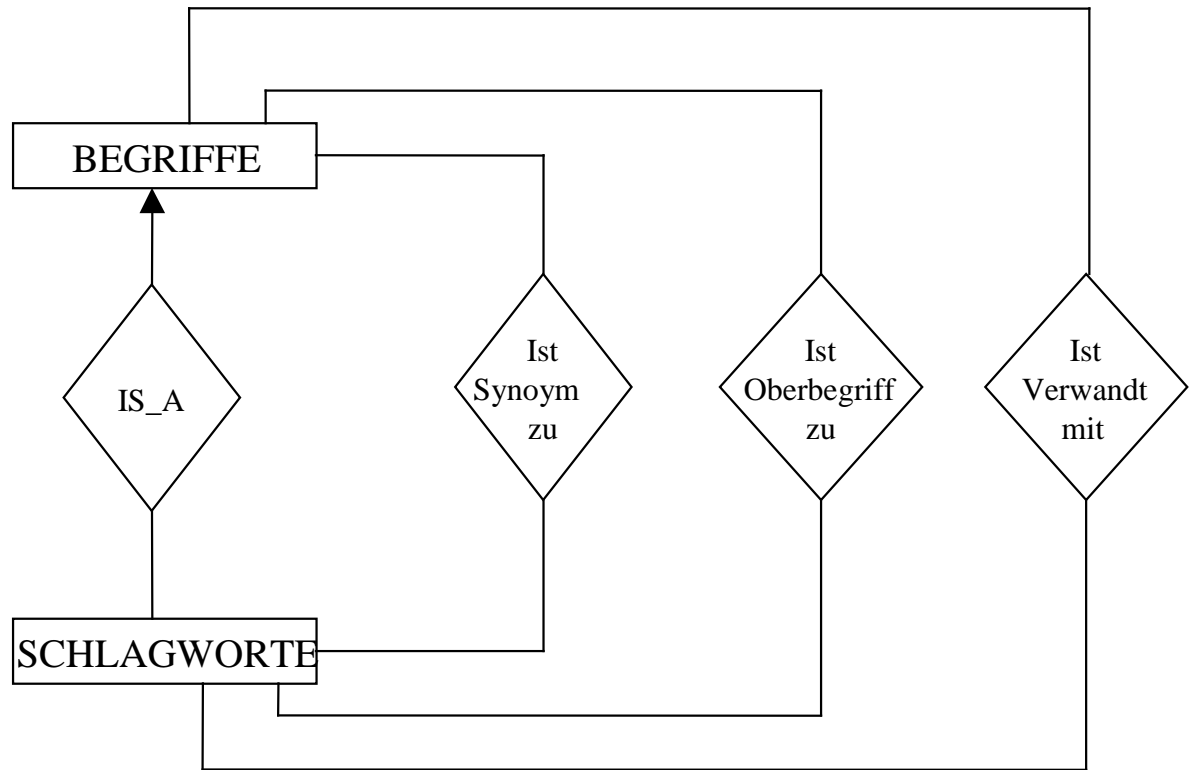
In [schm99] wird ein Vorteil für den Programmierer mit der Tatsache begründet, dass bei PHP eine erheblich einfache Programmierbarkeit vorliegt. Desweiteren werden die Vielzahl von vorhandenen Modulen, die systemübergreifende Verfügbarkeit und das Open Source Prinzip als Vorteil hervorgehoben.

3.2.3 Proprietäre Server-API

Programmierschnittstellen für Anwendungen die direkt vom Hersteller für seinen eigenen Webserver implementiert worden sind, nennt man proprietär. Vertreter dieser Gattung sind z.B das Internet Server Application Programming Interface (ISAPI) von Microsoft oder das Netscape Server Application Programming Interface (NSAPI). Hierbei werden die einzelnen Programme direkt in den Webserver eingebunden. Ein Vorteil bei dieser Vorgehensweise ist, dass nicht dauernd zwischen einzelnen Prozessen umgeschaltet werden muß. Dies bringt einen erheblichen Geschwindigkeitsvorteil mit sich. Auf der anderen Seite ist damit immer die Gefahr verbunden, daß der Webserver zum Absturz gebracht wird. Ein weiterer Nachteil entsteht durch die Serverabhängigkeit.

