

dbi-materialien

185

Das Sondersammelgebiets-
Fachinformationsprojekt
(SSG-FI)
Göttingen

Dokumentation – Teil 1

DEUTSCHES BIBLIOTHEKSINSTITUT



**Das Sondersammelgebiets-
Fachinformationsprojekt
(SSG-FI)**

der

**Niedersächsischen
Staats- und Universitätsbibliothek
Göttingen**

**GeoGuide, MathGuide, Anglo-American History Guide und
Anglo-American Literature Guide**

(www.SUB.Uni-Goettingen.de/ssgfi/)

**Dokumentation
Teil 1**

**Berlin 1999
DEUTSCHES BIBLIOTHEKSINSTITUT**

dbi-materialien ; 185

Schriften der Deutschen Forschungsgemeinschaft

Das Projekt „SSG-Informationssysteme im Internet“ wurde gefördert von der Deutschen Forschungsgemeinschaft.

Projektleiter: *Dr. Wilfried Enderle*

Redaktion: *Dr. Thomas Fischer, Dr. Heike Neuroth*

Die Deutsche Bibliothek– CIP-Einheitsaufnahme

Niedersächsische Staats- und Universitätsbibliothek <Göttingen>: Das Sonder-sammelgebiets-Fachinformationsprojekt (SSG-FI) der Niedersächsischen Staats- und Universitätsbibliothek Göttingen : GeoGuide, MathGuide, Anglo-American History Guide und Anglo-American Literature Guide (www.SUB.Uni-Goettingen.de/ssgfi/) ; Dokumentation / Deutsches Bibliotheksinstitut [Projektleiter: Wilfried Enderle]. – Berlin : Dt. Bibliotheksinst.

Teil 1 (1999)

(Dbi-materialien ; 185 : Schriften der Deutschen Forschungsgemeinschaft)

ISBN 3-87068-985-4

Berlin, April 1999

Herstellung und Vertrieb:

Deutsches Bibliotheksinstitut

Kurt-Schumacher-Damm 12 - 16

13405 Berlin

Druck: Ernst Knoth, 49303 Melle

Gedruckt auf alterungsbeständigem Papier.

Diese Veröffentlichung entstand mit Förderung des Bundes und der Länder

Vorwort

Die vorliegende umfangreiche Dokumentation des SSG-FI-Projektes bietet einen detaillierten Einblick in die inhaltliche wie technische Projektkonzeption. Zugleich wird versucht, auch den Kontext, die neuen Aufgaben und Probleme, mit denen Bibliotheken seit dem Siegeszug des Internet konfrontiert werden, zu thematisieren, um verständlich zu machen, warum dieses Projekt notwendig erschien und auch, warum es so und nicht anders konzipiert und realisiert wurde. Ziel der Dokumentation ist es mithin, sowohl dem allgemein an elektronischen Fachinformationen interessierten Bibliothekar und Dokumentar Informationen zum konzeptionellen Hintergrund des Projektes zu geben, als auch denjenigen, die dieses Konzept konkret nutzen wollen, detaillierte technische Anleitungen und konkrete Handreichungen zu bieten. Die Dokumentation wendet sich daher an unterschiedliche Leser, für die jeweils unterschiedliche Teile von Interesse sein dürften. Auch wenn natürlich nicht davon abgeraten werden soll, diese Dokumentation vollständig zu lesen, so soll doch vorab kurz erläutert werden, welche Kapitel sich speziell an welche Adressaten wenden.

1. *Allgemeine Informationen zur Problematik elektronischer Fachinformationen* und deren Erschließung sind in den Kapiteln 1 und 2 zu finden. Hier wird, allgemein formuliert, das Thema Internet und Bibliotheken behandelt.
2. Einen *Überblick über die Projektkonzeption*, eine Art Executive Summary für den etwas eiligeren Leser, bietet Kapitel 3.
3. Die *Suchmöglichkeiten aus Benutzersicht* sowie *die sachlichen Erschließungskonzepte* werden detailliert in den Kapiteln 4 und 5 vorgestellt. Wissenschaftliche Bibliothekare und Dokumentare, die das bibliothekarische Konzept en détail kennenlernen und ggf. vielleicht auch in der Praxis anwenden wollen, seien auf diese Kapitel verwiesen.
4. Das *technische Konzept* wird ausführlich in Kapitel 6 dargelegt. Dieses Kapitel ist für Systemadministratoren gedacht, die sich für technische Details der implementierten Datenbank- und Webserver-Lösung interessieren oder das System selbst in dieser Form nachnutzen, also selbst implementieren wollen.
5. Die laufende technische und inhaltliche *Pflege des Systems*, aber auch technische und konzeptionelle *Entwicklungsperspektiven* werden in den Kapiteln 7, 8 und 9 thematisiert. Sowohl praktische Anwender als auch allgemein interessierte Leser können sich hier informieren, wie Aufgaben und Probleme des Praxisbetriebs gelöst werden können.

Die einzelnen *Fachguides* des SSG-FI-Systems, also der Anglo-American History Guide, der Anglo-American Literature Guide sowie der GeoGuide und der

MathGuide werden in einem zweiten Teil, der separat veröffentlicht werden wird, vorgestellt werden. Die Beschreibungen dieser Fachguides wenden sich an den am jeweiligen Fach interessierten Leser, wobei das Projekt so angelegt wurde, daß sowohl geistes- und kulturwissenschaftliche (Anglistik, Geographie, Geschichte) als auch naturwissenschaftliche Disziplinen (Geowissenschaften) sowie Grundlagenfächer (Reine Mathematik) als Beispiel für benachbarte Bereiche fungieren können.

Inhalt

1 Sondersammelgebietsbibliotheken und das Internet	1
Von Wilfried Enderle	
1.1 Bibliotheken, digitale Medien und globale Wissenschaftsnetze	3
1.2 SSG-Bibliotheken und das Konzept der virtuellen Fachbibliotheken	7
1.3 Andere Subject-Gateway-Projekte	8
2 Erschließungsprobleme von Internet-Ressourcen	11
Von Wilfried Enderle, Thomas Fischer und Karsten Hülsemann	
2.1 Das Internet als wissenschaftliche Informationsquelle	11
2.2 Suchen im Internet: Hoffnungslos?	13
2.2.1 Besser suchen: Automatische Indizierung	13
2.2.2 Mehr Informationen bieten: Metadaten	14
2.2.3 Automatisierte versus intellektuelle Erschließung	16
2.3 Das Problem der Auswahl- und Bewertungskriterien	19
2.4 Infrastrukturelle Defizite des Internet	22
3 Die Konzeption des SSG-Fachinformationssystems im Überblick	25
Von Thomas Fischer, Karsten Hülsemann, Heike Neuroth und Norbert Pfurr	
3.1 Definition der Informationsquellen	28
3.2 Das Metadatenformat	29
3.3 Das sachliche Erschließungskonzept	30
3.4 Auswahl- und Bewertungskriterien	31
3.4.1 Die einzelnen Kategorien der Bewertung	32
4 Das SSG-Fachinformationssystem aus Benutzersicht	35
Von Thomas Fischer und Heike Neuroth	
4.1 Blättern („Browsen“) in den SSG-FI-Guides	37
4.2 Suche im SSG-FI System	41
4.2.1 Suche in den Guides des SSG-FI Projektes	42
Einfache freie Textsuche	42
Suche mit booleschen Operatoren	43
Suche in SSG-FI-Kategorien	44
4.2.2 Die Benutzerschnittstelle des Suchsystems	44
Die einfache Suche („Simple Search“)	44
Die komplexe Suche („Advanced Search“)	45
Das Suchergebnis	46

5 Das Erschließungskonzept des SSG-Fachinformationssystems 47

Von Thomas Fischer, Karsten Hülsemann, Heike Neuroth und Norbert Pfurr

5.1 Der Prozeß der Erschließung.....	47
5.2 Das SSG-FI Metadatenkonzept	49
5.2.1 Die Kategorien des SSG-FI-Systems.....	49
5.2.2 Dublin Core Metadaten	54
Die Syntax von Dublin Core in HTML.....	55
Die Semantik von Dublin Core	56
5.3 Die formale Erschließung.....	59
5.3.1 Der <i>Formal Key</i> und der <i>Source Type Catalog</i>	59
5.3.2 Ländercodes nach ISO 3166.....	62
5.3.3 Sprachcodes nach ISO 639-1	68
5.4 Das Grundkonzept der sachlichen Erschließung.....	70
5.4.1 Die Fachklassifikation	70
5.4.2 Basisklassifikation und Dewey Decimal Classification	71
5.4.3 Verbale sachliche Erschließung durch Schlagworte.....	72
5.5 Die Auswahl- und Bewertungskriterien im Projekt SSG-FI	72
5.5.1 Die einzelnen Kategorien der Bewertung	72
Content	73
Clarity.....	74
Index.....	75
Links	76
Level.....	77
Backlinks.....	78

6 Die technische Implementierung des SSG-Fachinformationssystems..... 81

Von Thomas Fischer

6.1 Datensammlung und -verwaltung: das Allegro-System	81
6.1.1 Die Dateien des Datenordners SSG-XYZ.....	82
6.1.2 Die Dateien des Ordners SSGFI	86
Anzeige und Voreinstellungen.....	88
Hilfsprogramme.....	89
Benutzte Datentypen.....	90
6.1.3 Eingabe in die Allegro-Datenbank	91
6.2 Datenpräsentation: der SSG-FI-Server	92
6.2.1 Die Dateistruktur des Servers	92
6.2.2 Die Dateien des Hauptordners	94
6.2.3 Die Suchdateien.....	96
6.3 Der Export der Daten in das Internet.....	96
6.3.1 Die Export-Routinen.....	96
6.3.2 Spezielle Datensätze	97
Ersetzung der Sprach- und Ländercodes	98
Ersetzung der Formal- und Themenschlüssel	98
Konkordanzen von GOK zu anderen Klassifikationen	99

6.3.3 Die Exportprogramme	99
Export der Infodateien	100
Export der Indexdateien	100
Export der Mini-Templates	100
Export der Verteilerdokumente	102
Gesamlexport mit Mini-Templates.....	103
Für den Export benutzte Tabellen	103
6.3.4 Einrichtung des Internetservers.....	104
6.4 Zusätzliche Einrichtungen: Suche und Formulareingabe.....	106
6.4.1 Die technische Implementierung der Suche	106
Das Indexprogramm <i>swish-e</i>	106
Die Bearbeitung der Suchanfrage: das CGI-Script <i>suche.cgi</i>	108
6.4.2 Die technische Implementierung der Dateneingabe über Internet	109
Dateneingabe per HTML-Formular	109
Die Bearbeitung der Benachrichtigung: das CGI-Script <i>ssgfi.cgi</i>	110

7 Inhaltliche und technische Verwaltung des SSG-Fachinformationssystems 111

Von Thomas Fischer, Karsten Hülsemann und Heike Neuroth

7.1 Recherchestrategien beim Aufbau der SSG-FI-Guides	111
7.1.1 Systematische Ansätze.....	112
7.1.2 Nutzung vorhandener Suchverzeichnisse	113
7.1.3 Nutzung kommerzieller Suchmaschinen	114
7.2 Pflege und Administration des Systems	115
7.2.1 Prüfung der Gültigkeit von URLs.....	116
7.2.2 Prüfung thematischer Sites auf Veränderungen	117
7.2.3 Vollständigkeit der thematischen Guides	118
7.3 Statistische Analysen.....	119
7.3.1 Generelles	121
7.3.2 Statistische Auswertung und Nutzerprofile am Beispiel der SSG-FI-Nutzung 1998.....	125

8 Modelle kooperativer Pflege der SSG-FI-Guides 137

Von Wilfried Enderle

9 Ausblick und weitere Entwicklungsmöglichkeiten 143

Von Wilfried Enderle

9.1 Dynamische http-Schnittstelle.....	143
9.2 Ausbau um die Daten gedruckter Informationsquellen	144
9.3 Einspielen der SSG-FI-Datensätze in den lokalen Online-Katalog	145
9.4 Das SSG-FI-System als Basis für die automatisierte Katalogisierung einzelner Internetdokumente.....	146

1 Sondersammelgebietsbibliotheken und das Internet

Von Wilfried Enderle

In Deutschland organisieren Bibliotheken kooperative Aufgaben auf zwei sich überlagernden und ergänzenden Ebenen: Der regionalen Ebene, zumeist begrenzt durch die jeweiligen Verbundsysteme, und der nationalen. Auf nationaler Ebene bildet das *System der überregionalen Literaturversorgung* eine Art *nationaler verteilter Forschungsbibliothek*, organisiert und mitfinanziert von der Deutschen Forschungsgemeinschaft (DFG) und getragen von vier zentralen Fachbibliotheken und 36 Sondersammelgebietsbibliotheken sowie einigen weiteren Sammelschwerpunktbibliotheken. Aufgabe der SSG-Bibliotheken ist es, für ihre jeweiligen Sondersammelgebiete die wissenschaftliche Literatur so komplett wie möglich zu sammeln, um zu gewährleisten, daß in Deutschland alle wissenschaftlich relevanten Publikationen zumindest in einem Exemplar vorhanden sind. Die SSG-Bibliotheken verpflichten sich, ihre mit Hilfe der DFG erworbenen Bestände in nationalen Direkt- und Fernleihdiensten allen wissenschaftlichen Nutzern zur Verfügung zu stellen. Mit der Erwerbung und Zurverfügung-Stellung ist zugleich auch die Aufgabe der dauerhaften Archivierung dieser Bestände verbunden¹.

Dieses hier grob skizzierte Konzept wurde entwickelt für das Medium des gedruckten Buches; die Dienstleistungen von SSG-Bibliotheken, die sich in den Jahrzehnten seit der Etablierung dieses Systems der überregionalen Literaturversorgung im Jahre 1949 in unterschiedlicher Ausprägung entfaltet haben, waren und sind hierauf zugeschnitten. Spätestens mit der Etablierung des World Wide Web wurde indes deutlich, daß sich die wissenschaftlichen Kommunikations- und Publikationsformen im Zeitalter digitaler Medien und globaler Wissenschaftsnetze wandeln und daß zunehmend wissenschaftliche Informationen und Publikationen auch oder auch nur in digitaler Form existieren. Mit der Digitalisierung der Medien war der rasante Siegeszug des Internet eines der wesentlichsten und die bisherigen Informationsstrukturen am einschneidendsten verändernden Ereignisse. Zum erstenmal wurden – zum Teil – wissenschaftliche Informationen frei auf Servern zur allgemeinen Nutzung aufgelegt; zugleich entstanden neue Kommunikations- und Publikationsformen, wie elektronische Diskussionsgruppen oder thematische Sites. Die DFG hat auf diese Veränderungen mit einem Memorandum zur „Weiterentwicklung der überregio-

1 Vgl. zum SSG-Konzept: Richtlinien zur Abgrenzung von Sondersammelgebieten und zur Beschaffung von Literatur, Deutsche Forschungsgemeinschaft, Bonn 1997; sowie zum Hintergrund: Überregionale Literaturversorgung von Wissenschaft und Forschung in der Bundesrepublik Deutschland. Denkschrift. Bibliotheksausschuß der Deutschen Forschungsgemeinschaft, Boppard 1975

nenal Literaturversorgung“ reagiert, um das SSG-Konzept den neuen informationstechnischen Rahmenbedingungen anzupassen.²

Verschiedene SSG-Bibliotheken entwickeln mittlerweile in einer Reihe von Projekten organisatorische und technische Formen zur Integration der neuen digitalen Medien und Informationsformen in das SSG-Konzept.³ Das Projekt Sondersammelgebiets-Fachinformationen (SSG-FI-Projekt) der Niedersächsischen Staats- und Universitätsbibliothek ist eines dieser Projekte. Ausgangspunkt beim Entwurf des Grundkonzeptes zu Beginn des Jahres 1996 war, daß im Internet zunehmend wissenschaftlich relevante Informationsquellen entstanden. Wollte eine SSG-Bibliothek weiterhin umfassend fachlichen „Bestandsaufbau“ betreiben, also alle relevanten Publikationen für ihr SSG dem Wissenschaftler zugänglich machen, wurde es notwendig, auch die Informationsmöglichkeiten des Internet in irgendeiner Form in ihr Dienstleistungsspektrum aufzunehmen.

Das SSG-FI-Projekt konzentrierte sich von vornherein auf wissenschaftlich relevante Fachinformationen im Internet. Ziel war es, sogenannte Subject Based Information Gateways für ausgewählte SSGs (Reine Mathematik, Geowissenschaften, Geographie und thematische Karten sowie Geschichte, Politik und Literatur des anglo-amerikanischen Kulturraums) aufzubauen – den MathGuide, GeoGuide sowie den Anglo-American History Guide und den Anglo-American Literature Guide. Es wurde mit dem Konzept der Subject Gateways also gezielt ein Teilbereich aus dem weiten Feld der digitalen Medien und neuen Informationssysteme herausgenommen, um hierfür eine prototypische, zukunftsfähige Lösung zu entwickeln. Um die Funktion und Bedeutung des SSG-FI-Systems innerhalb des gesamten Aufgabenkomplexes einer umfassenden digitalen Fachbibliothek einordnen zu können, sollen zunächst einmal die wesentlichen neuen Aufgaben skizziert werden, welche durch die Digitalisierung der Medien und die globalen Wissenschaftsnetze entstanden sind (Kap. 1.1.) sowie das Konzept der virtuellen Fachbibliotheken, wie es innerhalb der Subito-Initiative

2 Deutsche Forschungsgemeinschaft: Weiterentwicklung der überregionalen Literaturversorgung. Memorandum. In: Zeitschrift für Bibliothekswesen und Bibliographie, Bd. 45 (1998) S. 135-164. Siehe auch <http://www.dfg.de/foerder/biblio/memeeo.html>

3 Den Ausgangspunkt bildete in gewisser Hinsicht das Projekt WEBIS der Staats- und Universitätsbibliothek Hamburg (<http://webis.sub.uni-hamburg.de>), das einen allgemeinen Einstieg in die gesamten Dienste aller SSG-Bibliotheken über einen Webserver bietet. Ferner wären in diesem Kontext die SSG-S-Projekte der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB), der Senckenbergischen Bibliothek Frankfurt/Main, der Universitäts- und Landesbibliothek Saarbrücken und der Universitätsbibliothek Tübingen sowie das SSG-Current-Contents-Projekt der SUB Göttingen zu nennen. Dazu kommen mittlerweile die verschiedenen Projekte zum Aufbau virtueller Fachbibliotheken an der SUB Göttingen, der UB/TIB Hannover, der UStB Köln sowie der ULB Saarbrücken.

von der Subito-Arbeitsgruppe 2 erarbeitet wurde (Kap. 1.2.). Ein kurzer Überblick über vergleichbare internationale Projekte weist den derzeitigen Stand der Entwicklung in diesem Bereich auf (Kap. 1.3.), bevor auf die Erschließungsprobleme von Internet-Ressourcen (Kap. 2) und die Konzeption des SSG-FI-Systems detailliert eingegangen wird (ab Kap. 3).