4 Die Architektur des Prototyps

4.1 Das Datenmodell



BEGRIFFE = ({ suchbegriff, anzeigebegriff, swdnr })

SCHLAGWORTE = ({ swdnr, erluterung, alternative })

Abb 7.: Das Datenmodell

4.2 Die Tabellenstrukturen

Im nachfolgenden werden die einzelnen Strukturen der verwendeten Tabellen aufgelistet und kurz erläutert.

a) Tabelle: BEGRIFFE

Field	Type	Null	Key	Default
suchbegriff	text		MUL	
anzeigebegriff	text	YES		NULL
swdnr	varchar(10)	YES		NULL

Abb 8.: Tabellenstruktur BEGRIFFE

In dieser Tabelle werden sämtliche Begriffe gespeichert. Unter Begriff verstehe ich in diesem Zusammenhang, alle Sucheinstiege, die durch die Auswertung der belegten Datenfelder der einzelnen SWD-Sätzen entstehen. Dabei kann es sich um einzelne Worte, Wortfolgen oder mehrgliedrige Wendungen handeln. Die Felder suchbegriff und anzeigebegriff sind aufgrund von Sortierregeln, die im Abschnitt Datenbankloader erläutern werden, notwendig. Falls es sich bei dem Begriff um eine Ansetzung handelt, ist das Feld swdnr mit einer internen SWD-Nummer belegt.

b) Tabelle: ANSETZUNGEN

Field	Type	Null	Key	Default
swdnr	varchar(10)		MUL	
erlaeuterung	text	YES		NULL
alternative	text	YES		NULL

Abb 9.: Tabellenstruktur ANSETZUNGEN

In der Tabelle ANSETZUNGEN werden Erläuterungen und Definitionen zum Schlagwort selbst gehalten. Diese werden nicht zur Suche aufbereitet, da es sich um einen rein darstellenden Text handelt. Erläuterung und Ansetzung werden über die SWD-Nummer in den verschiedenen Tabellen zusammengehalten.

c) Tabelle: ISA_SYNONYM

Field	Type	Null	Key	Default
synonym	text		MUL	
swdnr	varchar(10)			

Abb 10.: Tabellenstruktur ISA_SYNONYM

Hier werden Verweisungen zu Schlagwörtern festgehalten.

d) Tabelle: ISA_VERWANDTSCHAFT

Field	Type	Null	Key	Default
verwandter	text			
swdnr	varchar(10)		MUL	

Abb 11.: Tabellenstruktur ISA_VERWANDTSCHAFT

In dieser Tabelle werden Verwandtschaftsbeziehungen zwischen Begriffen aufgezeigt.

e) Tabelle: ISA_OBERBEGRIFF

Field	Type	Null	Key	Default
oberbegriff	text			
swdnr	varchar(10)		MUL	

Abb 12.: Tabellenstruktur ISA_OBERBEGRIFF

In dieser Tabelle werden hierarchische Beziehungen zwischen zwei Begriffen aufgezeigt.

4.3 Administrationstools

In diesem Abschnitt werden die zwei Administrationstools dataextract und dataloader vorgestellt.

4.3.1 Das Tool dataextract

Dies Tool hat folgende Aufgaben zu verrichten:

- Extraktion der benötigten Daten aus MAB2
- Bearbeitung von Sonderzeichen
- Auswertung von Nicht-Sortierzeichen
- Erstellen von aufbereiteten Dateien zum Dataload

b) Extraktion der benötigten Daten aus MAB2

Die SWD wird im Maschinellen Austauschformat für Bibliotheken 2 (MAB2) geliefert. MAB2 ist ein spezielles Datenformat, das für die elektronische Verarbeitung bibliographischer Daten entwickelt wurde.

Ein Schlagwortsatz innerhalb der Schlagwortnormdatei enthält eine Vielzahl von Informationen, die zum größten Teil für unser Projekt nicht gebraucht werden. Dataextract hat somit die Aufgabe, die benötigten Begriffe heraus zu filtern. Alle anderen Informationen eines Schlagwortsatzes sind nicht von Nöten.

c) Bearbeitung von Sonderzeichen

Innerhalb unserer Datenbank sowie bei der Anzeige von HTML-Seiten auf dem Bildschirm arbeiten wir mit dem Zeichensatz ISO-8559-1. Er ist ein 8-Bit Zeichensatz, d. h. es können maximal 256 Zeichen dargestellt werden. Dies sind viel zu wenig um alle Zeichen, die in der SWD verwendet werden, darzustellen. Der ISO-8559-1 enthält die schriftspezifischen Zeichen für westeuropäische und amerikanische Sprachen. Dagegen können z.B. mit diesem Zeichensatz die

meisten Sonderzeichen und Diakritika der mitteleuropäischen, slawischen oder asiatischen Sprachen nicht abgebildet werden.