1.1 Bibliotheken, digitale Medien und globale Wissenschaftsnetze⁴

Zur Zeit erleben wir, wie die Bibliotheken der Gutenberg-Ära, der Gutenberg-Galaxis, wie man mit McLuhan⁵ sagen könnte, mit einer neuen, medientechnischen Revolution konfrontiert werden: derjenigen der Digitalisierung und globalen Vernetzung. Es ist keine Frage, daß sich auch der Charakter der Bibliotheken mit der Entstehung der neuen Medienform der digitalen Medien ändern wird, wie in der Vergangenheit mit dem Buchdruck. Doch was hat sich denn für die Bibliotheken geändert? Welche Folgen haben die neuen digitalen Medien und deren globale Vernetzung über das Internet für die Bibliotheken, und welche Konzeptionen entwickeln diese, um ihre Funktionen auch in einer neuen, medientechnischen Epoche wahrnehmen zu können?⁶

Um diese Fragen beantworten zu können, muß man sich zunächst einmal vor Augen halten, was konkret geschehen ist; in welcher Form sich die neuen Medien vom dem bisherigen, für die Bibliotheken zentralen Publikationsmedium des Buches unterscheiden. Dabei dürfte es sinnvoll sein, sich die wesentlichen Funktionen der bisherigen Medien und die - noch - existierende wissenschaftliche Informationsinfrastruktur anzusehen, wobei in diesem Rahmen freilich nur einige wesentliche und allgemeine Gesichtspunkte angesprochen werden können. Dabei ist zweierlei festzuhalten:

-
- 4 Ein Teil dieses Kapitels basiert auf: Wilfried Enderle, Quo vadis SSG? Die deutsche Sondersammelgebietsbibliothek auf dem Weg zur virtuellen Fachbibliothek. In: Die Rolle der Archive in Online-Informationssystemen. Workshop im Staatsarchiv Münster, 8.-9. Juli 1998, hrsg. von B. Bischoff u. W. Reininghaus (=Veröffentlichungen des Staatsarchivs Münster) Münster [voraussichtlich 1999]
 - 5 Vgl. Marshall McLuhan, Die Gutenberg-Galaxis. Das Ende des Buchzeitalters, [2. Aufl.], Bonn-Paris-Reading/Mass. 1995
 - 6 Die Literatur zum Thema digitale Bibliothek ist bereits nach wenigen Jahren fast unübersehbar geworden. Zum aktuellen Stand der Diskussion vgl. die Beiträge in: International Journal on Digital Libraries, Bd. 1ff, 1997ff; D-lib magazine. The Magazine of the Digital Library Research (<http://www.dlib.org>) Juni 1998. Siehe auch mit eher informationstechnischen Schwerpunkten: Digital Libraries. Research and Technology Advances, hrsg. von Nabil R. Adam u.a. (=Lecture Notes in Computer Science, Bd. 1082) Berlin-Heidelberg-New York u.a. 1996; Research and Advanced Technology for Digital Libraries, hrsg. von Carol Peters u.a. (=Lecture Notes in Computer Science, Bd. 1324) Berlin-Heidelberg-New York u.a. 1997

1. Das gedruckte Buch ist ein distinktes, abgeschlossenes und nicht mehr veränderbares Medium, das im wesentlichen eine lineare Rezeption begünstigt. Andere Medien wie Film oder Ton existieren separat und zusätzlich daneben. Dieses Medium setzt zur Sammlung wie zur Nutzung einen konkreten Ort, eben die Bibliothek, voraus.
2. Es hat sich im Laufe der letzten hundert Jahre eine feste informationstechnische Infrastruktur und Arbeitsteilung herausgebildet, bei der Bibliotheken eine definierte Rolle spielen und die zugleich für die Arbeit der Bibliothek, den zielgerichteten inhaltlichen Bestandsaufbau von entscheidender Bedeutung ist. Diese Infrastruktur sieht in idealtypischer Vereinfachung so aus: Für die Produktion von Büchern und Zeitschriften sind Verlage oder wissenschaftliche Kommissionen zuständig; Nationalbibliotheken sammeln und erfassen dieses Schrifttum komplett und zeigen es über laufend erscheinende Nationalbibliographien an. Fachinformationseinrichtungen erstellen auf dieser Basis Fachbibliographien; Rezensionsorgane verzeichnen die wesentlichste Fachliteratur. Auf dieser Basis können Bibliotheken schnell die wichtigste neu erschienene Literatur auswählen, sammeln und entsprechend den spezifischen Bedürfnissen ihrer Nutzer erschließen.⁷

Was hat sich nun mit dem Internet und der Digitalisierung der Medien verändert? Die bisherige technische Trennung verschiedener Medien - Schrift, Ton, Bild, Film - ist entfallen, da die technische Basis all dieser Medienformen mittlerweile dieselbe ist. Ob Schrift, Ton oder Bild, sie werden in technischer Hinsicht alle in digitaler Form transportiert, also über elektrische Zustände, welche die Zahlen 0 oder 1 repräsentieren. Die Konsequenzen davon sind: Modifizierbarkeit, Multimedialität, Hypertextualität und Vernetzbarkeit aller Komponenten.⁸ Dies bedeutet konkret, daß Daten technisch nicht mehr endgültig fixiert sind wie beim Buchdruck. Digitale Medien können im Prinzip ständig und an jedem Computer modifiziert werden. Dazu kommt, daß die Speicherform in ganz anderer Weise von der technischen Infrastruktur abhängt. Texte, Bilder und andere Medienformen liegen auf Servern, wobei es aufgrund der globalen Vernetzung nicht einmal notwendig ist, daß eine inhaltlich zusammengehörende Publikation physisch auf einem Server liegt. Durch Verlinkung können einzelne Teile einer solchen Publikation auf verschiedenen, weltweit verteilten Rechnern aufliegen. Anders als beim Buch, das, wenn es einmal produziert, also gedruckt

7 Einen guten Überblick über die Komplexität der modernen Informationsstruktur gibt: Grundlagen der praktischen Information und Dokumentation. Ein Handbuch zur Einführung in die fachliche Informationsarbeit, hrsg. von Marianne Bude u.a., Bde. 1-2, 4. Aufl., München-New Providence-London-Paris 1997

8 Vgl. als verständliche Einführung in die Struktur digitaler Medien: Norbert Gabriel, Kulturwissenschaften und neue Medien. Wissensvermittlung im digitalen Zeitalter, Darmstadt 1997, S. 49ff

und gebunden war, für seine Existenz nur einen Raum brauchte, wo es aufbewahrt wurde, benutzen digitale Internetpublikationen das dauernde Vorhandensein eines technisch komplexen, globalen Netzwerkes von Computern. Dazu kommt, daß die digitalen Medien auch in anderer technischer Form erschlossen und verwaltet werden können als ihre gedruckten Pendanten. Programme, Robots und Suchmaschinen können die Erschließung automatisieren und mit Hilfe standardisierter Metadaten ist auch eine automatisierte Administration digitaler Dokumente vorstellbar.

Diese Möglichkeiten werden erkauft durch eine hohe technische Komplexität, welche den digitalen Medien freilich auch eine ganz neue Mächtigkeit verschafft; ihre Funktionen hinsichtlich Darstellungsmöglichkeiten, Schnelligkeit der Produktion, die mit der Publikation fast identisch ist, sowie den Zugriffsmöglichkeiten für Leser, vor allem aber der Möglichkeit, direkte Links, Verknüpfungen zu anderen digitalen Medien in beliebiger Zahl und Komplexität zu erstellen, scheinen die Informationsmöglichkeiten gegenüber dem Printmedium ungeheuer zu potenzieren. Eine konkrete Folge ist, daß ganz neue Formen oder Typen von Publikationen entstehen, die es beim gedruckten Medium so nicht gab. Dazu gehören zum Beispiel die thematischen Sites im Internet, die auf eine neue Art und Weise unterschiedliche Informationsquellen zusammenfügen, wie bibliographische Datenbanken und Literaturlisten, Quellentexte, elektronische Diskussionslisten mit ihren Archiven, Preprints und Working Papers, aber auch elektronische Zeitschriften und Linklisten mit Verknüpfungen zu anderen thematisch relevanten Sites.

Zusammengefaßt haben sich mithin für Bibliotheken im wesentlichen zwei Aspekte verändert: Die technische Form des digitalen Mediums, welche eine neue Form der Erschließung und Verwaltung erfordert. Und zum anderen ist die organisatorische Infrastruktur für die neuen digitalen Medien selbst im Umbruch begriffen.

Da mittlerweile eine Fülle wissenschaftlich relevanter Informationen und Quellentexte im Internet aufliegt⁹, wird von der Prämisse ausgegangen, daß die Bibliotheken – unabhängig von der grundsätzlichen Problematik des neuen Mediums¹⁰ – sich darum bemühen müssen, dieses neue Medium in ihr Angebot

9 Vgl. nur für die Geschichtswissenschaft: Dennis A. Trinkle/Dorothy Auchter/Scott A. Merriman/Todd E. Larson, *The History Highway. A Guide to Internet Resources*, Armonk-London 1997; bedeutend schwächer das deutsche Pendant: Christian von Dittfurth, *Internet für Historiker*, 2. Aufl., Frankfurt-New York 1998

10 In diesem Beitrag wird nicht die grundsätzliche Frage thematisiert, ob das Internet überhaupt ein geeignetes Medium für wissenschaftliche Publikationen ist. Diese Frage ist zweifelsohne berechtigt und legitim, ja, mehr noch, sie ist es wert, intensiv diskutiert zu werden, intensiver jedenfalls, als dies gegenwärtig – zumindest im deutschen Bibliothekswesen – geschieht. Doch können andererseits – und von die-

zu integrieren. Für die Bibliotheken ergeben sich dabei freilich mehrere Schwierigkeiten, die zu einem großen Teil mit der sich noch entwickelnden organisatorischen Infrastruktur des Netzes zusammenhängen, zum Teil aber auch dem neuen Medium inhärent sind.

Ein Grundproblem ist, daß die Grenze zwischen Kommunikation und Publikation sehr viel unklarer ist als bei gedruckten Materialien. Dies liegt zum einen daran, daß im Netz sehr einfach und schnell Dinge veröffentlicht werden können. Der Erfolg des Internet, insbesondere des World Wide Web, beruht auch darauf, daß es mittlerweile in technischer Hinsicht sehr einfach ist, dort Texte und Bilder öffentlich zugänglich zu machen. Diese Veröffentlichungen haben aber oft nur vorübergehenden Charakter; es handelt sich hier eben auch um Ankündigungen, Projektbeschreibungen und ähnliches, im Grunde also nicht um eigentliche Publikationen, sondern um Informationsmaterial. Die Art und Weise, wie Publikationen im eigentlichen Sinne, also inhaltlich für Wissenschaftler relevantes Material veröffentlicht wird, unterscheidet sich indessen teilweise kaum von der Veröffentlichung dieses Informationsmaterials. Jedermann kann unter einer beliebigen Netzadresse alles auflegen, vom einfachen Informationsmaterial bis zu relevanten Publikationen. Gerade für die Selektion und Kenntlichmachung des letzteren fehlt derzeit aber noch die notwendige Infrastruktur wie sie für das gedruckte Buch existiert.

Dazu kommt des Weiteren, daß sich neue Veröffentlichungsformen entwickelt haben, sogenannte thematische Sites, für die es kein exaktes Pendant im Bereich der gedruckten Medien gibt. Bibliothekare müssen daher neue Konzepte und Regeln zur Typologisierung und Beschreibung dieser neuen Medienformen erarbeiten.

Ein drittes Problem, das mit der (noch) fehlenden organisatorischen Infrastruktur wie auch der technischen Struktur des Web zusammenhängt, ist die Schwierigkeit, einen Überblick über alle relevanten Informationsquellen im Netz zu finden und sich laufend systematisch über neu entstehende Sites zu informieren.

Im SSG-FI-Projekt wurde versucht, ein inhaltliches Konzept zu entwickeln, das für diese drei Problembereiche Lösungsansätze bietet und das – unabhängig von der jeweiligen lokalen technischen Infrastruktur einer SSG-Bibliothek – generell für alle SSGs nachnutzbar ist. Dabei ist freilich zu beachten, daß solche Lösungen immer auch vom technischen und vor allem vom organisatorischen Entwicklungsstand des Internet selbst abhängig sind. Solange, um nur ein Bei-

ser Prämisse wird hier ausgegangen – Bibliothekare nicht darüber entscheiden, welche Medienformen für Wissenschaftler geeignet sind. Sie müssen von dem Faktum ausgehen, daß derzeit das Internet sich als ein wissenschaftliches Kommunikations- und Publikationsmedium etabliert hat.

spiel zu nennen, es keine stabilen URNs gibt, sondern nur labile URLs, kann es keine befriedigende Lösung zur Gewährleistung der Validität des Zugriffs auf Internet-Ressourcen geben. Die Lösung eines solchen, das Internet als ganzes betreffenden Problems, kann jedoch nicht in einem speziellen Fachprojekt entwickelt werden.

1.2 SSG-Bibliotheken und das Konzept der virtuellen Fachbibliotheken.

Ein sehr allgemeines Konzept, wie in Zukunft SSG-Bibliotheken mit den neu entstandenen digitalen Medien umgehen sollen, wurde im Rahmen der Bund-Länder-Initiative SUBITO in der Arbeitsgruppe 2 entwickelt.¹¹ Dieses Modell geht von der Prämisse aus, daß in Zukunft für alle Fächer, wenn sicherlich auch mit unterschiedlicher Intensität, ein Großteil der Informations- und Dokumentbeschaffung über das Web erfolgen wird. Da digitale Medien in technischer Hinsicht nicht mehr standortgebunden sind, ergibt sich die Möglichkeit, diese Medien in Kooperation verschiedener Einrichtungen zu erschließen und den Zugriff darauf zu organisieren und auf Dauer zu gewährleisten. Zugleich wird aber davon ausgegangen, daß weiterhin eine organisatorische Infrastruktur vorhanden sein muß, welche eine derartige Kooperation organisiert.

Deshalb wurde vorgeschlagen, sogenannte „virtuelle Fachbibliotheken“ einzurichten, die jeweils von der zuständigen Fach- oder SSG-Bibliothek organisiert werden sollen. Eine solche virtuelle Fachbibliothek soll Fachinformationseinrichtungen, wissenschaftliche Fachgesellschaften, Verlage und weitere Bibliotheken zusammenschließen, um dem Nutzer einen direkten und möglichst vollständigen Zugriff auf die für ihn relevanten wissenschaftlichen Informationsressourcen zu ermöglichen. Damit soll es möglich sein, fachbibliographische Datenbanken direkt mit Bestandsnachweisen der in der SSG-Bibliothek vorhandenen Bestände zu verknüpfen; zugleich sollen in dieses Konzept digitale Medien, wie elektronische Zeitschriften, aber auch fachliche Internet-Ressourcen integriert werden (vgl. Abb. 1.1).

Eine virtuelle Fachbibliothek geht also, wie betont werden muß, weit über das Konzept eines Subject Gateway hinaus, wie es im Rahmen des SSG-FI-Projektes entwickelt wurde. Ein solcher Subject Gateway stellt freilich ein nicht unwichtiges Modul innerhalb einer virtuellen Fachbibliothek dar, da es über die in Datenbanken nachgewiesenen Einzeldokumente, seien sie gedruckt oder elektronisch, die thematischen Fachinformationen, die eventuell außerhalb des Konsortiums einer virtuellen Fachbibliothek im Internet vorhanden sind, mit e-

¹¹ Bund-Länder-Initiative zur Beschleunigung der Literatur- und Informationsversorgung: Die virtuelle Fachbibliothek. Empfehlungen der Arbeitsgruppe SUBITO.2, 25. März 1998. Diese Empfehlungen decken sich teilweise mit dem Memorandum der DFG (siehe Anm. 2).

schließt. Das SSG-FI-Projekt ist also ein Versuch, ein Modul einer virtuellen Fachbibliothek, wie sie in Zukunft von SSG-Bibliotheken organisiert werden sollen, zu realisieren.

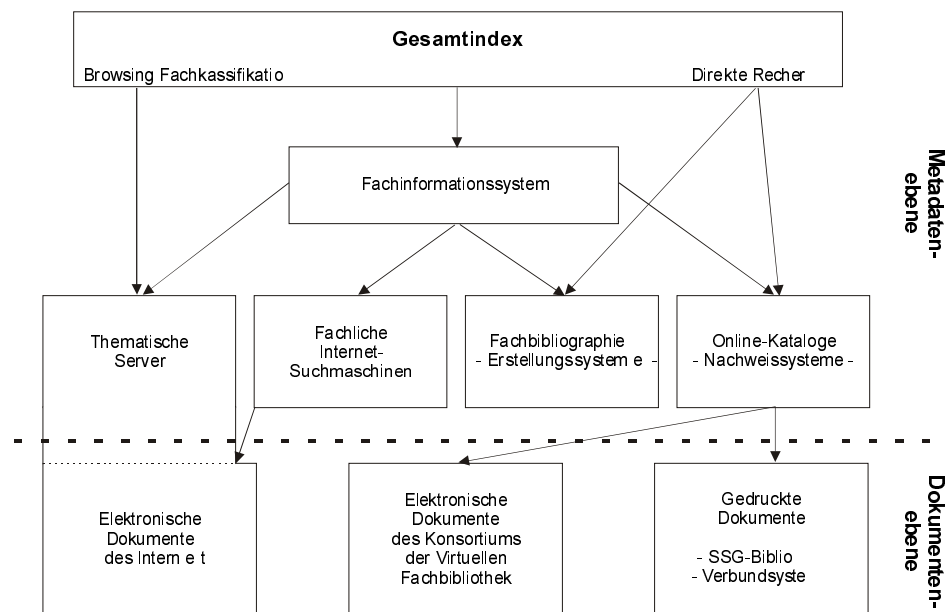


Abb. 1.1: Die Module einer virtuellen Fachbibliothek

1.3 Andere Subject-Gateway-Projekte

Auch wenn das SSG-FI-System innerhalb des deutschen Bibliothekswesens derzeit sicherlich das avancierteste Modell eines Subject Gateway realisiert hat, so gibt es natürlich im internationalen Rahmen vergleichbare Systeme und konzeptionelle Entwicklungen, die auch im Rahmen des SSG-FI-Projektes rezipiert worden sind. Eine wichtige Basis für die inhaltliche Konzeption bildeten die im *DESIRE-Projekt* erstellten Reports, welche unter anderem auch das Thema der Erschließung von Internet-Ressourcen behandeln. Für die Konzeption des SSG-FI-Projektes wurden dabei insbesondere mit berücksichtigt:

- A review of metadata: a survey of current resource description formats. (Deliverable D3.2 Part 1. March 1997)¹²
- Selection criteria for quality controlled information gateways. (Deliverable D3.2 Part 2; May 1997)¹³

12 URL <http://www.ukoln.ac.uk/metadata/DESIRE/overview/>

13 URL <http://www.ukoln.ac.uk/metadata/DESIRE/quality/>

- A list of quality selection criteria: a reference tool for Internet subject gate ways (interactive web version)¹⁴
- The role of classification schemes in Internet resource description and discovery (Deliverable D3.2 Part 3 including the Peer reviews, 1997-02-28)¹⁵.

Ein fortgeschrittenes Konzept verteilter Specialized Subject Gateways ist in Großbritannien im Rahmen der *Elib-Projekte*¹⁶ im Aufbau begriffen. Derzeit gibt es dort 10 Subject Gateways, welche im wesentlichen auf der technischen Lösung der *ROADS-Software* basieren. Zur Zeit sind dies:

- ADAM: Art, Design, Architecture & Media Information Gateway
- Biz/ed: Business Education on the Internet
- CAIN: Conflict Archive on the INternet
- CATRIONA II
- EEVL: Edinburgh Engineering Virtual Library
- IHR-Info (Institute of Historical Research)
- OMNI: Organising Medical Networked Information
- ROADS: Resource Organisation and Discovery in Subject-based services
- RUDI: Resources for Urban Design Information
- SOSIG: Social Science Information Gateway.

Darüber hinaus nutzen weitere Projekte außerhalb von Elib die ROADS-Software, wie zum Beispiel ALEX (Catalogue of Electronic Texts) oder EELS (Electronic Engineering Library Sweden)¹⁷.

Eine wesentliche konzeptionelle Komponente dieser Projekte ist intellektuelle Erfassung von Internet-Ressourcen mit verbaler wie fachsystematischer Erschließung. Diese Prämisse gilt partiell auch für das mit dem amerikanischen Projekt *Scout-Report* entstandenen, freilich universalen Subject Gateway (*Signpost*). Die Grundidee des Scout-Reports ist es, regelmäßig über neu entstandene, fachwissenschaftlich relevante Ressourcen im Internet zu berichten. Damit ergab sich automatisch, daß im Laufe der Zeit eine umfassendere Sammlung von Daten über solche Ressourcen entstand, welche dann auch dem direkten Zugriff des Nutzers geöffnet wurden.¹⁸

14 URL <http://www.sosig.ac.uk/desire/qindex.html>

15 URL <http://www.ub2.lu.se/desire/radar/reports/D3.2.3/>

16 Vgl. eLib. Electronic Libraries Programme (URL <http://www.ukoln.ac.uk/services/elib/>)

17 Vgl. dazu eine Übersicht existierender Subject Gateways auf der Basis der ROADS-Software: (URL <http://www.ilrt.bris.ac.uk/roads/who/>)

18 Vgl. zum Scout-Report: The Internet Scout Project. Computer Science Department, University of Wisconsin (URL <http://scout.cs.wisc.edu/scout/report/>) sowie Scout Report Signpost (URL <http://www.signpost.org/signpost/index.html>)

Eine andere technische Lösung eines spezialisierten Subject Gateway bietet *Argos: Limited Area Search of the Ancient and Medieval Internet*, welches sich selbst als „... the first peer-reviewed, limited area search engine (LASE) on the World-Wide Web“ bezeichnet.¹⁹ Argos stellt ein hybrides System intellektueller und automatisierter Erschließung dar, da auf der Basis intellektuell ausgewählter Sites zu deren Erschließung eine Suchmaschine eingesetzt wird und dem Nutzer ein Suchinterface angeboten wird, welches die Recherche über dort indizierte Suchtermini erlaubt. Technisch ähnliche Ansätze verfolgt das Elib-Projekt der UB Osnabrück, wenn auch hier nicht der Aufbau eines Subject Gateway bezweckt ist, sondern mit denselben Methoden eine elektronische Zeitschriftenbibliothek aufgebaut wird.²⁰ Das Projekt GERHARD des BIS Oldenburg ist insofern noch vergleichbar, als hier eine universale Suchmaschine für deutsche Internet-Ressourcen über die Universalklassifikation (UDC) automatisiert sachlich erschlossen wird.²¹

Generell machen diese Beispiele verwandter Systeme und Konzepte deutlich, daß allenthalben Bibliotheken oder andere Einrichtungen daran arbeiten, Subject Gateways oder Fachinformationssysteme zur Erschließung von Internet-Ressourcen aufzubauen. Gerade die Elib-Projekte lassen erkennen, daß dies in der ersten Phase auf nationaler Ebene erfolgt, wobei sich aufgrund der globalen Zugriffsmöglichkeiten im Internet eine transnationale Zusammenarbeit und Verlinkung fast automatisch ergibt. Das Beispiel von Argos zeigt zugleich, daß durchaus unterschiedliche technische und inhaltliche Konzeptionen für solche Subject Gateways möglich sind. Hierauf soll nun im folgenden 2. Kapitel eingegangen werden.