Aufgabe des dataextract ist, die nicht darstellbaren Zeichen auf ihre Grundzeichen herunterzubrechen. Als Beispiel sei hier die Ligatur aus i und j genannt, das durch die beiden Grundzeichen i und j ersetzt wird. Oder ein dumpfes türkisches i, das zu einem normalen i wird. Dieses Vorgehen entspricht bibliothekarischen Usus bei ähnlichen Projekten und der üblichen Zeichenverwendung bei Recherchen.

d) Auswertung von Nicht-Sortierzeichen

Bei der Ansetzung von Schlagwörtern kann es durchaus sein, dass bestimmte Zeichenketten des Schlagwortes nicht sortiert werden sollen (z.B. einleitende Artikel). Bei der alphabetischen Ordnung der Schlagworte werden diese Zeichenketten nicht berücksichtigt. Solche Zeichenketten werden mit sog. Nicht-Sortierzeichen gekennzeichnet. Hierbei ist zwischen Nicht-Sortierzeichen-Anfang „^“ und Nicht-Sortierzeichen-Ende „%“ zu unterscheiden. Die Zeichenkette, die zwischen diesen beiden Nicht-Sortierzeichen steht, gilt als nicht zu sortieren. Im Beispiel:

Walser, Martin / ^Die% Gallistl'sche Krankheit

Soll nicht die Zeichenkette „Walser, Martin / Die Gallistl'sche Krankheit“ sortiert werden, sondern die Zeichenkette „Walser, Martin / Gallistl'sche Krankheit“.

Bei einer Anzeige dieser Zeichenkette auf dem Bildschirm wird aber nicht die zu sortierende Zeichenkette dargestellt, sondern die erste Variante, also „Walser, Martin / Die Gallistl'sche Krankheit“.

Aus diesem Grund wurde die Tabelle BEGRIFFE auch in die Spalten suchbegriff und anzeigebegriff unterteilt. In der Spalte suchbegriff steht nun die Zeichenfolge „Walser, Martin / Gallistl'sche Krankheit“ und in der Spalte anzeigebegriff steht die Zeichenkette „Walser, Martin / Die Gallistl'sche Krankheit“.

Wie oben zu sehen ist, entfernt der dataextract auf der Seite des Suchbegriffs komplett die Nicht-Sortierzeichen und die nicht zu sortierende Zeichenfolge. Dagegen werden auf der Seite des Anzeigebegriffs nur die Nichtsortierzeichen entfernt.

d) Erstellen von aufbereiteten Dateien zum Dataload

Bei der Abarbeitung sämtlicher Schlagwortsätze aus der SWD erstellt dataextract die Dateien *begriffe.txt*, *ansetzungen.txt*, *isa_synonym.txt*, *isa_oberbegriff.txt* und *isa_verwandtschaft.txt*. Diese Dateien sind ein genaues Abbild der in Kapitel 4.2 erwähnten Tabellenstrukturen.

4.3.2 Das Tool dataloader

Beim dataloader handelt es sich um eine Batch-Datei. Die einzige Aufgabe des dataloaders ist die mit dem dataextract erzeugten Dateifiles mit dem Datenimportutility von MySQL „load data infile 'datei' into table TABLE;“ nach einander in die Datenbank zu laden.

5 Die softwaretechnische Realisierung

5.1 HTML-Templates

Bei der Programmierung der einzelnen HTML-Seiten wurde mit Templates gearbeitet. Durch die Templates wird eine Trennung von reinem HTML-Code zur Darstellung des BSZ-Layouts und der eigentlichen programmiertechnischen Lösung der Problemstellung erreicht.