19 URL <http://argos.evansville.edu/>

20 URL <http://elib.uni-osnabrueck.de/>

21 German Harvest Automated Retrieval and Directory (URL <http://gerhard.bis.uni-oldenburg.de>)

2 Erschließungsprobleme von Internet-Ressourcen

Von Wilfried Enderle, Thomas Fischer und Karsten Hülsemann

2.1 Das Internet als wissenschaftliche Informationsquelle

Das Internet ist unter anderem aus dem Versuch wissenschaftlicher Institutionen hervorgegangen, eine intensivere und direktere Kommunikation zu ermöglichen.²² Zusammen mit der umfassenden Vernetzung der Computer wurde das Internet zum weltumspannenden Informationsmedium:

- Es ist allgemein zugänglich, vom Büro und von zu Hause, aus Deutschland und der ganzen Welt.
- Das Internet ist unabhängig vom benutzten Rechner, ob UNIX, WINDOWS, Macintosh, Amiga oder Cray: Fast alle Rechnertypen können auf das Internet zugreifen, und für fast alle Programme auf diesen Rechnern stellen sich diese Dokumente auf die gleiche Weise dar.
- Das Internet ist komfortabel: Dateien können gesucht und gefundene Dokumente auf dem eigenen Rechner abgelegt und weiter bearbeitet werden.

E-Mail und Diskussionsgruppen im USENET hatten sich schon fest etabliert, als 1991 das World-Wide Web (WWW) von Tim Berners-Lee am European Laboratory for Particle Physics in Genf (CERN) entwickelt wurde.²³

Mit dem WWW wurde das Internet *multimedialfähig*: nicht nur Texte, auch Bilder, Töne, Animationen und Videosequenzen können übertragen werden. Damit begann eine dramatische Entwicklung: Die Zahl der „Webserver“ nahm und nimmt exponentiell zu, und das Internet dringt mit immer neuen Inhalten in immer weitere Bereiche der Öffentlichkeit vor (vgl. Abb. 2.1).

Auch der wissenschaftliche Bereich des Internet ist an diesem Wachstum stark beteiligt: Nicht nur ist mittlerweile (fast²⁴?) jede deutsche Universität im Netz präsent, sondern immer mehr Fakultäten, Fachbereiche und Institute stellen sich auf eigenen Servern dar und bieten zunehmend nicht nur organisatorische Informationen von Adressen bis Studienordnungen, sondern auch inhaltliche: Forschungsprojekte stellen sich vor, Artikel liegen in Auszügen oder vollständig bereit. Zusätzlich sind Forschungsinstitute und Fachgesellschaften präsent, Zeitschriften stellen Inhaltsverzeichnisse oder Artikel zur Verfügung, neue elektronische Zeitschriften werden gegründet usw.

22 Zur Geschichte des Internet vergleiche z.B. den Geschichtsbereich der Internet Society (URL <http://info.isoc.org/internet/history/index.shtml>).

23 Eine kurze Beschreibung der WWW-Entwicklung am CERN gibt Ben Segal (URL <http://wwwcn.cern.ch/pdp/ns/ben/TCPHIST.html>).

24 Daß nicht klar ist, wie sich herausfinden läßt, ob *alle* Universitäten im Internet präsent sind, zeigt einmal mehr die strukturellen Defizite des Internet.

Millionen WWW-Server

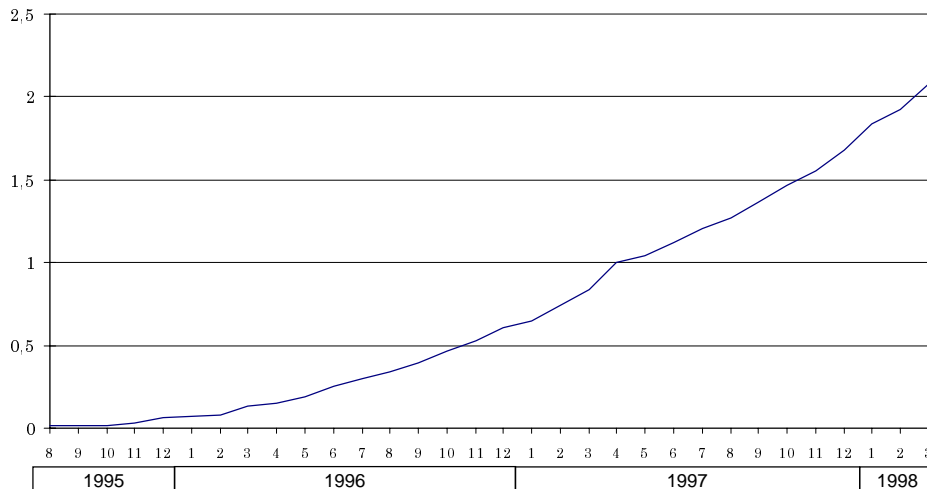


Abb. 2.1: Entwicklung der Zahl der WWW-Server (Quelle: Daten des Netcraft Web Server Survey, URL: <http://www.netcraft.com/survey/>)

Natürlich besteht nicht der Hauptteil, nicht einmal die Überzahl der Internetausdrucksquellen aus wissenschaftlichen Informationen. Aber sie bilden einen relevanten Teil des Angebots. Einen Eindruck über die Verteilung der Inhalte des WWW vermitteln die Einträge in den 14 Hauptkategorien von Yahoo (vgl. Abb. 2.2).

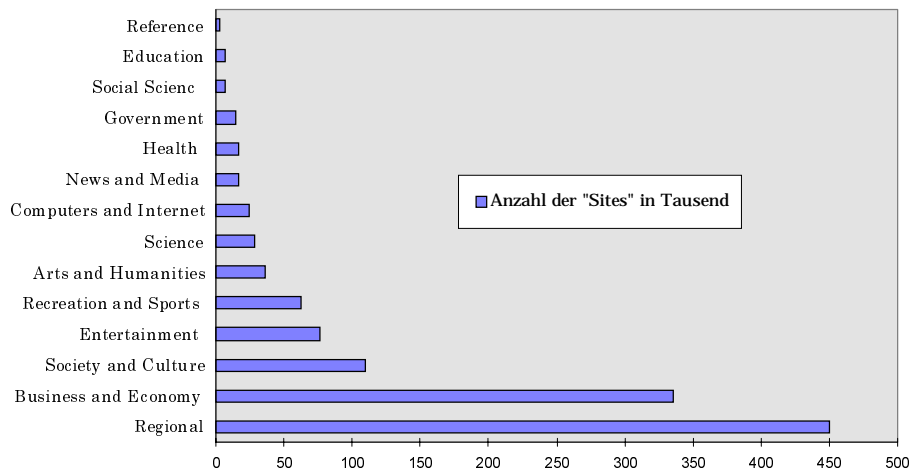


Abb. 2.2: Themenbereiche im WWW (Quelle: Eigene Berechnungen auf der Basis der Yahoo-Kategorien am 16. März 1998, URL: <http://www.yahoo.com>)

2.2 Suchen im Internet: Hoffnungslos?

Die Entwicklung des Internet insgesamt zu überschauen erscheint bei dieser Größenordnung unmöglich. Die Anzahl der WWW-Server überstieg im März 1996 die 100.000 und im April 1997 eine Million; im März 1998 wurden über zwei Millionen Server gezählt. Wie viele Seiten oder Dokumente auf diesen Servern liegen, kann man nur schätzen – ihre Zahl liegt aber sicher im Milliardenbereich. Von diesen sind wiederum vielleicht 10% wissenschaftlich relevant. So liefert eine Anfrage bei einer der klassischen Suchmaschinen, wie z.B. AltaVista (URL <http://altavista.digital.com/cgi-bin/query?pg=q&text=yes>) typischerweise Tausende von Einträgen, in denen sich die wenigen relevanten oft erfolgreich verstecken. Die Situation wird zunehmend als unbefriedigend empfunden, und es gibt verschiedene Vorschläge, wie ein Weg durch das Informationsdickicht gebahnt werden könnte. Im wesentlichen geht es um eine Annäherung aus drei Richtungen: Auf der einen Seite sollen die Suchmaschinen besser ausgestattet werden, so daß sie in stärkerem Maße in der Lage sind, aus einem Dokument relevante Inhalte und nicht nur einzelne Wörter zu entnehmen (Kap. 2.2.1). Auf der anderen Seite sollen die Dokumente selbst die Suche unterstützen und mit zusätzlichen Daten die nötigen Informationen liefern (Kap. 2.2.2.). Dazu kommt als dritter Ansatz – speziell für den Wissenschaftsbereich – der Versuch, über sogenannte Subject Gateways eine qualifizierte inhaltliche Auswahl zu treffen. Hier stellt sich die Frage, ob eine derart intellektuelle Auswahl und Erschließung im Vergleich zu den automatisierten Methoden der Suchmaschinen überhaupt sinnvoll und realisierbar ist (Kap. 2.2.3.).

2.2.1 Besser suchen: Automatische Indizierung

Die Übermacht von Zahl und Umfang der Dokumente hat dazu geführt, Hoffnungen auf die automatische Klassifikation der Internetdokumente durch verfeinerte Suchmaschinen zu setzen. Ein interessantes Projekt ist das „GERman Harvest Automated Retrieval and Directory“ (GERHARD) an der Universitätsbibliothek Oldenburg (URL <http://www.gerhard.de/>). Mittels der Harvest-Technologie sammelt GERHARD den Text von einer Million Dokumenten auf deutschen Servern von

- Universitäten, Fachhochschulen, Hochschulen,
- öffentlichen Forschungseinrichtungen (Max-Planck-Institute, Fraunhofer-Institute,...),
- dem deutschen Schul-Netz (schule.de),
- staatlichen Einrichtungen, Parteien auf Bundesebene.

Diese werden durch einen Vergleich mit Wörtern und Textphrasen auf entsprechende Einträge eines Klassifikationssystems (UDK: Universelle Dezimalklassifikation) abgebildet. Gefundene Zuordnungen werden durch verschiedene

statistische Verfahren unter Beachtung der strukturellen Information aus dem Dokument und dem Klassifikationssystem bewertet und nach Relevanz selektiert. Die so gewonnenen Daten werden in einer Datenbank gespeichert und können über Internet abgerufen werden. Die Klassifikationsprobleme sind jedoch gewaltig: Schon einfache Aufgaben wie die Feststellung von Titel oder AutorIn, die (wenn angegeben) einem Menschen keine Schwierigkeiten bereiten, können nur schwer programmiert werden. Die Zuordnung von Wort und Bedeutung ist nicht eindeutig und von dem jeweiligen Kontext abhängig. Die Betreiber haben jedoch (immer noch) die Hoffnung, daß in Zukunft mit dem GERHARD-Projekt ein Großteil der wissenschaftlich relevanten Quellen des deutschsprachigen Raums erfaßt und richtig klassifiziert werden kann.

2.2.2 Mehr Informationen bieten: Metadaten

Wenn es einer Maschine so schwer fällt, aus einem Dokument Sinn zu extrahieren, so müßte sie unterstützt werden. Diese Überlegung steht am Beginn einer umfangreichen Debatte, die unter dem Schlagwort „Metadaten“ geführt wird.²⁵ Die Grundidee ist hierbei, dem Dokument zusätzliche Daten zuzufügen, die Informationen über das Dokument enthalten. Mit den Metadaten werden Wörter und Begriffe in einen Kontext eingeordnet, der es erlaubt, zu unterscheiden, ob ein Wort Teil eines Titels, des Autorennamens oder einer Inhaltsbeschreibung ist. Dann wäre es zum Beispiel kein Problem, Dokumente über Fischfang von Dokumenten mit dem Autor „Fischer“ zu trennen.

Der international verbreitetste Vorschlag dazu ist das „Dublin Core“-Konzept, das zuerst 1995 auf dem Metadaten Workshop am Online Computer Library Center (OCLC) in Dublin, Ohio (USA) vorgestellt wurde. Seitdem ist dieses Konzept in verschiedenen internationalen Konferenzen weiter bearbeitet worden, zuletzt auf dem „6th Dublin Core Metadata Workshop“ in Washington Anfang November 1998²⁶. Die aktuelle Version dieser Metadaten sieht vor, daß in 15 Kategorien Zusatzinformationen zu einem Dokument angeboten werden, diese Kategorien werden teilweise durch Untergruppierungen („Subelements“) oder Qualifikatoren („Scheme“) verfeinert. Die Kategorien können grob in drei Teilbereiche eingeteilt werden, die sich auf den Inhalt, das geistige Eigentum und die Bereitstellung des Dokumentes beziehen (vgl. Tab. 2.1).

Ein Projekt, das konsequent auf Metadaten setzt, ist der „MathN Broker“, ein Regionalindex für mathematische Preprints (URL: <http://www.mathematik.uni-osnabrueck.de/harvest/brokers/MathN/>). Der MathN Broker kompiliert die Indi-

25 Zu fast allen Aspekten der Metadaten-Debatte(n) gibt es Material auf dem Göttinger Metadaten-Server (URL <http://www2.SUB.Uni-Goettingen.de/>).

26 Die neuesten Information zur Entwicklung des Dublin-Core-Standards sind auf dem Server der „Dublin Core Metadata Initiative“ zu finden (URL <http://purl.org/dc/>).

zes, die mit der Harvest-Technologie von verschiedenen mathematischen Institutionen erstellt werden. Hier ist es auch möglich, die Effektivität der Metadaten sozusagen „live“ zu studieren: Wenn man in einer der vorgegebenen Kategorien sucht (z.B. Autor, Schlagwort, Titel), bekommt man nur die Texte, die diese Angaben in den Metadaten enthalten - und diese liefern dann im Allgemeinen auch die anderen Grunddaten in übersichtlicher Form. Sucht man statt dessen im Bereich „Freier Text“, so werden alle Dokumente geliefert, die das entsprechende Wort enthalten. Das sind dann zwar viel mehr, die mitgelieferten Informationen zu den Fundstellen enthalten aber nicht die wesentlichen Metadaten: keinen Titel, kein Autor, keine Klassifikation. Statt dessen sehen die gefundenen Einträge so aus: „rot_293.ps on:kbibmp3.ub.uni-kl.de“. Diese Fundstelle verweist auf ein Dokument „rot_293.ps“ (wegen ps-Endung: wahrscheinlich eine PostScript-Datei) auf dem Server „kbibmp3.ub.uni-kl.de“ (wegen uni-kl vielleicht bei der Universität Kaiserslautern?). Wenn man mehr Informationen haben möchte, muß man das Dokument aufrufen. Man kann es auf seinem Rechner speichern und ausdrucken oder es ansehen, wenn man ein Programm hat, das PostScript-Dateien darstellt. Dieses langwierige Verfahren (die erwähnte Datei hat einen Umfang von 177 Kilobytes) für jede Fundstelle durchzuführen wäre aber zu aufwendig. So stochert man blind im Nebel des Gesamtindex.

Inhalt	Geistiges Eigentum	Bereitstellung
Title	Creator	Date
Subject	Publisher	Type
Description	Contributor	Format
Source	Rights	Identifizier
Language		
Relation		
Coverage		

Tab. 2.1: Basiskategorien der Dublin-Core-Metadaten (Quelle: Stuart L. Weibel, John A. Kunze, Carl Lagoze: Dublin Core Metadata for Simple Resource Discovery, URL <ftp://ftp.ietf.org/internet-drafts/draft-kunze-dc-02.txt>)

Somit zeigt sich, daß die Benutzung von Metadaten sehr effektiv ist, dafür müssen diese aber *vorhanden* sein. Und damit stellt sich die Frage: Wer legt die Kategorien fest, produziert die Metadaten, sorgt für ihre Konsistenz? Diese Fragen sind für das Internet noch nicht beantwortet. Projekte im Umkreis der Math-Bib-Net-Gruppe wie der MathN Broker versuchen, die Verwendung von Metadaten zur Voraussetzung einer Web-Veröffentlichung in ihrem Bereich zu machen. Dazu wird ein Programm („Mathematics Metadata Markup“) zur Verfügung gestellt, das es erlaubt, die Metadaten in ein Formular einzutragen und so den BenutzerInnen die HTML-Programmierung zu ersparen.

Ein anderer Vorschlag sieht vor, die Metadaten nicht von den AutorInnen der Web-Dokumente selbst, sondern von „vertrauenswürdigen Dritten“ erstellen zu lassen, ähnlich wie auch die Katalogisierung in einer Bibliothek durch spezielle Kräfte mit entsprechender Ausbildung erfolgt. Damit wäre gesichert, daß alle Metadaten einem einheitlichen Schema folgen, eine einheitliche Syntax hätten und somit bei Entwicklung entsprechender Suchmaschinen die Internet-Seiten optimal auffindbar wären.

2.2.3 Automatisierte versus intellektuelle Erschließung

Die Insuffizienz von Suchmaschinen für die wissenschaftliche Recherche und die – zumindest derzeit – noch deutlich zutage tretende Unmöglichkeit vollautomatischer inhaltlicher Klassifizierung ebenso wie die für die meisten elektronischen Dokumente und Webserver (noch) fehlenden Metadaten haben zu der Überlegung geführt, daß die Bibliothekare selbst eine Rolle bei der Erschließung des Internet spielen sollten. Relativ rasch mit der Etablierung des Web begann mithin unter Bibliothekaren und Fachdokumentaren eine Diskussion über die „Katalogisierung des Internet“²⁷.

Die Diskussion wurde dabei einerseits stark von dem scheinbaren Problem der unübersehbaren Fülle und des exponentiellen Wachstums des Netzes und andererseits von den automatisierungstechnischen Möglichkeiten der digitalen Medien geprägt. Dem gegenüber stand immer die konkrete Erfahrung, daß beim derzeitigen Stand der Struktur des Internet und den fehlenden Metadaten eine effiziente Recherche im Netz für Fachwissenschaftler nicht möglich war. Das Internet war und ist, wie ein amerikanischer Historiker formulierte „an electronic Library of Congress without a cataloging system“²⁸. Damit war und ist freilich das Surfen im Internet für einen Wissenschaftler, der systematisch Informationen zu sammeln sucht, wenig hilfreich. „Simply ‚surfing the Web‘ can result in wonderful discoveries; however, it takes far too much time to be a useful research method.“²⁹ Angesichts dieses Standes stellte sich die Frage, wie eine für fachwissenschaftliche Bedürfnisse sinnvolle Erschließung des Netzes

27 Vgl. nur in Auswahl: Norman Oder: Cataloging the Net: Can We Do It? In: *Library Journal*, Bd. 123(16) (October 1, 1998), S. 47-51; Erik Jul: Cataloging Internet Resources: An Assessment and Prospectus. In: *The serials librarian*, Bd. 34 (1998) S. 91-104; Kristin H. Gerhard: Cataloging Internet Resources Practical Issues and Concerns. In: *The serials librarian*, Bd. 32 (1997) S. 123-138; Sally E. Strutt: Cataloguing and the Internet: Considerations at the British Library. In: *Library review*, Bd. 46 (1997) S. 490-499

28 Ryan Johnson: Historical Research On-Line. A New Ball Game. In: Dennis A. Trinkle (Hrsg.): *Writing, Teaching, and Researching History in the Electronic Age. Historians and Computers*, Armonk – London 1998, S. 184

29 Ebenda, S. 185

durchgeführt werden kann. Da eine Erschließung für solche hochqualifizierten Bedürfnisse eine ausreichende Qualität der Erschließung voraussetzt, stellte sich für die Entwicklung fachlicher Internetkataloge und Suchmaschinen wiederum die Frage, in welcher Form eine intellektuelle Erschließung hier sinnvoll praktiziert werden könne.

In der Diskussion um die Alternative zwischen automatisierter und intellektueller Erschließung war freilich zumeist die Analyse des Problemfeldes oft etwas zu kurz gekommen. Zu pauschal wurden meist die enormen Zahlen existierender Webseiten im Internet als Beleg dafür angeführt, daß diese Informationsmenge intellektuell nicht mehr zu erschließen sei. Dabei wird indes meistens der Unterschied zwischen Information und Publikation übersehen, wobei man natürlich auch zugestehen muß, daß gerade die klare Definition dieses Unterschiedes im Internet schwierig geworden ist. Zum Verständnis des Problems dürfte es daher hilfreich sein, sich vor Augen zu halten, wo die Unterschiede zur derzeitigen Praxis bei gedruckten Publikationen liegen.