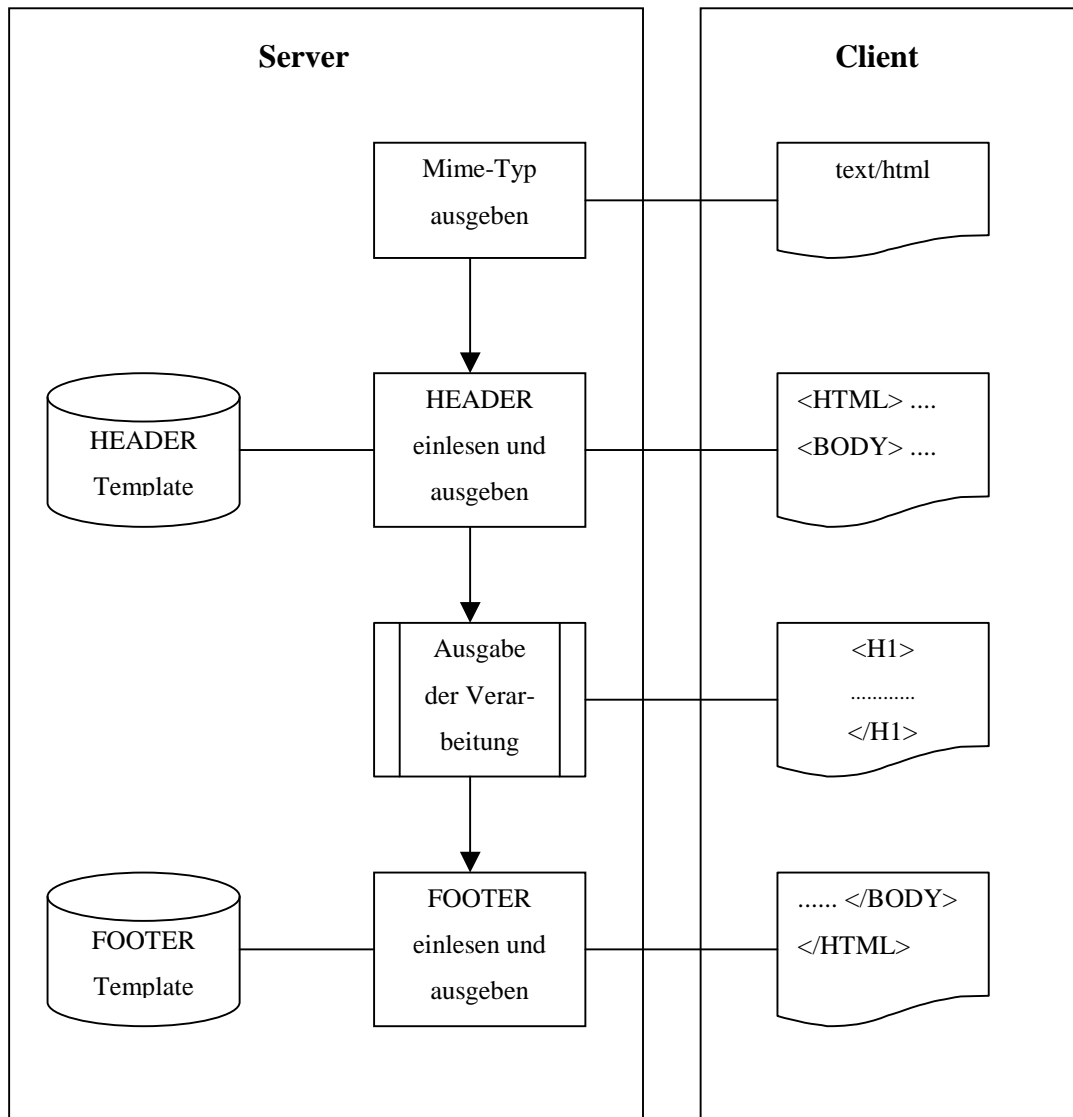


Abb 13.: Schematischer Ablauf von HTML-Templates

Diese Vorgehensweise hat mehrere Vorteile. Zum einem sind durch die Ausklammerung des BSZ-Layouts die CGI-Programme wesentlich übersichtlicher. HTML-Templates unterstützen die Lesbarkeit von Programmen. Ein weiterer Vorteil liegt in der Layoutunabhängigkeit der Programme. Ändert sich im Laufe der Zeit das Design der Homepage, so bleibt der eigentliche Programmcode davon unberührt.

Um diese Vorteile in Anspruch nehmen zu können, bestehen die generierten HTML-Seiten generell aus drei Teilen:

- eine HEADER-Datei (header.html)
- generiertes Ergebnis des Programmes
- eine FOOTER-datei (footer.html)

Der Schematische Ablauf von HTML-Templates kann aus der Abbildung 13 entnommen werden.

5.2 Die verschiedenen Suchmodi

5.2.1 Exakte Suche

Bei der exakten Suche werden alle Schlagworte in der Datenbanktabelle BEGRIFFE auf exakte vollständige Übereinstimmung, mit Ausnahme von Groß- und Kleinschreibung, mit dem eingegebenen Begriff überprüft.

Schlagworte sind alle Datensätze, bei denen eine SWD-Nummer (Spalte swdnr) vorhanden ist. Synonyme (ohne SWD-Nummer) werden in dieser Suchart nicht berücksichtigt.

Beispiel: Gesucht wird das Schlagwort „Gummibärchen“.

Wenn die Zeichenkette „gummib“ eingegeben wird, generiert das System eine Warnmeldung: Leider wurde kein Schlagwort gefunden! Dies liegt daran, dass „gummib“ kein Schlagwort ist.

Bei der Eingabe des Wortes „Gummibär“ wird dieselbe Warnmeldung generiert. Dies geschieht, obwohl „Gummibär“ in der Datenbanktabelle BEGRIFFE in der Spalte suchbegriffe exakt vorhanden ist. Dies ist auf die Tatsache zurückzuführen das „Gummibär“ keine SWD-Nummer besitzt. Dieser Ausdruck ist somit ein Synonym.

Eine Treffermeldung erscheint in diesem Beispiel nur, wenn exakt und vollständig „Gummibärchen“ eingegeben wird.

Aufgrund der Restriktion der genauer Übereinstimmung werden mit dieser Suche nur stark eingeschränkte Ergebnisse zu erzielen sein. Sie hilft besonders, eine exakte Form zu finden, die in vielen Sätzen verwendet, aber nur in genau einem Satz angesetzt wird. (z.B. Deutschland ist die Ansetzung für unseren Staat, kommt aber in ca. 5700 weiteren Sätzen einleitend vor, z.B. in deutschen Körperschaften.

5.2.2 Index Suche

Bei der Index Suche werden alle Schlagworte und Synonyme in der Datenbanktabelle BEGRIFFE aufgeblättert. Die eingebende Zeichenkette dient als Startwert, d.h. ab dem ersten Synonym oder Schlagwort, das mit der gewünschten Zeichenfolge beginnt, werden alle alphabetisch nachgeordneten Begriffe als Treffer auf dem Bildschirm ausgegeben.

Beispiel: Gesucht wird das Schlagwort „Gummibärchen“.

Wenn nun die Zeichenkette „gummib“ eingegeben wird, generiert das CGI-Programm einen Index ab dem Startwert „gummib“. Wobei in diesem Fall das Synonym „Gummibär“ der Startwert wäre. Als nächster Treffer wird schon das Schlagwort „Gummibärchen“ angezeigt. Da „Gummibärchen“ die nächst höhere alphabetische Ordnung von „Gummibär“ in der SWD ist, haben wir unseren gesuchten Begriff bereits gefunden.

Wie im Beispiel gesehen, kann man die Index Suche mit einem Buchindex vergleichen. Wenn man ungefähr weiß, wie das gesuchte Schlagwort beginnt, kann man es über die Index Suche sehr schnell finden.

Eine weitere Nutzung dieser Suchfunktion wäre das alphabetische „Durchblättern“ durch den Datenbestand der SWD. Bei weit über 1.000.000 Einträge (Schlagworte + Synonyme) würde dies aber ein ziemlich zeitaufwendige Beschäftigung sein.

Für den Fall, dass man weiß mit welcher Zeichenkette das gesuchte Schlagwort beginnt, ist diese Suchfunktion bestens geeignet.

5.2.3 Teilwortsuche

Bei der Teilwortsuche ist dem Interessenten ein Teilausdruck bekannt. Diese Suchfunktion ist so implementiert, daß sie alle Schlagwörter und Synonyme in der Datenbanktabelle BEGRIFFE durchsucht. Als Ergebnis werden alle Treffer zurückgegeben, bei denen die eingegebene Zeichenkette irgendwo im Synonym oder im Schlagwort befindet. Hierbei spielt es keine Rolle, ob die Zeichenkette am Anfang, in der Mitte oder am Ende des Begriffes gefunden wird.

Beispiel: Gesucht wird das Schlagwort „Gummibärchen“.

Wenn nun die Zeichenkette „gummi“ eingegeben wird, generiert das CGI-Programm eine alphabetisch sortierte HTML-Seite mit allen Treffern, in denen entweder im Schlagwort oder im Synonym die Zeichenkette „gummi“ vorhanden ist. Der erste Treffer liegt hier beim Schlagwort „Akaziengummi“. Unser gesuchtes Schlagwort befindet sich dagegen erst an Position 30.