Fachdokumentationseinrichtungen und Bibliotheken haben in den letzten Jahrzehnten aus der gesamten Zahl der jährlich erscheinenden Menge wissenschaftlicher Publikationen, die mittlerweile allein in Deutschland pro Jahr ungefähr 80.000 Einheiten beträgt, Datenbanken und Bibliotheksbestände aufgebaut, die einen jährlichen Zuwachs von mehreren zehntausend Datensätzen bzw. Bänden haben. Eine mittlere Universitätsbibliothek erwirbt in Deutschland pro Jahr 25.000 bis 40.000 Bände, an der SUB Göttingen sind es durchschnittlich 60.000³⁰. Diese Zahlen dokumentieren, daß Bibliotheken durchaus mit großen Informationsmengen umgehen können. Damit stellt sich die Frage, warum dies für den Bereich digitaler Informationsressourcen im Internet nicht möglich sein sollte. Hat sich dort die Zahl wissenschaftlich relevanter Informationsressourcen gegenüber den gedruckten Publikationen so enorm erhöht? Sicherlich ist es mit der Möglichkeit elektronischer Publikation leichter geworden, wissenschaftliche Ergebnisse zu publizieren, was im Endergebnis zu einer Zunahme wissenschaftlicher Publikationen führen wird. Vor allem aber dürfte es zu einer stärkeren Fragmentierung solcher Publikationen führen, das heißt, daß, wie gerade die Fächer Mathematik und Physik belegen, zunehmend Preprints und Working Papers auf diesem Wege veröffentlicht werden, wobei das Internet hierfür nochmals eine technische Vereinfachung bei der Publikation darstellt. Trotz dieser Zunahme wissenschaftlicher Publikationen aufgrund der in technischer Hinsicht leichteren Publikationsmöglichkeiten wird die Zahl wissenschaftlicher Institutionen und der dort tätigen Wissenschaftler auch in Zukunft nicht exponentiell zunehmen. Deren Zahl wird für das jeweilige Fach immer in einem einigermaßen überschaubaren Rahmen bleiben.

30 Vgl. Deutsche Bibliotheksstatistik: DBS, Teil B.: Wissenschaftliche Bibliotheken 1997, hrsg. vom Deutschen Bibliotheksinstitut, Berlin 1998

Das eigentliche Problem, was die Erschließung wissenschaftlicher Internet-Ressourcen angeht, ist also nicht in erster Linie ein Problem der Quantität, auch wenn dies natürlich eine gewisse Rolle spielt, sondern ein Problem der Organisation, bzw. genauer der fehlenden organisatorischen Infrastruktur des Netzes. Daß Dokumentare und Bibliothekare große Mengen gedruckter Daten verarbeiten können, liegt wesentlich daran, daß es eingeführte und effiziente Strukturen gibt, die es erlauben, sich rasch und systematisch über Neuerscheinungen zu informieren. Dazu gehören vor allem die Nationalbibliographien, aber auch Rezensionsorgane und Buchhandelsinformationen. Gäbe es vergleichbare Strukturen im Internet, könnte es durchaus möglich sein, allein mit Hilfe intellektueller Erschließung die Informationsressourcen zu erfassen, die die Qualität wissenschaftlicher Publikationen besitzen.

Dies bedeutet aber nicht zugleich, daß dies auch sinnvoll ist. Schließlich sollte man nicht die Strukturen des Printbereichs direkt und unmittelbar auf die neuen digitalen Medien übertragen, da man damit die diesem Medium inhärenten neuen technischen Möglichkeiten nicht nutzen und ausschöpfen würde. Man muß sich nur darüber im klaren sein, daß es unter bestimmten Umständen möglich ist, die Informationsressourcen des Internet intellektuell zu überschauen und daß gerade die Qualifizierung und Auswahl wissenschaftlicher Ressourcen immer auch eine intellektuelle Bewertung und Arbeit in irgendeiner Form erfordert. Sei es, daß man solche Ressourcen direkt bewertet, sei es, daß man das fachliche Wissen in Expertensysteme einbringt, welche dann die eigentliche Erschließung der Ressourcen durchführen.

Der derzeitige Stand der Entwicklung der Specialized Subject Gateways im Internet zeigt denn auch, daß derzeit hybride Systeme entwickelt werden, die versuchen, intellektuelle Auswahl mit automatisierten Verfahren zu verknüpfen³¹. Dies bedeutet, daß fachlich relevante Websites intellektuell ausgewählt werden, ggf. sogar eine Analyse bis auf Verzeichnisebene erfolgt, die Erschließung einzelner Seiten dann aber auf der Grundlage dieser vorgegebenen Adressen durch Robots durchgeführt wird. Diese Erschließung selbst kann dann analog zu den universalen Suchmaschinen nur Stichwortindizes umfassen, also nur die Möglichkeit einer verbalen Recherche bieten. Bei einem zweiten Modell werden diese Indizes zusätzlich mit Hilfe von Expertensystemen automatisch durch Begriffsabgleich um Notationen erweitert, so daß auch ein Einstieg über einen systematischen Katalog möglich ist. Im ersten Fall muß die intellektuelle Auswahl und Analyse der Sites gründlicher und präziser sein, damit eine rein verbale Suchmaschine ausreichend qualifizierte Ergebnisse liefern kann; im zweiten Fall kann die Auswahl der Sites gröber sein, wobei beim derzeitigen Stand der Entwicklung deutlich ist, daß solche Systeme, was ihre E r-

31 Vgl. die Ausführungen oben in Kap. 1.3

gebnisse angeht, bei weitem nicht mit der Qualität intellektueller Erschließung konkurrieren können.

Die zukünftige Bedeutung und Qualität solcher Hybridsysteme könnte auch noch durch die zunehmende Verbreitung des Konzeptes standardisierter Metadaten gesteigert werden. In diesem Fall liegt der Schwerpunkt wieder stärker auf der intellektuellen Erschließung, nur daß diese von Dokumentaren und Bibliothekaren hin zum Autor selbst verlagert wird. Die von ihm gelieferten Metadaten könnten dann automatisiert ausgewertet werden, das heißt, daß die Produktion von Fachdatenbanken und Bibliothekskatalogen in diesem Bereich weitgehend automatisiert würde. Ein Problem bleibt dabei die Qualität der sachlichen Erschließung, da ein Fachautor diese nicht in dem Maße gewährleisten kann wie ein Fachdokumentar oder wissenschaftlicher Bibliothekar.

Das SSG-FI-System ist so konzipiert, daß es eine Ausgangsbasis bilden kann für den Einsatz automatisierter Methoden. Da nur Informationsressourcen, thematische Sites, nicht aber einzelne Dokumente erschlossen werden, können diese Sites die Datenbasis bilden, die dann in Zukunft auf Dokumentenebene detailliert durch entsprechende Programme erschlossen wird. Dieses Vorgehen wird derzeit im Rahmen des EULER-Projektes der Europäischen Kommission an Hand der Daten des MathGuide getestet.

Da, wie noch detaillierter gezeigt wird, das SSG-FI-System auf standardisierten Metadaten aufbaut, gibt es noch eine zweite Möglichkeit der Automatisierung. Sowie im Netz entsprechende Fachressourcen bereits vom Produzenten über Metadaten erschlossen werden, können diese direkt in das System übernommen werden, wobei im Einzelfall voraussichtlich entsprechende intellektuelle Ergänzungen, die den spezifischen Bedürfnissen des SSG-FI-Systems Rechnung tragen, gemacht werden müssen. Dazu kommt als drittes, daß auf der Basis eines mit Metadaten arbeitenden Fachinformationssystems Kooperationsmöglichkeiten mit anderen Subject Gateways gegeben sind³².

Unabhängig davon, inwieweit das SSG-FI-System Basis für weitergehende automatisierte Erschließungen sein kann, ist der Grundbestand intellektuell ausgewählt und erschlossen. Damit stellt sich die Frage, welche Auswahl- und Bewertungskriterien für die Erschließung von Internet-Ressourcen angewandt werden können und sollen.

2.3 Das Problem der Auswahl- und Bewertungskriterien

Bibliothekaren wie Wissenschaftlern ist zunehmend bewußt geworden, daß die Informationsmöglichkeiten des Internet nur dann sinnvoll genutzt werden können, wenn fachliche Suchmaschinen oder Kataloge vorhanden sind, welche nur

32 Vgl. dazu Kap. 8

die wissenschaftlich relevanten, also ausgewählte und positiv bewertete Ressourcen erfassen. Dies zeigt die zunehmende Zahl einschlägiger Publikationen³³ wie auch der Aufbau entsprechender fachlicher Suchmaschinen und Kataloge³⁴. Diese Erkenntnis entbehrt freilich nicht einer gewissen Banalität, ist es doch im Grunde für Bibliothekare wie Fachdokumentare eine alltägliche Praxis, die von ihnen gekauften oder erschlossenen Ressourcen nach spezifischen Gesichtspunkten auszuwählen und damit auch in ihrer wissenschaftlichen Qualität und Nutzbarkeit zu bewerten. Bibliothekare haben immer schon sehr stark aus der großen Zahl gedruckter Publikationen nach bestimmten Kriterien ausgewählt, wobei nicht nur die jeweilige Qualität eines Buches von Belang war, sondern ebenso die vorhandene Bestandsstruktur und der Sammelauftrag der Bibliothek.

Im Unterschied zur Auswahl gedruckter Literatur birgt die Erstellung eines Katalogs von Internet-Ressourcen zunächst nicht dasselbe finanzielle Risiko, weil (zumindest im Konzept von SSG-FI und ähnlichen Projekten) mit der Erstellung eines Katalogs und eines sog. *pointer site* bzw. eines *Subject Gateway* zunächst keine Kosten für die Bereitstellung der Ressource *an sich* verbunden sind. Es bleibt somit das Problem der fachlichen Qualität. Dieses fachliche Auswahl- und Bewertungsproblem stellt sich bei Internet-Ressourcen freilich in ausgeprägterer Form. Nicht nur fehlen bei Informationsquellen im Internet zumeist die bibliographischen Angaben (Autor, Herausgeber, Verleger, Erscheinungsjahr/-datum und -ort), die für die meisten gedruckten Ressourcen typischerweise relativ einfach zu ermitteln sind. Es fehlt zudem an der gesamten Infrastruktur, die eine geleitete Auswahl ermöglicht. Wo der „klassische“ Bibliothekar sich auf Nationalbibliographien, ein etabliertes Verlagswesen und die rigorose Qualitätskontrolle in wissenschaftlichen Zeitschriften durch ein „peer review“ (d.h. ein Gutachtersystem) stützen kann, ist der „Internet Librarian“ darauf angewiesen, all diese Qualitätskontrollen selbst durchzuführen.

Dazu kommt, daß eben nicht nur eine Infrastruktur fehlt, welche bereits dem Bibliothekar eine Vorauswahl nach bestimmten Kriterien aus der Vielzahl wissenschaftlicher Publikationen anbietet, sondern daß die elektronischen Ressourcen auch teilweise nach anderen Kriterien zu beurteilen sind als gedruckte

33 Vgl. zum Stand der Literatur George McMurdo: Evaluating Web information and design. In: *Journal of Information Science*, Bd. 24, 1998, S. 192f. Zum deutschen Raum siehe nur als Beispiel: Margarete Payer, „Wie kann man die Qualität von Internetressourcen für den wissenschaftlichen Bereich beurteilen? Hinterfragung ausgewählter Vorschläge“; Vortrag am 11. Juni 1997, HBI Stuttgart. Fassung vom 9. Juni 1997, URL: <http://machno.hbi-stuttgart.de/~payer/infoq.html>

34 Vgl. Anthony F. Beavers: Evaluating Search Engine Models for Scholarly Purposes. A Report from the Internet Applications Laboratory. In: *D-Lib Magazine*, December 1998, URL: <http://www.dlib.org/dlib/december98/12beavers.htm>

Bücher und Zeitschriften. Dabei liegt es natürlich nahe, bei den Auswahl- und Bewertungskriterien mit den vom gedruckten Medium her bekannten Maßstäben anzusetzen. Es steht auch außer Frage, daß dort, wo das digitale Medium letztlich nur ein Äquivalent eines gedruckten Buches ist, also zum Beispiel bei digitalen Editionen oder Online-Bibliographien, im Grunde auch dieselben Kriterien angewandt werden können. Schwieriger wird es dann, wenn elektronische Ressourcen neue Formen annehmen, wie zum Beispiel bei virtuellen Bibliotheken oder fachlichen Informationsservern, welche ganz unterschiedliche Informationsformen vereinen.

Zunächst aber ist das Kernkriterium für eine Auswahl einer zu erschließenden Internet-Ressource, der Zielsetzung der SSG-FI-Guides entsprechend, deren wissenschaftliche Nutzbarkeit. Dieses zentrale Qualitätskriterium wird so auch von anderen Projekten definiert: „A high quality Internet resource is one that meets the information needs of the user“ (Definition von SOSIG/DESIRE). Für ein fachwissenschaftliches Informationssystem bedeutet dies, daß die erschlossenen Ressourcen Informationen enthalten müssen, die *für wissenschaftliche Forschung von Bedeutung* sind. Was dies im einzelnen bedeutet, muß dann jeweils für das einzelne Fach, also den einzelnen Guide, definiert werden, da die fachwissenschaftlichen Bedürfnisse sich doch sehr stark unterscheiden.³⁵ Es gibt freilich ein allgemeines, formales Raster, nach dem unabhängig vom Fach die Internet-Ressourcen beschrieben werden können. Dieses Raster umfaßt folgende Aspekte:

Content	Clarity	Backlinks
Index	Links	Level

Der Ablauf bei der Auswahl einer Ressource ist somit folgender:

1. Aufgrund fachspezifischer Kriterien wird entschieden, daß eine Ressource für wissenschaftliche Zwecke nutzbar ist. Mit der Selektion von Ressourcen ist eine erste, freilich *implizite* Form von Bewertung gegeben. Das heißt, die Aufnahme einer Ressource in einen Subject Gateway unterstellt im Regelfall, daß es sich dabei um eine qualitativ hochwertige Ressource handelt. „Im Regelfall“ deshalb, weil durchaus im Rahmen einzelner Guides unter fachspezifischen Erwägungen vieles dafür sprechen kann, bestimmte Bereiche oder Arten von Ressourcen möglichst *in ihrer Gesamtheit* zu erfassen. Eine solche Vereinbarung wurde beispielsweise für den Anglo-American Literature Guide getroffen. Hier wird versucht, alle Seminare bzw. Universitätsinstitute im Bereich der Anglistik und Amerikanistik im deutschsprachigen Raum aufzunehmen, um Nutzern eine komplette Übersicht zu geben³⁶. Der

35 Hierauf wird im einzelnen in Teil 2 eingegangen

36 Vgl. ebenda

Umkehrschluß gilt allerdings nicht zwingend: Wenn eine Ressource nicht in einem (oder mehreren) SSG-FI katalogisiert und beschrieben ist, so spricht das nicht notwendigerweise für einen Mangel an Qualität. Es kann vielmehr auch sein, daß das für diesen Guide zuständige Team noch nicht auf diese Ressource aufmerksam geworden ist bzw. ein entsprechender Hinweis durch Fachwissenschaftler oder andere (noch) nicht erfolgt ist. Angesichts der Veränderlichkeit des Informationsraumes Internet ist auch für die Zukunft nicht davon auszugehen, daß ein Fehlen im Katalog von SSG-FI ein eindeutiges Zeichen für schlechte Informationsqualität ist.

2. Nach der Entscheidung, eine Ressource zu erfassen, wird diese zunächst nach formalen und fachlichen Kriterien typisiert bzw. - in klassischer Formulierung - inhaltlich bzw. sachlich erschlossen. Also es wird festgelegt, ob es sich um ein Quellenarchiv oder eine fachliche Linkliste etc. handelt³⁷; dazu kommt die Einordnung in ein bestimmtes Wissensgebiet (fachliche Klassifizierung)³⁸. Vor dem Hintergrund dieser inhaltlichen Erschließung werden dann die oben genannten Aspekte geprüft und die Informationsressource erhält für jedes bewertbare Kriterium (linke und mittlere Spalte) einen bis drei Sterne.

Insgesamt geht es dabei nicht, wie betont werden muß, um eine gleichsam rezensierende Bewertung. Die Tatsache, daß ein Site aufgenommen wurde, dokumentiert bereits, daß er für die wissenschaftliche Forschung relevant und nutzbar ist. Die Analyse nach dem beschriebenen Raster dient vor allem dazu, dem Nutzer vorab einen Eindruck davon zu vermitteln, was ihm dieser Site konkret bietet, ob er also zahlreiche weiterführende Links enthält, eine lokale Suchmaschine und ähnliches. Daß darin eine evaluierende Komponente enthalten ist, ist unstrittig. Diese dient aber nur dazu, verschiedene Sites, die grundsätzlich positiv bewertet wurden, nach möglichst formalisierten Kriterien näher zu charakterisieren.³⁹

2.4 Infrastrukturelle Defizite des Internet

Umfang und Qualität von Special Subject Gateways im Internet hängen nicht nur von der Qualität und Konsistenz der Subject Gateways selber ab, sondern ebenso von dem Grad der infrastrukturellen Organisation des Internet. Dessen Defizite können natürlich nicht mit einzelnen kleinen Projekten zum Aufbau fachlicher Informationssysteme gelöst werden, man muß aber diese Defizite benennen, damit erkennbar wird, wo derzeit die Grenzen dieser Informationssysteme notgedrungen liegen (müssen). Im folgenden wird auch nicht das g e-

37 Zur formalen Erschließung siehe Kap. 5.3

38 Zur fachlichen Erschließung siehe Kap. 5.4

39 Siehe hierzu en détail Kap. 3.3

samte Feld der Probleme angesprochen, welche das Internet und die digitalen Medien mit sich bringen, sondern nur kurz diejenigen wesentlichen Umstände, die für den Aufbau und Betrieb eines Fachinformationssystems im Internet von Belang sind. Im wesentlichen sind dies derzeit drei Problembereiche: Die fehlende organisatorische Infrastruktur, das Problem der Validität der Netzadressen und die noch fehlenden Standards für Beschreibung und Gestaltung von Informationsressourcen im Internet.

Das Problem der fehlenden organisatorischen Infrastruktur im Hinblick auf wissenschaftlich relevante Fachinformationen wurde bereits angesprochen. Eine der Hauptschwierigkeiten beim Versuch, einen Subject Gateway aktuell zu halten, ist derzeit, daß es keine Institutionen gibt, welche umfassend über neu entstandene Informationsressourcen unterrichten, sei es auf nationaler oder fachlicher Ebene⁴⁰. Derzeit muß man regelmäßig andere Subject Gateways und fachliche Linklisten durchsehen, was dazu führt, daß man immer wieder auf dieselben Ressourcen stößt; ferner ist es notwendig, in einer Vielzahl fachlicher Diskussionslisten subskribiert zu sein oder deren Archive regelmäßig durchzusehen, da dort in der Regel neue Ressourcen angezeigt werden.

Notwendig wäre also, daß zunächst auf nationaler Ebene, analog zu dem Modell der CIP-Katalogisate einer Nationalbibliographie, eine umfassende, aber grob fachlich gegliederte Information über neue Informationssites erfolgt oder auch über wesentliche Veränderungen auf diesen Sites. Dazu wäre es von Bedeutung, wenn sich zentrale fachliche Dienste etablieren würden, bei denen die Wissenschaftler von ihnen selbst erstellte Ressourcen anmelden können. Für die letztere Funktion können auch SSG-Bibliotheken im Kontext einer virtuellen Fachbibliothek Aufgaben übernehmen. Gerade das SSG-FI-System sieht auch die Anmeldung neuer Ressourcen durch Wissenschaftler selbst vor.

Der zweite Problembereich der sich oft verändernden URLs (Uniform Resource Locator) ist allgemein bekannt und bedarf einer Lösung auf der Ebene der für die Standards des Netzes zuständigen Gremien. Mit dem Modell der URNs (Uniform Resource Number) gibt es auch einen Lösungsvorschlag, ebenso wie mit den DOIs (Document Object Identifiers). Entscheidend wird sein, daß sich zukünftig ein international anerkanntes Modell durchsetzt. Wenn dies der Fall wäre, würde die laufende Pflege eines Subject Gateway beträchtlich vereinfacht.

Für die raschere Erschließung von Informationsressourcen im Internet wäre es zum dritten notwendig, daß sich einheitliche Standards bei der Gestaltung solcher Internet-Ressourcen durchsetzen. Daß solche Ressourcen durch die Pro-

40 Das Internet Scout Project hat genau dieses Defizit erkannt und versucht dafür eine Lösung zu bieten, der freilich zur Zeit noch die entsprechende institutionelle Absicherung fehlt, welche deren Dauer gewährleisten könnte.

duzenten selbst mit Metadaten versehen werden, die automatisiert ausgewertet werden können, ist dabei nur ein, wenn auch wichtiger Gesichtspunkt. Ebenso notwendig wäre es aber auch, daß sich Layoutstandards etablierten, die rascher als es bislang oft der Fall ist, die wesentlichen Informationen über einen Site erkennen lassen⁴¹. Wenn die inhaltlich verantwortlichen Einrichtungen und Autoren sofort erkennbar wären, wenn darüber hinaus der Site selbst, was Umfang und Inhalt, aber auch die strukturelle Organisation seiner Daten angeht, Informationen auf der ersten Ebene, also der Einstiegsseite enthielte, könnte die intellektuelle Bewertung rascher erfolgen als dies derzeit der Fall ist. Derzeit ist es oft notwendig, über verschiedene Ebenen des Site zu recherchieren, um einen Eindruck von dessen fachlicher Qualität und Organisation zu erhalten.

41 Ein Beispiel für diese Standardisierung ist die MathNet-Seite der Konferenz der mathematischen Fachbereiche, die eine (weitgehend) einheitliche Struktur zum Einstieg in die Server von Mathematik-Fachbereichen anbietet.