Die Teilwortsuche wird aufgrund ihrer Fähigkeit, über die Identifikation von Wortteilen große Resultatmengen zu erzielen, beliebt sein. Denn sie stellt genau den Fall da, bei dem der Anwender irgendeine Silbe oder Wortteile von gesuchten Schlagwörtern weiß. Ihre Grenze wird aber dort liegen, wo zu unspezifisch gesucht wird, das heißt mit zu kleinen oder mit zu oft vorkommenden Zeichenfolgen.

Die drei Suchmodi wurden in enger Abstimmung mit den befaßten Bibliothekaren entwickelt. Sie gehen davon aus, daß damit ein guter Mittelweg bestritten ist, der bibliothekarische Fachkenntnisse bei der Suche nicht voraussetzt, aber gleichzeitig die richtige Hilfestellung gewährt, um übergroße Suchergebnisse zu vermeiden. Wert wurde darauf gelegt, bei der Präsentation der Suchergebnisse die zu verwendende Begrifflichkeit optisch und logisch deutlich herauszustellen.

5.3 Aufgabenbeschreibung der Programme

5.3.1 oswd-suchabfrage.pl



Abb 14.: Bildschirm-Maske oswd-suchabfrage.pl

Das Programm *oswd-suchabfrage.pl* generiert die Einstiegsmaske für die Recherche in der Online-Schlagwortnormdatei. Als erstes kann eine der in Kapitel 5.2 beschriebenen Suchfunktionen ausgewählt werden. Standardmäßig ist die Teilwortsuche voreingestellt.

Bei der Teilwortsuche ist im Pop-Up Menü Suchmodus der Listeninhalt Teilwortsuche aktiviert. Darüber hinaus muss die Check Box Alphabetische Liste deaktiviert sein.

Ist eine Index Suche gewünscht, so muss das im Pop-Up Menü Suchmodus der Listeninhalt Teilwortsuche eingestellt werden und zusätzlich muß die Check Box Alphabetische Liste aktiviert werden.

Falls mit der Exakten Suche gearbeitet werden soll, muß das Pup-Up Menü Suchmodus auf den Listeninhalt Exakte Suche eingestellt werden. Ist dies der Fall,

spielt es keine Rolle ob die Check Box Alphabetische Liste aktiviert oder deaktiviert ist. Es wird auf jeden Fall nach der exakten Übereinstimmung mit dem Suchbegriff und einem Schlagwort gesucht.

Im Editierfeld muss der zu suchende Begriff eingetragen werden. Naturgemäß hängt das zu erwartende Ergebnis von dieser Eingabe und dem ausgewählten Suchmodus ab.

Durch drücken des „Abfrage Starten“-Buttons wird eine Suchanfrage angestoßen, d.h. der ausgewählte Suchmodus und der eingetippte Suchbegriff werden an das Skript *oswd-trefferanzeige.pl* übergeben.

Der Button „Zurücksetzen“ dient zur Einstellung des Default Wertes.

5.3.2 *oswd-trefferanzeige.pl*

Dieses Skript liest über die Umgebungsvariablen den Suchmodus und den zu suchenden Begriff aus. Als nächstes führt sie die gewünschte Suchart nach dem gewünschten Suchbegriff in der Datenbank aus.

Im Falle einer erfolglosen Abfrage generiert dieses Skript die nachfolgende HTML-Seite:



Abb 15.: Warnmeldung über fehlgeschlagene Suche

Fällt das Ergebnis dagegen positiv aus, so werden die ersten 20 Treffer, falls vorhanden, ausgegeben. Die Suchtreffer werden in Tabellenform aufgelistet. Jedes gefundene Schlagwort bzw. Synonym wird am Anfang einer neuen Tabellenzeile angezeigt. Handelt es sich bei dem Treffer um ein Schlagwort, so wird dieses als Link dargestellt. Der Rest dieser Zeile bleibt leer. Ist der Treffer dagegen ein Synonym, so wird dieses ganz normal dargestellt. Hinter dem Synonym wird die Zeichenkette „verwende:“ ausgegeben. Ihr wiederum folgt ein verlinkter Begriff. In der Abb. 16 kann dieses Anzeigeschemata nachvollzogen werden.

Suchergebnis aus dem OSWD - Schlagwortvokabular
Bibliotheksservice-Zentrum Baden-Württemberg

Gefundene Schlagwörter:

Gummibär	verwende:	Gummibärchen
Gummibärchen		
Gummiband		
Gummibolzenkupplung	verwende:	Bolzenkupplung
Gummidichtung		
Gummidruck	verwende:	Flexodruck
Gummidrucktuch		
Gummielastizität		
Gummierung		
Gummierzeugnis	verwende:	Gummiprodukt
Gummife der		
Gummiformteil		
Gummihandel		
Gummierstellung		
Gummindustrie		
Gummiknüppel	verwende:	Schlagstock
Gummikupplung		
Gummilnse	verwende:	Zoomobjektiv
Gummimarkt		
Gummimehl		

[Weiter](#)

Abb 16.: Bildschirm-Maske oswd-trefferanzeige.pl

Hinter den Links verbergen sich die zu verwendenden Schlagwörter.

Wird bei der Suche auf ein Schlagwort getroffen, wird aus der Datenbanktabelle BEGRIFFE die dazugehörige SWD-Nummer ausgelesen. Das Schlagwort wird in einen Link umgewandelt, der auf das Skript *oswd-einzelanzeige.pl* verweist. Zusätzlich wird diesem Link per Get-Methode die SWD-Nummer angehängt. Dadurch kann das aufgerufene Skript die SWD-Nummer aus der Umgebungsvariable auslesen.

Trifft man bei der Suche auf ein Synonym, wird erst einmal über die Datenbanktabelle ISA_SYNONYM, die SWD-Nummer ermittelt. Über die SWD-Nummer wird dann das gesuchte Schlagwort abgefragt. Ab hier erfolgt die schon oben erwähnte Prozedur zur Linkerstellung eines Schlagwortes.

Wenn man nach bestimmten Begriffen durch die Teilwortsuchfunktion fahndet, wird oben rechts in der Ecke die Anzahl der gefundenen Treffer angezeigt. Links neben dieser Anzeige steht, in welchem Zählraum sich die aktuelle Anzeige gerade befindet, z. B. 1 bis 20 von 165. Bei der Index Suche entfällt dagegen diese Zählfunktion.

Wie oben bereits erwähnt, werden immer nur 20 Suchtreffer – falls vorhanden - auf dem Bildschirm angezeigt. Möchte man zu den nächsten 20 oder vielleicht zu den vorigen 20 Treffern gelangen, muß man mit Hilfe des Zurück-Links bzw. Weiter-Links durch die Gesamttrefferanzahl navigieren.

Hierbei wird dasselbe Skript noch einmal aufgerufen. Allerdings mit anderen Start- und Endparametern.

Es ist noch zu erläutern, warum die Anzeige der Treffer auf jeweils 20 beschränkt ist. Die Schlagwortnormdatei ist urheberrechtlich geschützt. Das Bibliotheksservice-Zentrum Baden-Württemberg ist im Besitz einer Lizenz der SWD. Somit darf innerhalb des BSZ mit der Schlagwortnormdatei gearbeitet werden. Es muß aber sichergestellt sein, dass über das WWW in einfacher Weise kein kompletter Abzug der SWD erhalten werden kann. Da innerhalb dieses Projekts mit Links und der Get-Methode gearbeitet wird, stecken in der Navigationsleiste des Browsers mögliche Angriffspunkte. Dies soll anhand der folgenden Abbildung gezeigt werden.

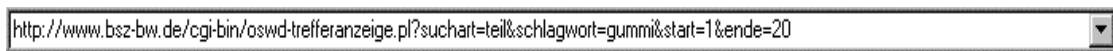


Abb 17.: Navigationsleiste eines Browsers

Ein potentieller Angreifer kann in der Navigationsleiste des Browsers die Start- und Endwerte manuell verändern, um ein anderes Anzeigeverhalten zu erreichen.

Das Skript *oswd-trefferanzeige.pl* ist gegen so einen Angriff gerüstet. Es überprüft am Programmfang ob die Start- und Endwerte manipuliert worden sind. Falls ja, werden die Manipulationen in dem Umfang unschädlich gemacht, dass ab dem Startwert trotz erhöhtem Endwert nur zwanzig Treffer angezeigt werden. Dem Urheberrecht ist somit Genüge getan.