3 Die Konzeption des SSG-Fachinformationssystems im Überblick

Von Thomas Fischer, Karsten Hülsemann, Heike Neuroth und Norbert Pfurr

Ausgangspunkt für die Konzeption des Projektes waren Mitte der neunziger Jahre folgende Erkenntnisse:

- Im Internet werden im zunehmenden Maße wissenschaftlich relevante Informationen bereitgestellt.
- Für eine Sondersammelgebietsbibliothek mit dem Auftrag, die wissenschaftlichen Publikationen der Disziplin ihres Sondersammelgebietes vollständig zu sammeln und für den überregionalen Leihverkehr zu erschließen, sind auch diese Informationsressourcen von Bedeutung. Es sollte daher ein Konzept entwickelt werden, das die elektronischen Fachinformationen in das Angebot einer SSG-Bibliothek integriert.

Die Erfahrung zeigt, daß die Fülle des Informationsangebots im Internet beim Einsatz von vollautomatischen Verfahren zum Sammeln und Erschließen elektronischer Fachinformationen (Suchmaschinen) zu Ergebnissen führt, die nicht den Ansprüchen wissenschaftlicher Recherche genügen (vgl. Kap. 2). Hier ist eine Selektion nach den Kriterien wissenschaftliche Qualität und Relevanz wünschenswert. So ergab sich die Notwendigkeit, durch intellektuelle Recherche und Bewertung einen Überblick über die relevanten elektronischen Fachinformationen zu gewinnen, um sie dann dem Nutzer in strukturierter Form anbieten zu können.

Daraus resultierten für das SSG-FI Projekt folgende Leitlinien:

- Ziel des SSG-FI-Projektes ist es, dem Wissenschaftler oder Studenten die Fülle vorhandener wissenschaftlicher Informationsquellen im WWW, aber in Zukunft auch gedruckte oder als CD-ROM vorliegende Auskunftsmittel in übersichtlicher und qualifizierter Form zu erschließen.
- Es sollen im Rahmen des Systems nur sekundäre Fachinformationen erschlossen werden, also Informationen über Informationsressourcen. Für das Online-Angebot sind dies thematische Sites, nicht aber einzelne elektronische Dokumente (vgl. Kap. 3.1).
- Die thematischen Sites werden nach qualitativen Kriterien ausgewählt und bewertend beschrieben. Es werden somit nicht pauschal alle thematisch einschlägigen Sites erfaßt.
- Das Projekt erfordert damit auch die Erarbeitung von Regeln, wie thematische Sites erfaßt, beschrieben und bewertet werden sollen (vgl. Kap. 3.3).
- Das SSG-FI-System soll einen qualifizierten Überblick über vorhandene Informationsressourcen geben und damit auch eine Basis für die ggf. notwendige automatisierte Erschließung elektronischer Primärdokumente schaffen.

- Der Fachinformationsführer stellt eine zusätzliche SSG-Dienstleistung dar, eine Art Bibliographie der Fachinformationsmittel oder Auskunftsmittel, wie sie in der Regel von SSG-Bibliotheken nicht angeboten wurden und ergänzt und komplettiert damit vorhandene SSG-Dienste. In einer virtuellen Fachbibliothek ist der SSG-Fachinformationsführer in einer Ebene über den Suchmaschinen und Online-Katalogen, also Dateneinheiten zur Beschreibung primärer Informationseinheiten (wie z.B. elektronische Texte, Bilder, Software), angesiedelt.
- Das Konzept für die Erfassung von Internet-Fachinformationen oder thematischen Quellen soll nachnutzbar sein für andere Sondersammelgebiete, d.h. es soll konzeptionell so offen sein, daß auch andere Fachgebiete es nutzen können (vgl. Kap. 8). Es soll auch keinerlei technische Einschränkungen in dem Sinne geben, daß bestimmte DV-Systeme eingesetzt werden müssen. Das wesentliche Ziel war die Erarbeitung einer Konzeption, die mit unterschiedlichen technischen Systemen realisiert und damit ohne größeren Aufwand auch von kleineren SSG-Bibliotheken umgesetzt werden kann.

Die Fülle vorhandener, wissenschaftlich relevanter Informationsquellen im WWW, aber auch gedruckt oder als CD-ROM vorliegende Bibliographien und Datenquellen sollen also in übersichtlicher und qualifizierter Form erschlossen werden.

Im SSG-Fachinformationsprojekt (SSG-FI) werden wissenschaftlich relevante Online-Ressourcen (in der Regel thematische Sites, vgl. Kap. 3.1) recherchiert, katalogisiert und beschrieben. Der bibliographischen Beschreibung liegt pro Internet-Ressource ein Satz Metadaten zugrunde (vgl. Kap. 3.2), der den Titel, den Autor, den Verleger etc. erfaßt, eine Beschreibung und Schlagwörter liefert und Auskunft über Zugangsberechtigungen wie z.B. etwaige Kosten liefert. Das Metadatenkonzept basiert auf dem „Dublin Core Metadata Element Set“ und ist um einige SSG-FI spezifische Kategorien erweitert worden. Die Internet-Ressourcen werden von Fachwissenschaftlern und bibliothekarischen Fachreferenten ausgewählt und möglichst nach einem fachübergreifenden und objektiven Maßstab bewertet (vgl. Kap. 3.3).

Der Fachinformationsführer bietet verschiedene Einstiegs- bzw. Suchmöglichkeiten an (vgl. Kap. 4): Ein Nutzer kann zunächst einmal, wie er das zum Beispiel von klassischen gedruckten Bibliographien gewohnt ist, nach thematisch interessanten Informationsquellen in einer Fachklassifikation blättern, dem sogenannten *Subject Catalog*. Ferner gibt es einen *Source Type Catalog*, der die Informationsquellen nach formalen Kriterien auflistet, und für die Suche eine *lokale Suchmaschine*, in die Schlag- oder Stichwörter sowie Fachnotationen etc. direkt eingegeben werden und miteinander verknüpft werden können (mehrdimensionale Suche).

Technisch liegt dem Fachinformationssystem eine Allegro-Datenbank zugrunde, von der aus die Datensätze in das Internetformat HTML umgesetzt werden (vgl. Kap. 6).

Abb. 3.1 gibt einen Überblick über die Struktur des Fachinformationssystems. Auf der linken Abbildungsfläche ist die Betreibersicht schematisch dargestellt: Die Datensätze eines „SSG-FI Guides“ werden in einem Datenbanksystem gesammelt und verwaltet. Für die Beschreibung der Internet-Sites stehen ca. 40 Kategorien zur Verfügung, in denen die Grunddaten zu jeder Informationsquelle und Zusatzinformationen erfasst werden. Diese Metadaten werden aus der Datenbank heraus in das WWW exportiert. Im Hinblick auf die Nutzbarkeit (weite Verbreitung von Allegro im Bibliothekswesen, geringe Lizenzgebühr) wird im SSG-FI Projekt die Allegro V15 Datenbank eingesetzt. Aus Allegro werden die Daten für die Internet-Präsentation in feste HTML-Dokumente übersetzt.

Auf der rechten Abbildungsfläche ist schematisch die Nutzersicht dargestellt, die für jeden Guide identisch ist: Es gibt auf der ersten Hierarchieebene drei Einstiegsmöglichkeiten für die Recherche im Guide (Subject Catalog, Source Type Catalog, Local Search). Auf derselben Ebene finden sich Hinweise auf allgemeine Informationen.

Der Nutzer oder die Nutzerin können nun in dieser Struktur *blättern* („browse“) oder *suchen*. Sie erhalten sie eine kurze Beschreibung der Internet-Ressourcen, jeweils mit der Möglichkeit, ein ausführliches Metadaten-Formular einzusehen oder sich zu der Ressource selbst „durchzuklicken“.

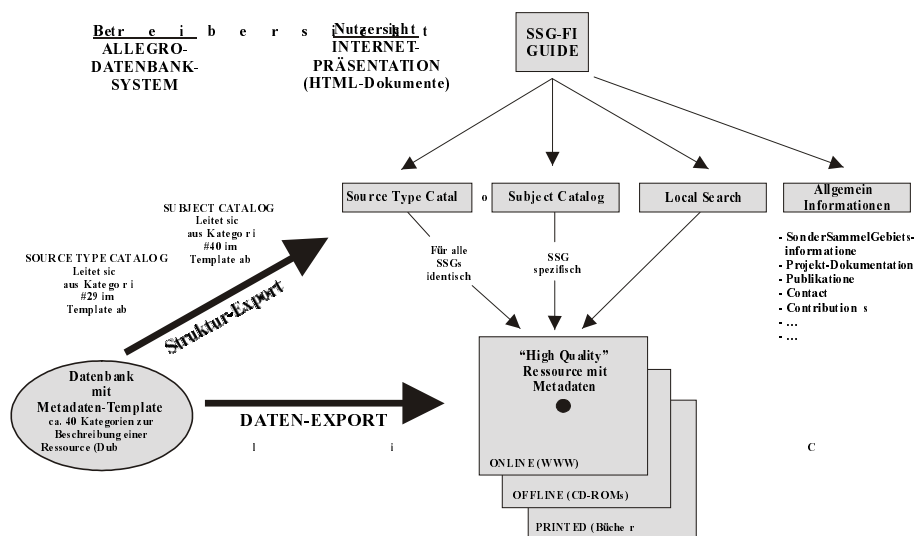


Abb. 3.1: Struktur des SSG-FI-Systems

3.1 Definition der Informationsquellen

Da automatische Verfahren derzeit keine Erfassung des Internet leisten, die wissenschaftlichen Kriterien genügen, muß „das Internet“ intellektuell erfaßt werden (vgl. Kap. 2). Dabei muß im Vorfeld geklärt sein, welche Informationen überhaupt erfaßt werden sollen. Im SSG-FI-Projekt werden nur sekundäre Informationsquellen erfaßt, also keine elektronischen Einzeldokumente; dafür wurde der Begriff des **thematischen Site** eingeführt. Ein thematischer Site ist eine Ansammlung miteinander verbundener Dokumente, die von einer Person oder Institution im Internet veröffentlicht werden und einem fachlichen Thema gewidmet sind.

Dies ist ein nicht immer unproblematischer Begriff, da er

- nicht dem technischen Server entspricht, der der Vermittlung der Dokumente ins Internet dient,
- damit auch schwer abzugrenzen ist, da nicht alles, was mit einem thematischen Site verbunden ist, ihm zuzurechnen ist, und schließlich
- ein thematischer Site einen anderen mit engerem Thema umfassen kann.

Als Beispiel mag der Site einer Fachgesellschaft dienen, der somit ein thematischer Site ist. Dieser Internet-Bereich kann so groß sein, daß für eine übersichtliche Erfassung Untereinheiten (z.B. große Tagungen mit Kurzfassungen vorgesehener Referate, Liste der veröffentlichten Zeitschriften oder große strukturierte Linksammlungen) oder gar deren Teilbereiche (z.B. einzelne Zeitschriften) getrennt voneinander beschrieben werden sollten.

Die Auswahl der zu erfassenden thematischen Sites spielt bei der Menge der Internet-Ressourcen ebenfalls eine Rolle (vgl. Kap. 2.3). Möchte man den SSG-FI Guide übersichtlich und für den Nutzer überschaubar gestalten, so steht Qualität vor Quantität. Bevor entschieden wird, welcher thematische Site in das System aufgenommen wird, muß man sich einen Überblick über möglichst viele der vorhandenen Sites mit dem speziellen Themengebiet verschaffen und von diesen dann die wissenschaftlich relevantesten Sites aufnehmen und beschreiben. Kapitel 7.1 gibt einen Überblick über die Recherchestrategien beim Aufbau der SSG-FI-Guides.

Zusammenfassend kann festgehalten werden, daß in erster Linie nur die wissenschaftlich relevantesten thematischen Sites in das System integriert werden, eventuell wird ein zu umfangreicher Site bei der formalen und/oder thematischen Erfassung in mehrere Unterpunkte untergliedert. Hinweise auf erwähnenswerte Einzeldokumente sollten in der Beschreibung erfolgen.

Das SSG-FI System ist mithin ein Konzept zur Beschreibung wissenschaftlicher Informationen im Internet. Beschrieben werden sollen typischerweise *thematische Sites*, also nicht einzelne Texte oder Quellen, sondern Informations-

vermittler, die auf solche Informationen verweisen oder sie beinhalten. Für jeden dieser Sites wird ein Metadatensatz angelegt.

3.2 Das Metadatenformat

Metadaten sind Daten über Daten, d.h. Metadaten bezeichnen das, was einer Ressource an beschreibenden Zusatzinformationen gegeben wird, um sie besser zu erschließen und dadurch gezielter auffindbar zu machen. Da die sinnvolle automatisierte Katalogisierung des Internet bisher noch nicht gelungen ist, stand die Überlegung am Beginn einer international geführten Debatte, daß mitgelieferte Metadaten zu den Internet-Seiten das erfolgreiche Suchen im Internet optimieren könnten.

Das SSG-FI Projekt benutzt Dublin Core als Grundlage, die um einige SSG-FI-spezifische Kategorien erweitert worden ist. Zur Beschreibung und Klassifikation der Ressourcen wurde eine Liste von etwa 40 Kategorien erarbeitet, die sich wie folgt zusammensetzt:

- von den 15 Dublin Core Kernkategorien werden im SSG-FI-System 13 benutzt, die z.T. weiter untergliedert sind; Contributor und Source entfallen, es kommen hinzu:
- der Formalschlüssel der Ressource,
- das Land, in dem die Ressource technisch bereit gestellt wird,
- statistische Daten zum Umfang des Site und seiner Netz-Einbindung,
- eine Bewertung der wissenschaftlichen Relevanz und Übersichtlichkeit,
- zusätzliche Kommentare und Bemerkungen, z. B. zu den Zugangsberechtigungen und -voraussetzungen (EDV-Anforderungen), Kosten etc.

Außerdem werden intern noch einige Kategorien zur Verwaltung der Daten benutzt. Die Kategorien sind in acht Bereiche eingeteilt: Bibliographische Daten, Klassifikation, zeitliche Daten, Verfügbarkeit, Bewertung, Statistik und interne Kategorien (siehe Kapitel 5.2.1).

Diese Kategorien sind an Dublin Core ausgerichtet, weil dies derzeit der international erfolgreichste Versuch ist, einen einheitlichen Standard zur Beschreibung von Informationsquellen zu entwickeln. Der Dublin-Core-Standard ist einerseits einfach genug, daß er auch für Laien handhabbar bleibt, und andererseits hinreichend komplex, um vielfältige Informationen aufnehmen und organisieren zu können (siehe Kapitel 5.2.2).

Nach dem benutzten Medium wird die *Art* der Informationsquellen unterschieden in on(line) = Internet, off(line) = CD-ROM, pr(inted) = Buch/ Zeitschrift. Der *Typ* dieser Quellen wird durch den *Source Type* (früher Formal Key) klassifiziert, mit dem versucht wird, eine für alle drei Bereiche und alle Fächer einheitliche Kategorisierung wissenschaftlicher Informationsquellen zu leisten (siehe Kapitel 5.3.1). Die *fachliche Klassifikation* erfolgt hingegen fachspezifisch, dar-

um kann hier auch keine Vorlage geliefert werden. Für einige Einträge werden Tabellen internationaler Standardabkürzungen benutzt: ISO 3166 für das Land, ISO 639-1 für die Sprache (siehe Kapitel 5.3.2 und Kapitel 5.3.3).

Allerdings ergeben sich auch einige Probleme im Umgang mit Metadaten, auf die hier nur kurz eingegangen wird:

- Copyrightfragen bei Übernahme schon vorhandener Metadaten entweder von Dritten oder vom Autor der Seite,
- Umgang mit vorhandenen Metadaten, die aber unvollständig, falsch, wenig aussagekräftig oder sonst inkompatibel zu dem SSG-FI-System sind,
- Fehlen eines verbindlichen internationalen Metadatenkonzepts, sowohl was die Semantik als auch die Syntax bzw. HTML-Darstellung angeht,
- Metadaten-Kategorien (z.B. DC.Creator) sind noch nicht im Internet suchbar, weil kommerzielle Suchmaschinen wie AltaVista etc. die Metadaten nicht indizieren,
- umgekehrt verzichten daher viele Webautoren auf die Bereitstellung von Metadaten.

3.3 Das sachliche Erschließungskonzept

Während die formale Erfassung gemäß dem „Dublin Core Metadata Set“ für die Bibliothek eine wichtige Entscheidung über die zukünftige Nutzbarkeit der Daten darstellt, ist aus der Sicht der Nutzer entscheidend, welche sachliche Erschließung ihm geboten wird. *Subject* und *Source Type Catalog* bilden schließlich für ihn den systematischen Zugang zu den Daten. Diese basieren auf zwei besonderen Kategorien, dem Formalschlüssel („Source Type“) und der Fachklassifikation („Subject“), die zur Strukturierung des Guide benutzt werden. Der Formalschlüssel ist in Anlehnung an die „Regeln für den Schlagwortkatalog“⁴² aufgebaut worden, um thematische Sites formal zu klassifizieren, umfaßt über 40 Dokumentarten und ist fachübergreifend definiert. Typische Ausprägungen sind der Server eines Fachbereichs, einer Fachgesellschaft, eines Forschungsinstituten oder einer Zeitschrift, aber auch Wörterbücher, Enzyklopädien, Datensammlungen, Adreßbücher, Bibliographien, Zeitschriftenverzeichnisse, Diskussionslisten usw. (vgl. Kap. 5.3.1). Da die SSG-FI-Guides international ausgerichtet sind, liegen sie in englischer Sprache vor, die Formalschlagwörter sind daher übersetzt worden.

Die Fachklassifikation muß für jedes Fach einzeln gewählt bzw. gestaltet werden (vgl. Kap. 5.4). Wegen der internationalen Ausrichtung sollte möglichst eine entsprechend anerkannte und verbreitete Klassifikation gewählt werden, die jedoch für viele Fächer nicht vorliegt. So wurde für den MathGuide die *Mathe-*

⁴² Regeln für den Schlagwortkatalog (RSWK), berab. von der Kommission des Deutschen Bibliotheksinstituts für Sacherschließung, 2. erw. Auflage, Berlin 1991

matics Subject Classification gewählt, die von den *Math Reviews* und dem *Zentralblatt der Mathematik* benutzt wird und in der Reinen Mathematik international angewandt wird. Für die anderen Bereiche wurde als Primärklassifikation die *Göttinger Online Klassifikation (GOK)* gewählt, da keine international akzeptierten Schemata vorlagen. Diese wird noch um die amerikanische *Dewey Decimal Classification (DDC)* und die *Basisklassifikation (BK)* des norddeutschen Gemeinsamen Bibliotheksverbundes (GBV) ergänzt. Aus Sicht der NutzerInnen bietet der möglichst internationale Standards aufgreifende fachklassifikatorische Einstieg die wichtigste und primäre Suchmöglichkeit.

Der Formalschlüssel kann als eine alternative Auswahlkategorie zum Einstieg in das Informationssystem genutzt werden, die z.B. dem Bibliothekar eine Übersicht über thematische Sites einer bestimmten Kategorie verschafft oder dem Fachwissenschaftler eine Liste der Fachgesellschaften zur Verfügung stellt. Einen dritten Zugang bietet die Suchfunktion, die es erlaubt, nach Wörtern im allgemeinen Text oder in bestimmten Kategorien zu suchen.

3.4 Auswahl- und Bewertungskriterien

Neben der sachlichen Erschließung durch Formalschlüssel und Fachklassifikation wird jeder thematische Site auch noch nach möglichst formalisierten Kriterien beschrieben. Ferner wird eine kurze Beschreibung (Abstract) des Site erstellt.

Wie bereits in Kapitel 2.3 dargelegt, ist mit der Selektion von Ressourcen für einen Katalog bereits eine erste Bewertung gegeben. Das heißt, daß durch die Aufnahme einer Ressource in den Guide unterstellt wird, daß es sich dabei um eine qualitativ hochwertige Ressource handelt. Im Rahmen von SSG-FI werden die Ressourcen *zusätzlich* bewertet. Das scheint deshalb sinnvoll, weil nicht alle katalogisierten Ressourcen in jeder Hinsicht und durchgängig qualitativ hochwertig sind. Es werden in einzelnen Fällen auch Informationsquellen katalogisiert und erschlossen, die bei einer strikten Auslegung der Kriterien nicht zwingend berücksichtigt würden. Das geschieht, um den Fachinformationsführer für die jeweiligen Fachwissenschaftler so „rund“ wie möglich zu gestalten. Ein Beispiel: die WWW-Seiten einer wichtigen renommierten Fachgesellschaft werden auch dann katalogisiert, wenn sie wenig mehr als Anschriften und Telefonnummern enthalten, weil ein fachwissenschaftlicher Nutzer eines SSG-FI davon ausgehen muß, eine solche Organisation im Katalog finden zu können. Mit einer solchen Vorgabe ist aber klar, daß das hinter einer solchen WWW-Seite stehende Informationsangebot genauer beschrieben werden sollte.

Im SSG-FI werden allerdings nicht die Forschungsinfrastruktur und ihre Institutionen evaluiert, sondern Fachinformationen. Mit anderen Worten: Es wird nicht die fachwissenschaftliche Bedeutung von Organisationen und Individuen *an sich* bewertet, sondern das inhaltliche Angebot der beschriebenen Ressou-

ce(n). Im Ergebnis bedeutet dies, daß der Site eines History Departments einer relativ unbekanntem und/oder „unbedeutenden“ Universität sehr wohl exzellente Fachinformationen liefern kann. Im Umkehrschluß kann vom Vorliegen hochwertiger Informationen auf einer bestimmten Site nicht zwingend geschlossen werden, daß es sich um eine in der „academic community“ bedeutende Organisation handelt; auch ist es keineswegs immer so, daß renommierte Fachgesellschaften über ihre Server hochwertige Fachinformationen bereitstellen.

3.4.1 Die einzelnen Kategorien der Bewertung





Das Team des SSG-FI Projekts hat sich dafür entschieden, jede katalogisierte Ressource nach verschiedenen Kriterien zu evaluieren. Damit soll unterschiedlichen Anforderungen, die fachwissenschaftliche Nutzer an elektronische Ressourcen stellen können, im einzelnen Rechnung getragen werden. Die Bewertungskategorien sind:

- Content Inhaltliche Relevanz,
- Clarity Übersichtlichkeit und Layout,
- Index Indizierung und
- Links Verweise auf weitere Ressourcen.

Dazu kommen als indirekte Kriterien:

- Level Zielgruppe, für die die Ressource geeignet ist und
- Backlinks Anzahl der Verweise auf diese Ressource von anderen Quellen.

Für die vier erstgenannten Aspekte sieht SSG-FI grundsätzlich folgende Bewertungsstufen vor:

	fair
	good
	excellent
	none (nur in Kategorie „Links“ möglich)

Mit der Kategorie **Level** soll die ungefähre Zielgruppe einer Informationsquelle angegeben werden:

- popular:* populärwissenschaftlich, d.h. ohne fachliche Vorkenntnisse verständlich
- undergraduate:* mit wissenschaftlichem Anspruch, aber eher auf Grundstudiums-Niveau
- graduate:* mit wissenschaftlichem Anspruch, auf Hauptstudiums-Niveau
- professional:* gerichtet an und geeignet für Fachwissenschaftler in diesem Bereich

Mit der Kategorie **Backlinks** wird die Anzahl der Verweise („links“) auf die Resource von anderen Sites im Internet aus angegeben. Die Anzahl der Treffer wird im Regelfall in Zehnerschritten auf- bzw. abgerundet. Die Kategorie kann mit einem „Citation Index“ im Bereich gedruckter Werke verglichen werden.

Tab. 3.2 gibt einen Überblick über das Bewertungsschema des SSG-FI Projektes. Die einzelnen Punkte werden in Kapitel 5.5 detailliert erläutert.

Kategorie	excellent ★★★	good ★★	fair ★
Content	„erschöpfende“ Behandlung des Themas; außergewöhnliche Qualität	qualitativ überdurchschnittliche Abhandlung des Themas	„ordentliche“ Behandlung des Themas
Clarity	optimales Layout: einer intuitiven Nutzung entgegenkommend	sehr übersichtliche Gestaltung; klare Gliederung	übersichtliche Gestaltung
Index	lokale Suchmaschine	Register („Site Index“)	kein Index
Links	viele, gut strukturierte und/oder annotierte Links	etliche (ggf.) geordnete und/oder annotierte Links	wenige, typischerweise nicht annotierte Links (Eingabe „-“ / none, wenn keine Hyperlinks)

Tab. 3.2: Bewertungsschema SSG-FI

4 Das SSG-Fachinformationssystem aus Benutzersicht

Von Thomas Fischer und Heike Neuroth

Unter dem Haupteinstieg (der „Homepage“) des SSG-FI-Bereiches kann auf die einzelnen Guides zugegriffen werden (vgl. Abb. 4.1).

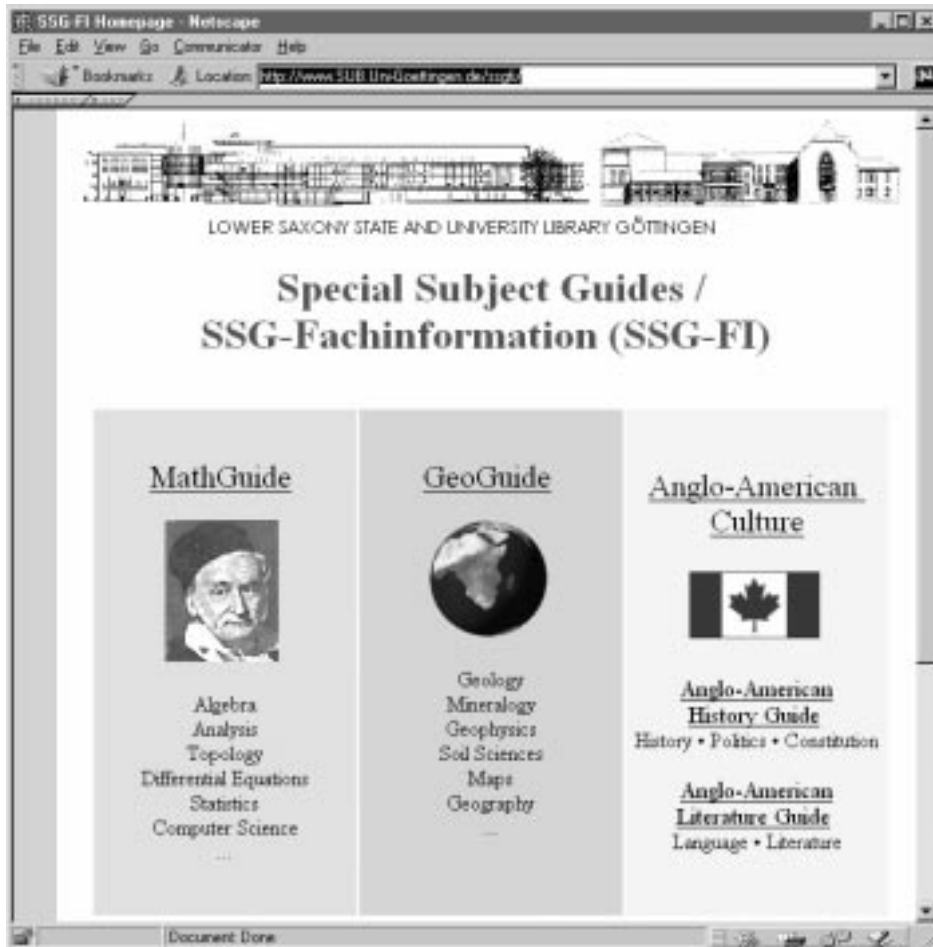


Abb. 4.1: Einstiegsseite von SSG-FI

Die Struktur des SSG-FI-Systems ist hierarchisch gegliedert und möglichst flach gehalten. Zunächst ist das gewünschte Fach auf der Homepage des SSG-FI-Projektes auszuwählen (vgl. Abb. 4.2). Hier stehen eventuell auch al-

gemeine Systeminformationen zur Verfügung, die aber zur eigentlichen *Benutzung* nicht wichtig sind. Dies ist die *globale* Auswahl:

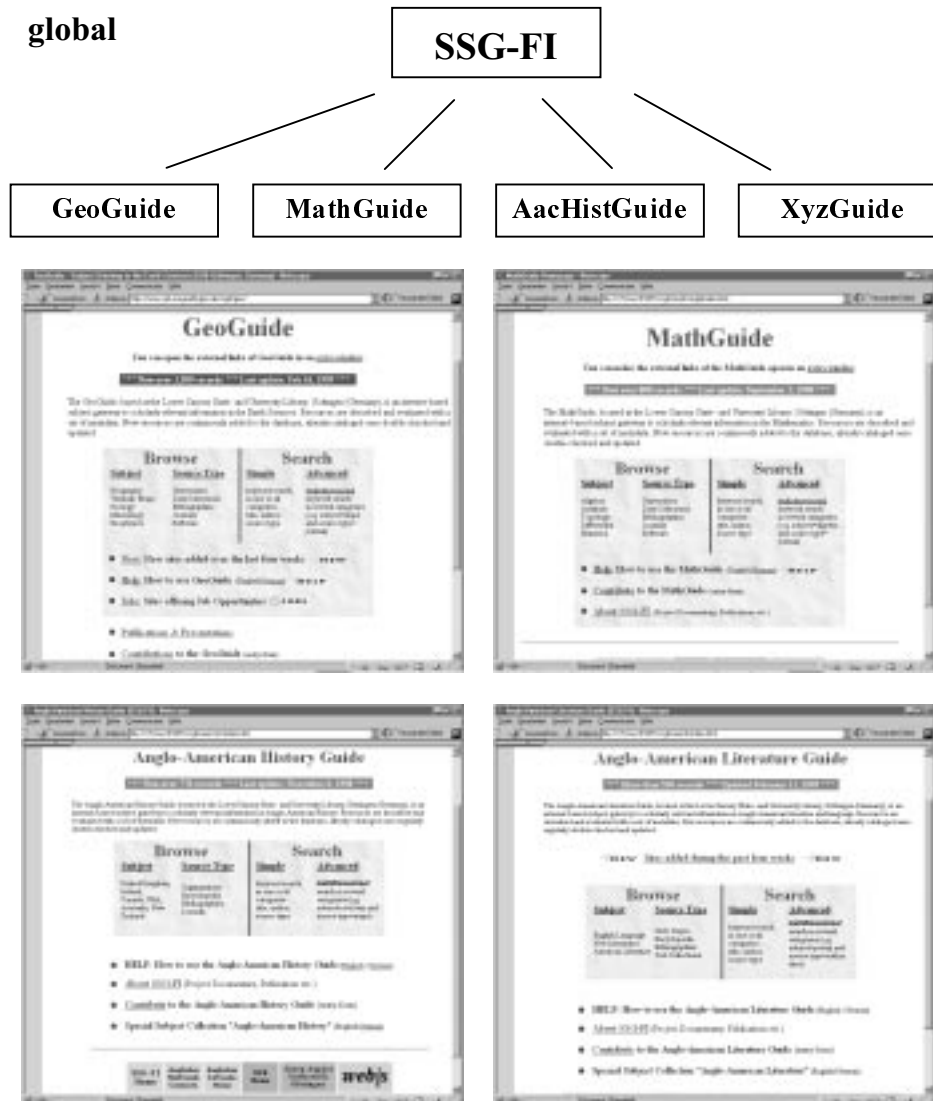


Abb. 4.2: Einstieg in die bereits vorliegenden Fach-Guides

Da gleichzeitig die Quellentypen Internet, CD-ROM und Buch unterschieden werden, ergibt sich eine mehrdimensionale Datenstruktur (vgl. Abb.4.3).

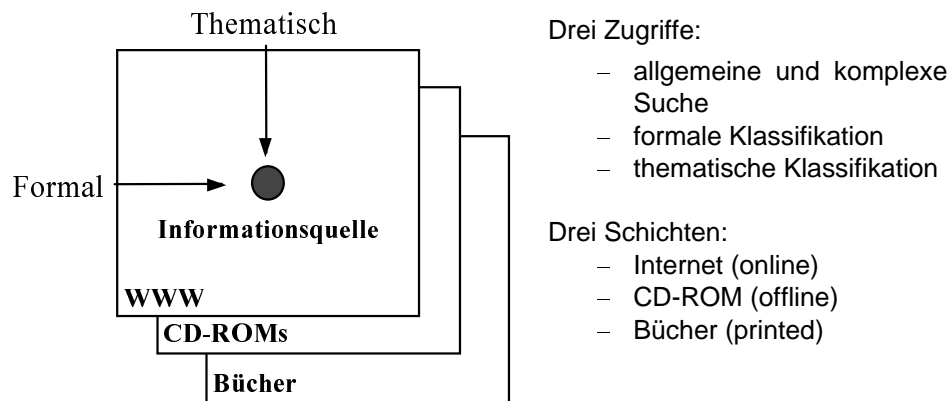


Abb. 4.3: Die mehrdimensionale Datenstruktur im SSG-FI System

Bei jeder der drei möglichen Einstiege wird der Nutzer bei framefähigen Browsern auf der linken Seite von einem Navigator begleitet, der ihm den Wechsel zwischen den einzelnen Seiten bequem ermöglicht. Für nicht framefähige Browser stehen jeweils am Ende einer Seite, die ebenfalls das Wechseln zwischen den Haupteinstiegsseiten erlauben.

4.1 Blättern („Browsen“) in den SSG-FI-Guides

Um dem Nutzer mehrere Einstiegsmöglichkeiten in den Guide anzubieten, wird ihm ein thematisch aufgebauter Katalog (Subject Catalog, vgl. Abb. 4.5) und ein formaler, nach Dokumenttyp sortierter Katalog (Source Type Catalog, vgl. Abb. 4.6) angeboten. Diese bieten also die Möglichkeit in den Datenbeständen zu blättern („Browsen“) und so einen Überblick über schon aufgenommenen Ressourcen zu bekommen. Außerdem gibt es verschiedenen Suchmöglichkeiten (vgl. Kap. 4.2)

Jeder Guide ist intern auf die gleiche Weise gegliedert, so daß sich in der lokalen Auswahl immer drei Möglichkeiten des Einstiegs über die Themenstruktur des Faches (Subject Catalog), die formale Beschreibung der Informationsquelle (Source Type Catalog) sowie zwei Arten der Suche (Simple Search und Advanced Search) bieten (vgl. Abb. 4.4).

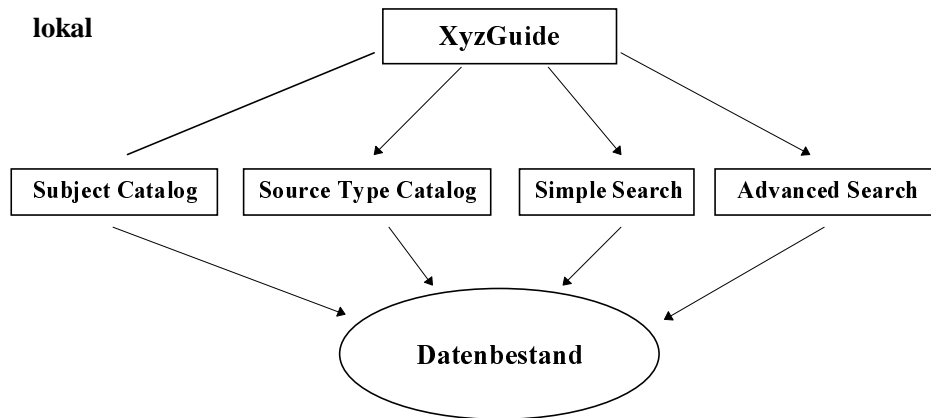


Abb. 4.4: Auswahl in einem SSG-FI Guide

Dem Subject (vgl. Abb. 4.5) und Source Type Catalog (Abb. 4.6) wird jeweils eine Kurzübersicht über den eigentlichen Catalog vorangestellt.

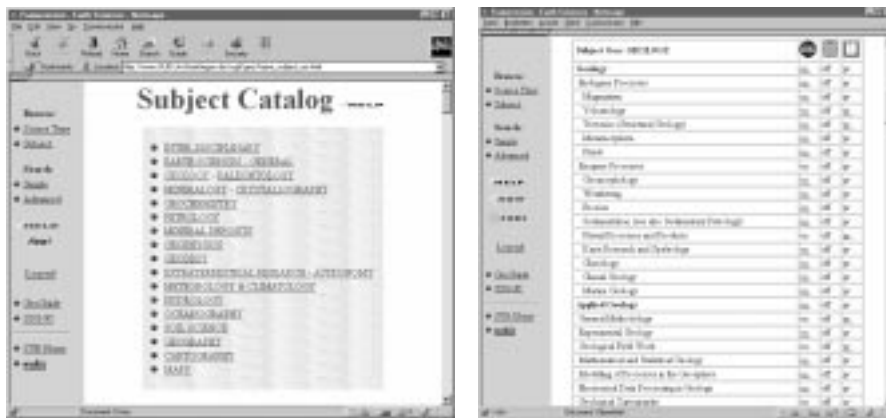


Abb. 4.5: Einstiegsseite des Subject Catalogs (GeoGuide) mit Kurzübersicht und Themenstruktur

Der Zugriff über den *Subject Catalog* erfolgt über die hierarchische Gliederung der Fachdisziplin (vgl. Kap. 5.4). Sinnvollerweise orientiert sich die Struktur an einem Klassifikationssystem, um über die Notationsstruktur die automatische Generierung der Ergebnislisten zu ermöglichen. Ein einheitlicher Ansatz für die fachliche Gliederung in den verschiedenen Wissensgebieten ist jedoch nicht zu formulieren, denn die Klassifizierung in verschiedenen Sondersammelgebietsfächern kann sehr unterschiedlich sein. Deshalb liegt es im Ermessen des jeweiligen SSG, ob auf eine wissenschaftlich anerkannte fachspezifische Klassifi-

kation oder ein bibliothekarisches Universalklassifikationssystem zurückgegriffen wird. Natürlich können auch mehrere Klassifikationssysteme verwendet werden, allerdings muß eine Fachklassifikation für den Aufbau des Subject Catalog gewählt werden (Primärklassifikation).

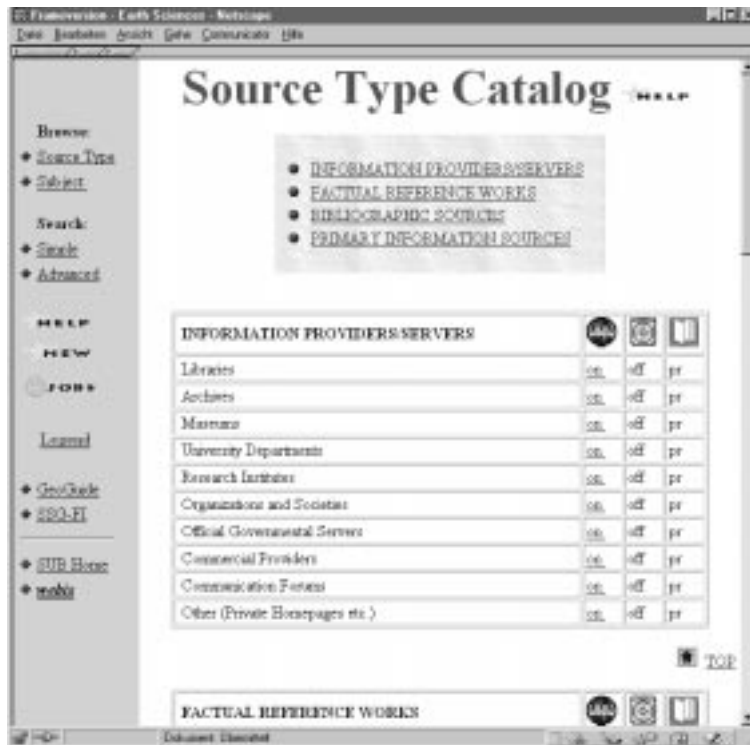


Abb. 4.6: Einstiegsseite des Source Type Catalogs

Im Source Type Catalog sind die einzelnen Dokumenttypen in vier Blöcken zusammengefaßt: Im ersten Block finden sich die verschiedenen Informationsanbieter, die bevorzugt als Online-Quellen formal kategorisiert werden können (Museen, Institute etc.). Im zweiten Block sind die verschiedenen Sachauskunftsmittel zusammengefaßt (Nachschlagewerke, Datenbanken). Im dritten Block finden sich die Literatursuchmittel (Bibliographien). Und im vierten Block werden spezielle primäre Informationsquellen angeboten (Handbücher, Zeitschriften).

Einzelne Kategorien können fachspezifisch weiter untergliedert oder ausdifferenziert werden, wenn die entsprechenden Listen zu groß werden. So kann eine Kategorie formal weiter untergliedert (z.B. die Zeitschriften nach dem Grad

der Verfügbarkeit) oder die Einträge einer Kategorie nach Ländern aufgeteilt werden (z.B. die Fachbereichsserver).

Die Zusammenfassungen der Datensätze werden beim Subject und Source Type Catalog als sogenannte Mini-Templates angeboten, die kurz über den Inhalt informieren sollen (vgl. Abb. 4.7). Folgende Angaben stehen dabei zur Verfügung: Titel, thematische bzw. formale Einhängung, Schlagwörter, URL, Niveau und Bewertung. Man kann sich nun zu einer Ressource das ausführliche Metadaten-Template ansehen oder sich direkt zu der Ressource „durchklicken“. Auf der linken Seite wird bei Auswahl der Legende in der Navigationsleiste über die Bedeutung der verschiedenen Symbole informiert.



Abb. 4.7: Layout der Mini-Templates am Beispiel eines Formalschlüssels

Zu jedem Zeitpunkt ist es möglich, eine speziell an den jeweiligen Guide angepaßte Hilfedatei aufzurufen, die dem Nutzer auch Ratschläge über die Handhabung und spezielle Tips gibt (vgl. Abb. 4.8).