5.3.3 *oswd-einzelanzeige.pl*

Das Skript *oswd-einzelanzeige.pl* erhält per Umgebungsvariable eine SWD-Nummer. Die Aufgabe dieses Skriptes ist es, alle Informationen z.B. Schlagwort, Synonyme, Oberbegriff oder verwandte Begriffe, die zu dieser SWD-Nummer gehören anzuzeigen. Hierzu werden alle vorhandenen Datenbanktabellen durchsucht. Ein mögliches Suchergebnis könnte wie folgt aussehen.

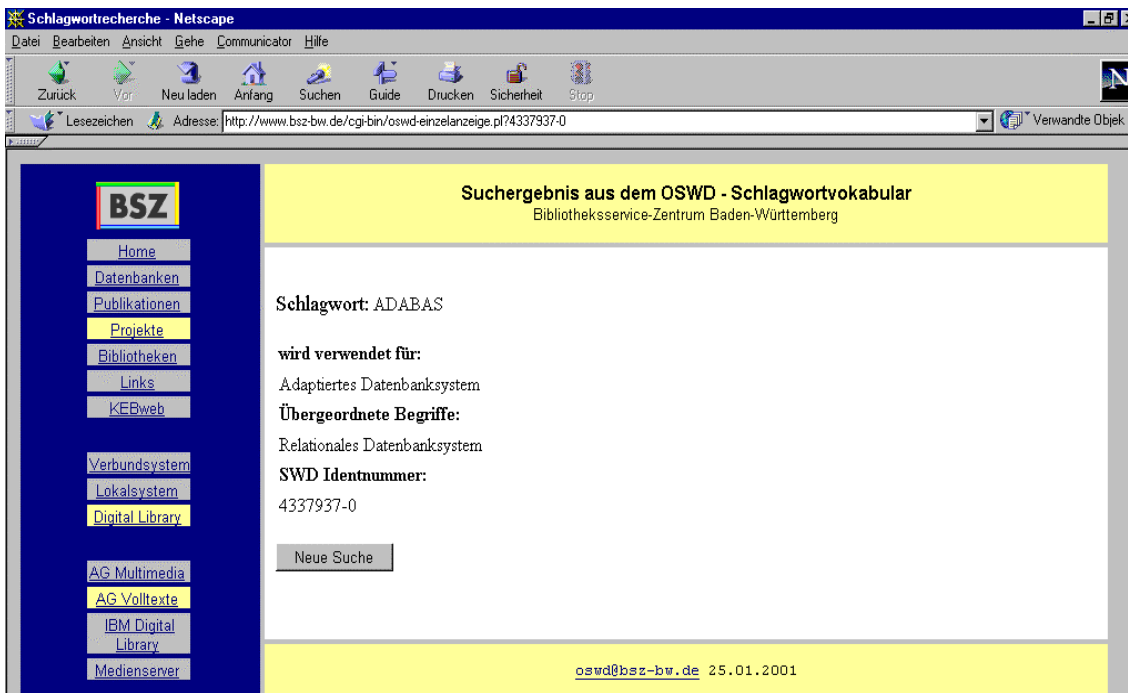


Abb 18.: Bildschirm-Maske *oswd-einzelanzeige.pl*

Durch das Drücken auf den „Neue Suche“ Button gelangt man wieder auf die Einstiegsmaske *oswd-suchabfrage.pl*

Quellenverzeichnis

- [HACK92] Hacker R.: Bibliothekarisches Grundwissen
K G Saur München London New York Paris, 1992
6., völlig neu bearbeitete Auflage, ISBN 3-598-11078-2
- [HAJJ98] Hajji, F.:Perl: Einführung, Anwendungen, Referenz
Bonn: Addison Wesley, 1998
ISBN 3-8372-1335-x
- [KRUE00] Krüger, G.: Go to Java 2
Addison-Wesley, 2000
2. Auflage, ISBN 3-8273-1710-X
- [LAUR99] Laurie B. & Laurie P.:Apache, Das umfassende Referenzwerk
O'Reily, 1999
Deutsche Ausgabe der 2. Auflage, ISBN 3-89721-127-0
- [MAUR96] Maurer R.: HTML und CGI PROGRAMMIERUNG
dpunkt, 1996
ISBN 3-920993-28-4
- [MUEN98] Münz S.: HTML 4.0 Das Handbuch
Franzis Verlag GmbH, 1998
ISBN: 3-772375-14-6
- [SCHM99] Schmid E., Cartus C., Blume R.: php dynamische webauftritte
professionell realisieren
Markt & Technik, 1999
ISBN 3-8272-5524-4