Abb. 4.8: Einstiegsseite in die Hilfe-Datei der Guides

4.2 Suche im SSG-FI System

Für die effektive Nutzung der SSG-FI-Daten ist eine Suchfunktion notwendig. In der Allegro-Datenbank selbst gibt es dazu zwei Möglichkeiten:

- In den Registern werden alphabetische Listen einzelner Kategorien bereitgestellt. Dadurch können die zugehörigen Datensätze aufgerufen bzw. in ihnen geblättert werden.
- Die Freitextsuche liefert zu einem Suchbegriff (eventuell in einer bestimmten Kategorie) alle ihn enthaltenden Datensätze, wobei die Allegro-Suchfunktionen und die stattfindenden Konvertierungen zu beachten sind.

Bei dem Export auf den Webserver gehen diese Möglichkeiten verloren, und es müssen neue Suchmöglichkeiten aufgebaut werden. Dazu wird das Programm *swish-e* benutzt, das einerseits den Index aufbaut und andererseits die Suche in diesem Index durchführt.

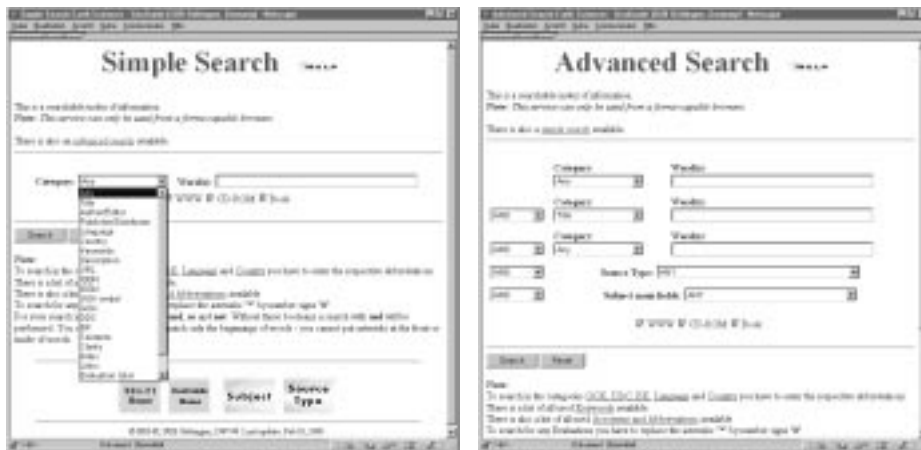


Abb. 4.9: Einstiegsseite der lokalen Suchmaschinen (Simple/Advanced Search)

4.2.1 Suche in den Guides des SSG-FI Projektes

Einfache freie Textsuche

Die einfachste Suchform ist die Suche nach einzelnen Ausdrücken: ein Wort wird eingegeben und der Suchmechanismus liefert eine Liste der entsprechenden Fundstellen. Dies sollte ohne weitere Zusatzüberlegungen zu nutzen sein, z.B. sollten Sonderzeichen wie Umlaute ohne Konvertierung oder Auflösung eingegeben werden können. Andererseits sollten auch Benutzer, die nicht ohne weiteres auf diese Zeichen zugreifen können, ohne Probleme suchen können. Dazu werden die Sonderzeichen auf ihre Grundzeichen reduziert, d.h. ä zu a, ö zu o usw. München wird also durch „Munchen“ gefunden. Das gleiche gilt auch für die anderen benutzten Sonderzeichen mit Akzent, Tilde oder „^o“, oder für æ und œ. Zu beachten ist nur, daß „ß“ nicht aufgelöst wird und daher schwerer zu finden ist.

Da diese automatische Reduzierung von dem Indexprogramm durchgeführt wird, erschien es nicht notwendig, zusätzliche Listen von Standardformen anzulegen. Die Groß/Kleinschreibung wird bei dieser Suche ignoriert, auch das erscheint kein großes Problem bei den gegebenen Daten.

Die Möglichkeiten der Wortsuche werden erheblich erweitert, wenn eine Suche mit Platzhaltern möglich ist. Hier erlaubt das SSG-FI-System leider nur das Abschneiden („Trunkierung“) von Wörtern: Ein Sternchen am Ende des Ausdruck erlaubt es, alle Wörter mit dem Ausdruck am Anfang zu finden. Bei einem Sternchen innerhalb des Wortes wird das Ende des Wortes hinter dem Sternchen ignoriert; Suche nach Wortenden mit einem Sternchen am Anfang ist nicht möglich.

Die spezielle Bedeutung des Sternchens in der Suche führt dazu, daß nach Sternen direkt nicht gesucht werden kann: um nach den Bewertungen zu suchen, die mit Sternen ausgedrückt werden, muß für „*“ ein „#“ eingegeben werden.

Wegen der inneren Struktur des aufgebauten Index ist es nicht möglich, nach Ausdrücken zu suchen, die einen Doppelpunkt enthalten. Das ist ein Problem, da vollständige Internet-Adressen aus dem Protokoll (http, ftp; news) einem *Doppelpunkt*, ein oder zwei Schrägstrichen und der URL bestehen. Diese Ausdrücke können so also nicht gesucht werden. Die Suche wurde so implementiert, daß nach der eigentlichen Adresse, dem Ausdruck hinter dem Schrägstrich gesucht werden kann. Damit kann allerdings nicht nach allen Datensätzen mit einem bestimmten Protokoll (z.B. allen ftp-Servern oder newsgroups) gesucht werden.

Nicht möglich ist die Suche nach zusammengesetzten Ausdrücken wie „Earth Science“ oder „Commutative Algebra“, da der Index nur aus einzelnen Wörtern besteht. Daß die Suche nach solchen Begriffen trotzdem normalerweise das gewünschte liefert, liegt an der automatischen AND-Verknüpfung bei so einer Anfrage.

Suche mit booleschen Operatoren

Eine speziellere Suche ist mit den booleschen Operatoren AND, OR, AND NOT möglich. Die Voreinstellung der Suche ist so, daß einzelne Wörter automatisch mit „AND“ verknüpft werden, die Eingabe von „Earth Science“ z.B. also die Suche „Earth **AND** Science“ startet. Das liefert auf jeden Fall die Dokumente, die den Begriff „Earth Science“ enthalten, außerdem aber auch alle Dokumente, in denen „Earth“ und „Science“ unabhängig voneinander vorkommen. Die Verknüpfung mit „OR“ und „AND NOT“ ermöglicht die Suche nach Alternativen oder das Ausschließen unerwünschter Begriffe.

Für noch komplexere Suchanfragen können diese Operatoren zusammengesetzt und geklammert werden, so daß z.B. eine Anfrage „Commutative AND

(Algebra OR Rings)“ möglich ist. Bei all diesen Möglichkeiten können die Suchbegriffe auch mit dem Sternchen abgeschnitten werden.

Suche in SSG-FI-Kategorien

Während die bisher beschriebenen Suchfunktionen eine allgemeine Textsuche durchführen, gibt es auch die Möglichkeit, in den einzelnen Kategorien des SSG-FI-Projektes zu suchen. Die Syntax der Anfrage ist so, daß die Bezeichnung des zugehörigen Meta-Tags mit einem Gleichheitszeichen mit der Suchanfrage verbunden wird: „DC.Subject=migration“ sucht in den Schlagwörtern nach dem Wort „migration“. Die Meta-Tag-Umsetzung der Allegro-Kategorien wird in Abschnitt 5.2.1 beschrieben.

Die auf diese Weise gebildeten Ausdrücke können dann mit den booleschen Operatoren verknüpft oder auch durch Sternchen abgeschnitten werden; auch Kombinationen mit Klammern sind möglich.

Während auf diese Weise alle Möglichkeiten der Suche gegeben sind, ist es doch recht unbequem, sich die Meta-Tags zusätzlich zu den benutzten Kategoriebezeichnungen zu merken. Damit das nicht nötig ist, wurde die Benutzerschnittstelle der Suche entsprechend eingerichtet.

4.2.2 Die Benutzerschnittstelle des Suchsystems

Im Internet wird die Suche in den SSG-FI-Guides durch Eingabeformulare unterstützt, die erlauben, die Suchbegriffe einzugeben und die gewünschte SSG-FI-Kategorie auszuwählen. Es stehen dabei in jedem Bereich zwei Formulare zur Verfügung, die einfache und die komplexe Suche. Die „einfache Suche“ bietet einen möglichst unkomplizierten Einstieg in die Suche, die „komplexe Suche“ vielfältige Möglichkeiten einer sehr gezielten Suche.

Die einfache Suche („Simple Search“)

Das Formular der einfachen Suche bietet nur zwei Haupteingabemöglichkeiten: die Feldauswahl und das Eingabefeld. In das Eingabefeld können beliebige Ausdrücke der unter Abschnitt 4.2.1 beschriebenen Form eingetragen werden, und die Suche liefert das Gewünschte.

Der zusätzliche Komfort liegt in der Feldauswahl. In der Standardeinstellung „Any“ wird in allen Kategorien, genauer gesagt im gesamten sichtbaren Text der Infodateien gesucht. Alternativ kann eine der Kategorien des SSG-FI-Projektes ausgewählt werden, dann wird der eingegebene Ausdruck nur in der gewünschten Kategorie gesucht. Dabei werden die Kategorien mit der Bezeichnung angegeben, mit der sie in den Infodateien erscheinen, so daß die interne Codierung nicht bekannt sein muß. Bei der Ausgabe der Suche wird die umcodierte Suchanfrage mit ausgegeben, so daß sichtbar ist, ob die Anfrage auf die gewünschte Weise umgesetzt wurde.

Auch hier stehen das Abschneiden und die booleschen Operatoren zur Verfügung. Die Kategorien, die zur Auswahl angeboten werden, unterscheiden sich in den verschiedenen SSG-FI-Guides. Da für eine einfache Suche nicht alle Kategorien angeboten werden sollen, stehen hier nur die wichtigsten wie Titel, Schlagwörter, URL, Sprache und Land zur Verfügung.

Für die Suche in den Kategorien Country und Language ist zu beachten, daß diese als Abkürzungen (nach ISO 3166 bzw. nach ISO 639-1) eingetragen werden müssen, also nach Großbritannien als „GB“ und nach finnisch als „fi“ gesucht wird (vgl. Kap. 5.3.2 bzw. 5.3.3). Dies ist auch ein Unterschied zwischen der Suche in Kategorien und der freien Textsuche („Any“): Im sichtbaren Bereich werden diese Abkürzung übersetzt, so daß nicht „fi“, wohl aber „Finnish“ gefunden wird.

Das gleiche gilt für andere Bereiche, wo Abkürzungen verwandt werden, z.B. für den Formalschlüssel („Source Type“). Da die Suche nach einem Formalschlüssel insofern nichts Neues ergibt, als die entsprechenden Dokumente auch über den „Source Type Catalog“ erreichbar sind, wird diese Möglichkeit nicht bei allen Guides angeboten. In einem anderen Formular wird die Eingabe dieser SSG-FI-spezifischen Abkürzungen durch ein Auswahlmenü ersetzt.

Zusätzlich zu den beschriebenen Eingabe- und Auswahlmöglichkeiten stehen in dem einfachen Suchformular noch drei Auswahlkästchen zur Verfügung, die standardmäßig alle eingeschaltet sind: *WWW*, *CD-ROM* und *Book*. Damit kann die Suche auf die verschiedenen Dokumentarten im SSG-FI-Projekt eingeschränkt werden. Sind alle ausgeschaltet, so wird diese Option nicht genutzt (also im gesamten SSG-FI-Guide gesucht).

Für die Suche im Bereich des Anglo-Amerikanischen Kulturraums steht noch eine weitere Option zur Verfügung: Es kann ausgewählt werden, ob im Literatur- oder Geschichtsbereich oder in beiden gesucht wird; die Suche gibt dann an, in welchem Index die Suche welche Einträge gefunden hat. Diese Option wurde aufgrund der inhaltlichen Nähe der beiden Guides eingeführt: literarische Quellen können auch von geschichtlicher Bedeutung sein und umgekehrt. Technisch könnte auch bei den anderen Guides so eine Auswahl vorhanden sein, aber es ist nicht klar, warum jemand in Mathematik und Geowissenschaften gleichzeitig suchen sollte.

Die komplexe Suche („Advanced Search“)

Mit dem Formular der komplexen Suche werden die Möglichkeiten der einfachen Suche nicht wesentlich erweitert, wohl aber umfassender unterstützt. Es gibt hier drei Eingabefelder, die jeweils mit einem Auswahlmenü für die zugehörige Kategorie verbunden sind; außerdem kann der verbindende boolesche Operator ausgewählt werden (Voreinstellung: AND), so daß die Anfrage „DC.Title=Algebra and SSG.Country=DE and DC.Subject= commutative“ ein-

fach durch Auswahl von „Title“, „Country“ und „Keywords“ sowie der entsprechenden Operatoren formuliert werden kann. Dabei müssen auch hier für Sprache und Land die entsprechenden Abkürzungen eingegeben werden.

Zusätzlich gibt es bei den unterschiedlichen SSG-FI-Guides noch weitere Auswahlfelder für den Formalschlüssel („Source Type“), den Themenschlüssel („Subject“) oder die fachlichen Hauptthemen („Subject Main Fields“). Dabei werden intern die zugehörigen Abkürzungen benutzt, die im letzten Fall extra für die Suche erzeugt werden und somit unsichtbar bleiben.

Die einzelnen Eingabefelder haben die gleiche Funktionalität wie bei der einfachen Suche, so daß insbesondere weiterhin die Möglichkeit besteht, Begriffe mit booleschen Operatoren zu verbinden. Auch eine zusätzliche Klammerung ist möglich - wegen der vielen Kombinationsmöglichkeiten aber eigentlich nicht notwendig.

Das Suchergebnis

Das Ergebnis jeder Suche ist eine Auflistung der gefundenen Datensätze. Zusätzlich wird die von dem Suchsystem verarbeitete Anfrage angezeigt, die ja bei den Auswahlmenüs und der Klammerung nicht unbedingt vorhersehbar ist. Das erlaubt eine elementare Kontrolle der Suchanfrage.

Als Ergebnis werden die gefundenen Titel angezeigt, die mit der jeweiligen Info-datei verbunden sind, so daß durch Anklicken die Datei aufgerufen wird. Derzeit können diese Fundstellen nicht nacheinander durchgeblättert werden, sondern die Fundliste dient als Verteiler zu den Datensätzen, zu dem immer wieder zurückgesprungen werden muß. Vorstellbar wäre ein Programm (mit JAVA oder JavaScript geschrieben), das es ermöglicht, die Fundstellen mittels Vor- und Zurück-Tasten abzuarbeiten.

Die Titel werden in alphabetischer Reihenfolge angezeigt, da das Suchprogramm keine andere sinnvolle Option bietet. Intern werden außer dem Titel noch die Größe der gefundenen Datei (jeweils etwa 5 kB), der Dateityp (immer HTML) und eine Trefferbewertung („Score“, Bedeutung ist nicht dokumentiert) angeboten, alles Daten, die für die spezielle Situation der SSG-FI-Guides nicht sinnvoll verwertet werden können.

Es wäre wünschenswert, die Treffer nach andern Kriterien (z.B. Land oder Bewertung) anzuzeigen oder auch zusätzliche Informationen zu liefern, wie sie in den Verteilerdokumenten gegeben werden.

Die Suche selbst erfolgt schnell und zuverlässig und ist so ein wichtiges Hilfsmittel bei der Orientierung in den SSG-FI-Guides.

5 Das Erschließungskonzept des SSG-Fachinformationssystems

Von Thomas Fischer, Karsten Hülsemann, Heike Neuroth und Norbert Pfurr

5.1 Der Prozeß der Erschließung

Es ist an dieser Stelle wichtig zu wiederholen, daß SSG-FI nicht anstrebt, Einzeldokumente zu erschließen, sondern „thematische Sites“ (vgl. Kap. 3.1). Was im einzelnen darunter verstanden wird ist z.T. sehr heterogen. Es kann sich um eine Bibliographie gedruckter Literatur zu einem abgegrenzten Thema handeln, die im technischen Sinne aus einer einzelnen (evtl. umfangreichen) HTML-Datei besteht. Es kann sich allerdings auch um den Server einer Fachgesellschaft handeln, auf dem sich Hunderte, ggf. Tausende von mehr oder weniger stark miteinander inhaltlich verbundenen und technisch verknüpften („inter-linked“) Dateien befinden. Die Methode, mit der SSG-FI dieser Heterogenität Rechnung trägt, ist die zusätzliche formale Erschließung nach Ressourcentyp (vgl. dazu unten). Damit ist es zum einen möglich, verschiedene Site-Typen zu unterscheiden und den Nutzern bereits über die formale Klassifikation eine erste Information zur (wahrscheinlichen) Komplexität des Site zu liefern. Zum anderen aber kann damit ein komplexer Site „zerlegt“ werden. So ist es möglich, einen umfangreichen thematischen Site durch die Erstellung von mehreren Datensätzen genauer zu erfassen. Dabei wird der Site in mehrere Teilbereiche aufgegliedert, die einzeln aufgenommen werden. Ein Beispiel dafür ist der komplexe Site „U.S. Geological Survey“. Dieser kann einmal als allgemeiner Einstiegspunkt beschrieben und unter dem formalen Schlüssel „Official Governmental Servers“ eingeordnet werden. Teilbereiche des Site können aber auch noch unter den Formalschlüsseln „Data Collections“, „(Current) Research Projects“, oder „List of Government Reports“ klassifiziert werden und sind damit über den Source Type Catalog separat und somit gezielter zugänglich.

Bei solch „großen“ Sites stellt sich gleichwohl in jedem Fall die Frage, in welcher „Tiefe“ diese erschlossen werden sollen. Um eine Analogie zur bibliothekarischen Arbeit im Bereich gedruckter Werke herzustellen: Wo der traditionelle Bibliothekar auf Verlagsmitteilungen und etablierte Rezensionsorgane zurückgreifen kann, um eine Entscheidung über die Aufnahme (Kauf bzw. Bestellung) zu treffen, sind solche Hilfsmittel im Online-Bereich nur ansatzweise vorhanden.⁴³ In aller Regel muß sich der „Internet Librarian“ eine Information über Online-Ressourcen anhand der Ressource selbst verschaffen. *Idealerweise* sollten Internet-Sites dann wenigstens einen Bestandteil mit Informationen zum Site selbst („about this site“ o.ä.) haben, *typischerweise* haben sie dies jedoch nicht.

43 Ein gutes Beispiel eines solchen Rezensionsorgans ist der bereits erwähnte *Internet Scout Report*.

Das heißt, es muß zunächst eine grundsätzliche Durchsicht erfolgen, um überhaupt eine Entscheidung über Aufnahme bzw. Nichtaufnahme in den Katalog (d.h. die Datenbank) zu treffen. Ist die Entscheidung *für* eine Aufnahme gefallen, beginnt die eigentliche Erschließung der Ressource. Diese Tätigkeit entspricht wiederum der traditionellen bibliothekarischen Sacherschließung. So wie der Fach-Bibliothekar die von ihm ausgewählten Titel nicht von der ersten bis zur letzten Seite durcharbeiten kann, bevor er sie sachlich erschließt, so kann auch der „Internet Librarian“ nicht jedem Verweis (d.h. Hyperlink) innerhalb eines Site folgen, um eine Übersicht über die Inhalte zu bekommen.

Natürlich ist eine vollständige Durchsicht in den meisten Fällen nicht grundsätzlich unmöglich. Doch ein vernünftiges Verhältnis von (zeitlichem und damit personellem) Aufwand und (Daten-)Ertrag läßt sich damit nicht erreichen. Ökonomisch gesprochen: Der informationelle Grenznutzen ist tendenziell abnehmend. Wenn die Alternative besteht, einen Site vollständig „durchzuklicken“ oder die Zeit zu nutzen, um zusätzliche/weitere Ressourcen zu erschließen, so ist im Regelfall die letztere Vorgehensweise effektiver. In Anlehnung an das geflügelte Wort vom „librarian who reads is lost“ ließe sich sagen: „The Internet librarian who follows every hyperlink is lost as well“.

Diese Problematik läßt sich anhand verschiedener Site-Typen verdeutlichen: Wenn für die SSG-FIs die einschlägigen „University Departments“ erfaßt werden, ist es in aller Regel wenig einträglich, den üblicherweise vorhandenen Links zu „faculty“ oder „graduate studies“ im einzelnen zu folgen. Natürlich besteht damit das Risiko, bestimmte Informationsangebote, die etwa (nur) über die Homepage eines Fakultätsmitglieds oder unter einem Punkt „studentische Web-Projekte“ zugänglich sind, zu übersehen. Es ist aber weder die Absicht noch die Aufgabe der SSG-FI, Defizite von Informationsanbietern in deren Benutzerführung zu kompensieren. Hinzu kommt noch, daß jederzeit vom Betreiber eines Servers Dokumente hinzugefügt, entfernt oder aktualisiert werden (können).

Auf der anderen Seite bietet die Hyperlink-Struktur des Internet Anderen, die an einer ähnlichen Thematik arbeiten, die Möglichkeit, von ihren Sites auf solche Informationen zu verweisen. Die Erfahrung bei SSG-FI hat gezeigt, daß die Internet Community von dieser Möglichkeit durchaus rege und effektiv Gebrauch macht. Ein Beispiel mag das verdeutlichen: Auf dem Server des Department of History der University of Rochester (<http://www.history.rochester.edu/>) liegen mehrere Jahrgänge des *Penny Magazine*, einer für Historiker durchaus interessanten Zeitschrift des 19. Jahrhunderts, im Volltext bereit. Von der Homepage und der „Links“-Seite des Departments selbst allerdings finden sich zwar Verweise auf verschiedene lokale Projekte mit relevanten Materialien (z.B. zur Lokalgeschichte und zum Eerie Canal), aber es findet sich keinerlei Verweis auf das *Penny Magazine*. Die „Entdeckung“ dieser Ressource für SSG-FI erfolgte

über einen anderen Subject Gateway, das „Electronic Text Center der University of Virginia“⁴⁴, welches darauf verweist.

5.2 Das SSG-FI Metadatenkonzept

Das Erschließungskonzept mit Dublin Core und SSG-FI Kategorien, mit formalen Kategorien (Source Type Catalog) und der thematischen Erschließung (Subject Catalog) soll zum Ziel haben, thematische Sites möglichst umfassend zu beschreiben und zu katalogisieren.

Für die Beschreibung thematischer Sites im Internet wurde im SSG-FI-Projekt ein Konzept der formalen und thematischen Erfassung entwickelt. Ausgehend von der Dublin-Core-Debatte wurden zu Anfang der Projektlaufzeit (1996) ca. 40 Kategorien festgelegt, mit denen die Informationsressource nach Form und Inhalt beschrieben wird. Diese Kategorien werden für die Beschreibung der einzelnen Sites von Wissenschaftlern bzw. Bibliothekaren benutzt. Der für jeden Site erstellte Datensatz wird in einer Allegro-Datenbank gespeichert.

5.2.1 Die Kategorien des SSG-FI-Systems

Die Kategorien des SSG-FI-Systems sind in sieben Blöcke eingeteilt: Bibliographische Daten, Klassifikation, Daten, Verfügbarkeit, Bewertung, Statistiken, Interna. Bis auf die Interna werden diese Angaben in der Internet-Version der SSG-FI-Guides veröffentlicht und so als Beschreibung der erfaßten Sites zur Verfügung gestellt.

In der folgenden Tabelle werden die Kategorien kurz vorgestellt. Dazu gehört die (interne) Kategoriennummer, der (englische) Namen der Kategorie, eine kurze Beschreibung des Inhalts und die Kodierung der Kategorie für die Meta-Tags der HTML-Dokumente. Die Erfassung der Sites erfolgt auf englisch, und es wird möglichst eine englischsprachige Seite als Adresse angegeben. Die Schreibweise von Namen und Orten wird (soweit möglich) in der Landessprache gegeben. Wenn in einer Kategorie *mehrere Einträge* erfolgen, sind diese durch *Semikolon und Leerzeichen* von einander zu trennen. Kategorien, in denen Mehrfacheinträge üblicherweise möglich sind, sind in der Tabelle mit einem „*“ gekennzeichnet.

Persönliche Namen sollten im Format „Nachnamen, Vornamen (E-Mail)“ angegeben werden; die Namen von Organisationen, Gesellschaften o.ä. werden als „Name, Ort“ angegeben; das Land wird hinzugefügt, falls es von dem in Kategorie 17 abweicht. Zeiteinträge erfolgen im Format JJJJMMTT, z.B. 19980104. Wenn Tag oder Monat nicht zu ermitteln sind, wird entsprechend abgeschnitten, ist das Datum gar nicht angegeben, wird „unknown“ eingefügt.

⁴⁴ URL: <http://etext.lib.virginia.edu/>

Nr.	Name	Inhalt	Kodierung
Bibliographische Daten			
00	File Number	automatisch vergebene interne Ressourcennummer, ergibt Namen der Infodatei.	nur interne Information
10	Title	Title wird benutzt, um das zugehörige Info-Dokument in den Verteilerlisten anzuzeigen und muß daher das Dokument (möglichst) eindeutig charakterisieren. 1. Der vom Verfasser, Urheber oder Verleger vergebene Name der Ressource. In HTML-Dokumenten der Titel laut Titelleiste oder Überschrift. 2. Wenn das nicht aussagekräftig ist, dann aus dem Dokument abgeleitet.	DC.Title
11	Author*	Die Personen, die intellektuell für den Inhalt verantwortlich sind (Webmaster z.B. sind keine Autoren). Eintrag <i>nur bei gesicherten Autoren</i> eines Dokuments.	DC.Creator
12	Editor*	Editoren oder Editorengremium, verantwortlich für den intellektuellen Inhalt eines Dokuments, i.A. alternativ zu Author (Webmaster etc. sind auch keine Editoren, dieses Feld kann auch freigelassen werden).	DC.Creator
13	Publisher	Verlag, Gesellschaft, Fakultät etc., verantwortlich für den Vertrieb eines Dokuments aus inhaltlicher Sicht.	DC.Publisher
14	Distributor	Betreiber des Servers (Fakultät/ Universitätsserver/ Rechenzentrum), verantwortlich für den Vertrieb eines Dokuments aus technischer Sicht, zusätzlich Angabe der URL, wenn nicht in #21 oder #27 enthalten.	DC.Publisher
15	Type	Art des Dokumentes: on(line), off(line) oder pr(inted) (für WWW, CD-ROM, Buch)	DC.Type
16	Language*	Sprache des Dokumentes nach ISO 639-1 (immer in Kleinbuchstaben).	DC.Language
17	Country	Ländercode nach ISO 3166:1993 (immer in Großbuchstaben), (deutsche Fassung DIN EN 23166:1995).	SSG.Country

18	Format*	Bei Internet-Quellen das datentechnische Format der Ressource (in Anlehnung an den MIME-Typ), z.B. text/html (Textdatei), jpeg, gif (Bilddaten), tar, zip (komprimierte Daten). Bei CD-ROMs: benötigte Software, Betriebssystem, Mindestvoraussetzungen (Arbeitsspeicher etc.); bei gedruckten Quellen Angabe über die Buchgröße (Oktav, Quart, Folio).	DC.Format
19	Keywords*	Schlagwörter, die das Thema und den Inhalt beschreiben, bis auf Eigennamen klein geschrieben. Möglichst mindestens je ein Schlagwort zu Region/ Ort, Servertyp, Ressourcentyp, Name, zur fachspezifischen Beschreibung möglichst einen entsprechenden internationalen Thesaurus verwenden.	DC.Subject
20	Description	Beschreibung des Site („abstract“).	DC.Description
21	URL	Bei Online-Dokumenten eindeutige Identifikation des Site durch die WWW-Adresse (mit Protokoll: http, ftp, gopher).	DC.Identifier (URL ist Standard)
22	ISSN/ISBN	Zur eindeutigen Identifikation <i>gedruckter</i> Ressourcen (International Standard Book Number) bzw. <i>gedruckter oder elektronischer Periodika</i> (International Standard Serial Number), evtl. analoge Daten für periodisch erscheinende CD-ROMs.	DC.Identifier SCHEME="ISSN/ISBN"
24	Call Number	Nur bei CD-ROMs oder gedruckten Ressourcen: Signatur der Bibliothek (# 28).	DC.Identifier SCHEME="CallNumber"
25	Mirrors*	WWW-Adresse(n) gespiegelter Versionen oder einer Auswahlliste von Spiegeln. Dieses Feld kann auch freigelassen werden.	DC.Relation SCHEME="URL" ROLE="mirror"
26	Derived from	Dient zur Darstellung der Bezüge zwischen Ressourcentypen; falls dem Dokument eine andere Version zugrunde liegt: ISSN, ISBN, URL. Dieses Feld kann auch freigelassen werden.	DC.Relation SCHEME="ISSN", "ISBN", "URL"

27	Contained in	URL der Quelle, in der die Informationen eingebunden sind, übergeordneter Site, z.B. Server der Universität oder Gesellschaft, falls URL abweichend von #13, #14, #21. Bei CD-ROMs oder gedruckten Ressourcen: Eintrag, wenn es sich z.B. um einen Stücktitel einer Serie handelt. Dieses Feld kann auch freigelassen werden.	DC.Relation SCHEME= "URL" ROLE= "parent"
28	Archived by	Bibliothek, Archiv; <i>notwendig</i> bei CD-ROMs oder gedruckten Ressourcen, bei Online-Ressourcen kann dieses Feld auch freigelassen werden.	DC.Relation ROLE= "archive"
29	Source Type*	Dokumenttyp (Formalschlüssel), wichtig für die Verteiler des Source Type Catalog (z.B. Datenbanken, Bibliographien...,vgl. Kap. 5.3.1).	SSG.Formal
Klassifikation			
40	KAT verbal*	Fachliche Klassenbezeichnung, für Mathematik KAT = MSC, sonst GOK	DC.Subject SCHEME= "KATVerbal"
41	KAT*	Fachliche Primärklassifikation, wichtig für die Verteiler des Subject Catalog (vgl. Kap. 5.4.1)	DC.Subject SCHEME= "KAT"
42	DDC*	Dewey Decimal Classification, international (vgl. Kap. 5.4.2)	DC.Subject SCHEME= "DDC"
43	BK*	Basisklassifikation des Gemeinsamen Bibliotheksverbundes (GBV) (vgl. Kap. 5.4.2)	DC.Subject SCHEME= "BK"
Daten			
50	Date	Bei Internet-Ressourcen das Datum, an dem ein Dokument zum letzten Mal geändert wurde, bei CD-ROMs oder gedruckten Ressourcen Erscheinungsdatum.	DC.Date
51	Course	Nur bei Periodika: Veröffentlichungszyklus einer Ressource, z.B. biweekly, monthly, continuous, irregular	SSG.Course
52	Temporal Coverage	Zeitraum, für den Informationen bereitliegen (dieses Feld kann auch freigelassen werden)	DC.Coverage. t.min bis DC.Coverage. t.max

Verfügbarkeit			
60	Access	Angaben zur Verfügbarkeit des Dokumentes und über mögliche Kosten für den Benutzer. Wenn keine Einschränkungen vorhanden sind, dann: „free“. Bei CD-ROMs oder gedruckten Ressourcen: available via SUB Göttingen, GBV Verbundkatalog, Karlsruher Virtueller Katalog (KVK) etc.	DC.Rights. Access
61	Restriction	Angaben über sonstige Beschränkungen: z.B. über spezielle Zugangsberechtigungen etc., wenn keine Beschränkungen vorhanden sind, dann: „none“. Bei CD-ROMs und gedruckten Ressourcen ist der Eintrag abhängig von den SUB-Bestimmungen bzw. der Richtlinien anderer Bibliotheken (eingeschränkte Ausleihfristen, nur in der SUB einsehbar etc.).	DC.Rights. Restrictions
62	Remarks	Zusatzinformationen zu Art bzw. Aufhebung der Beschränkungen: Bei Online-Ressourcen auch Angaben über benötigte Software; bei CD-ROMs über Subskriptionsraten, ansonsten wird das Feld freigelassen.	DC.Rights. Remarks
Bewertung (vgl. Kap. 3.3)			
70	Contents	Eintrag von * bis ***: Bewertet wird die inhaltliche Relevanz der Informationsquelle für das entsprechende Thema und Niveau.	SSG.Evaluati- on.Contents
71	Clarity	Eintrag von * bis ***: Bewertet wird die Übersichtlichkeit, das Layout, die Darstellung (Optik).	SSG.Evaluati- on.Clarity
72	Index	Eintrag von * bis ***: Bewertet wird die Indizierung: Gliederung, Inhaltsverzeichnis, Register, Suchmaschine, Abfragemöglichkeiten bei einer Datenbank.	SSG.Evaluati- on.Index
73	Links	Eintrag von * bis ***: Bewertet wird die Anzahl, Bedeutung und Beschreibung der Links zu verwandten Themen (nur bei Online-Ressourcen).	SSG.Evaluati- on.Links
74	Level*	Angabe über das Niveau der Informationsquelle und die Anforderungen an Vorkenntnisse. Einträge: popular; undergraduate; graduate; professional.	SSG.Evaluati- on.Level

Statistiken			
80	Size	Bei Online-Ressourcen: Umfang von Datenbanken, Angaben über Suchmaschinen, Anzahl von Verteilern, Photos etc.; bei CD-ROMs: Anzahl der Datenbankeinträge etc.; bei gedruckten Dokumenten Seitenzahl, Tabellen etc. (dieses Feld kann auch freigelassen werden).	SSG.Size
82	Backlinks	Anzahl der Verweise auf diesen Site von anderen Sites aus, wenn keine gefunden, dann: „0“ (nur bei Online-Ressourcen).	SSG.Backlinks
90	Notes	Weitere Bemerkungen zu der Informationsquelle, z.B. ob Teile davon noch im Aufbau sind, eine neue Version geplant ist; Angabe von Begleitmaterial, Hinweis auf Unterreihen etc. (dieses Feld kann auch freigelassen werden).	SSG.Notes
91	Time Stamp	Datum, an dem das Formular über die Informationsquelle ausgefüllt wurde.	SSG. Timestamp
Interna			
93	Comment	Anmerkungen, Fragen etc. des Bearbeiters.	nur interne Information
94	Status	Daten durchgesehen und zur Veröffentlichung freigegeben, „stop“ verhindert den Export.	nur interne Information

5.2.2 Dublin Core Metadaten

Die Kategorien des SSG-FI-Projektes lehnen sich an die Kategorien von Dublin Core an und werden mit dem von Dublin Core vorgeschlagenen Mechanismus innerhalb der HTML-Dateien veröffentlicht. Daher muß sowohl die Semantik als auch die Syntax bei der Erstellung und dem Export der Daten beachtet werden. Die relevanten Informationen werden hier kurz zusammengefaßt.⁴⁵

Hauptquellen:

- Dublin Core Metadata Element Set: Reference Description
http://purl.org/DC/about/element_set.htm
- Dublin Core Sub-elements (Paul Miller)
<http://www.mailbase.ac.uk/lists/dc-subelements/files/helsinki.html>

⁴⁵ Die Homepage von Dublin Core findet sich unter der URL <http://purl.org/DC/>.

- Dublin Core Qualifiers/Substructure (Rebecca Guenther)
<http://www.loc.gov/marc/dcqualif.html>
- Tags for the Identification of Languages (H. Alvestrand)
<http://www.cis.ohio-state.edu/htbin/rfc/rfc1766.html>
- Dublin Core Element: Coverage (Ad Hoc Working Group: Coverage Element)
<http://www.sdc.ucsb.edu/~mary/coverage.htm>

Die Syntax von Dublin Core in HTML

Die Dublin Core-Syntax im Allgemeinen und zur Implementierung in HTML im Besonderen ist (noch) nicht festgelegt, es bilden sich aber Regeln heraus, an denen sich das SSG-FI-Projekt orientiert.

Für die Beschreibung von HTML-Dokumenten ist im Kopfbereich (Header) die sogenannte Meta-Marke (Meta Tag) vorgesehen. Sie hat die Form

```
< META NAME="Name" CONTENT="Inhalt">
```

es wird also der Name und Inhalt der Meta-Marke in Anführungszeichen angegeben. Dieser Mechanismus wird für die Einfügung von Dublin-Core-Informationen in HTML angewandt und zum Teil noch modifiziert.

Die Dublin Core Beschreibung besteht aus 15 Kategorien (elements), die jeweils Namen und Wert (contents) haben und durch zusätzliche Eigenschaften (qualifiers) verfeinert werden können. Der derzeitige Standard sind die drei sogenannten „ Canberra Qualifiers“ (nach DC4): SCHEME, TYPE und LANG (Schema, Typ und Sprache). Eine Sonderrolle nimmt TYPE ein: Diese Eigenschaft wird jetzt als Unterkategorie (Sub-element) bezeichnet und mit einer besonderen (Punkt-) Syntax angegeben. Die Eigenschaften SCHEME und LANG werden einfach zusätzlich mit Werten in Anführungszeichen in das Meta-Tag eingefügt.

Beispiele:

```
<META NAME="DC.Creator" LANG="fr" CONTENT="Arnaud Le Hors">
```

```
<META NAME="DC.Subject" SCHEME="DDC" CONTENT="541.34">
```

Die Eigenschaften ROLE und FLAG sind derzeit nicht mehr in der Diskussion.

Die Unterkategorien werden mit einem Punkt getrennt an den Namen der DC-Kategorie angehängt. Dabei sollen keine Leerzeichen, Bindestriche oder Unterstreichungen benutzt werden, sondern die einzelnen Wörter mit großen Anfangsbuchstaben direkt aneinander gehängt werden. Darüber, ob das erste Wort groß oder klein geschrieben wird, gibt es unterschiedliche Angaben: von Dublin Core wird klein empfohlen, in Paul Millers Papier groß, im SSG-FI-Projekt werden Großbuchstaben bevorzugt.

Beispiele:

```
DC.Creator.PersonalName
```

```
DC.Creator.PersonalName.Address
```


Damit ergeben sich insgesamt Ausdrücke der Art:

```
<META NAME="DC.Creator.PersonalName" SCHEME="LCNAF" CONTENT="Le Hors, Arnaud">
```

(LCNAF = Library of Congress Name Authority File)

Die Semantik von Dublin Core

Die hier gegebene Zusammenfassung beruht auf den oben angegebenen englischen Quellen und wurde nicht übersetzt.

DC.Title The name given to the resource by the creator or publisher

Sub-Element: .Alternative

DC.Creator The person or organization primarily responsible for creating the intellectual content of the resource. For example, authors in the case of written documents, artists, photographers, or illustrators in the case of visual resources.

Sub-Element: .PersonalName, .PersonalName.Address, .CorporateName, .CorporateName.Address

Scheme: LCNAF (Library of Congress Name Authority File)

DC.Subject The topic of the resource. Typically, subject will be expressed as keywords or phrases that describe the subject or content of the resource. The use of controlled vocabularies and formal classification schemes is encouraged.

Scheme: Unqualified: Keyword is default, LCSH (Library of Congress Subject Headings), MeSH (Medical Subject Headings, AAT (Art and Architecture Thesaurus), LCNAF (Library of Congress Name Authority File): for names used as subjects, DDC (Dewey Decimal Classification), LCC (Library of Congress Classification), NLM (National Library of Medicine Classification), UDC (Universal Decimal Classification)

DC.Description A textual description of the content of the resource, including abstracts in the case of document-like objects or content descriptions in the case of visual resources.

Scheme: Abstract is default, URL

DC.Publisher The entity responsible for making the resource available in its present form, such as a publishing house, a university department, or a corporate entity.

Sub-Element: As DC.Creator.

DC.Contributor A person or organization not specified in a creator element who has made significant intellectual contributions to the resource but whose contribution is secondary to any person or organization specified in a creator element (for example, editor, translator, and illustrator).

Sub-Element: As DC.Creator.

Scheme: LCNAF (Library of Congress Name Authority File) At DC4 it was decided that role was not needed. It has thus not been included here. If needed, it could be included as a local extension.

DC.Date The date the resource was made available in its present form. Recommended best practice is an 8 digit number in the form YYYY-MM-DD as defined in ISO 8601. In this scheme, the date element 1994-11-05 corresponds to November 5, 1994. Many other schema are possible, but if used, they should be identified in an unambiguous manner.

Scheme: (Default is ISO 8601), ANSI X3.30, IETF RFC 822, (Others?)

DC.Type The category of the resource, such as home page, novel, poem, working paper, technical report, essay, dictionary. For the sake of interoperability, type should be selected from an enumerated list that is under development in the workshop series at the time of publication of this document.

Scheme: list of types is planned, minimal: Text, Image, Sound, Software, Data, Interactive; for current thinking on the application of this element. see <http://sunsite.berkeley.edu/Metadata/types.html>

DC.Format The data format of the resource, used to identify the software and possibly hardware that might be needed to display or operate the resource. For the sake of interoperability, format should be selected from an enumerated list that is under development in the workshop series at the time of publication of this document.

Sub-Element: No sub-elements at present, although it was recommended that either an existing SCHEME (such as the Getty's Art & Architecture Thesaurus) or a new one be recommended in order to better control the use of this element.

Scheme: IMT (i.e. MIME), DCPMT (Dublin Core Physical Medium Type)

DC.Identifier String or number used to uniquely identify the resource. Examples for networked resources include URLs and URNs (when implemented). Other globally unique identifiers, such as International Standard Book Numbers (ISBN) or other formal names would also be candidates for this element in the case of off-line resources.

Scheme: URL is default, URN (Uniform Resource Name), ISBN (International Standard Book Number), ISSN (International Standard Serial Number), SICI (Serial Item and Contribution Identifier), FPI (Formal Public Identifier)

DC.Source A string or number used to uniquely identify the work from which this resource was derived, if applicable. For example, a PDF version of a novel might have a source element containing an ISBN number for the physical book from which the PDF version was derived.

Scheme: Free text is default, URL, URN, ISBN, ISSN

DC.Language Language(s) of the intellectual content of the resource. Where practical, the content of this field should coincide with RFC 1766.

Scheme: IETF RFC 1766, Z39.53, ISO 639-1, ISO 639-2/B (after final publication)

DC.Relation The relationship of this resource to other resources. The intent of this element is to provide a means to express relationships among resources that have formal relationships to others, but exist as discrete resources themselves. For example, images in a document, chapters in a book, or items in a collection. Formal specification of relation is currently under development. Users and developers should understand that use of this element is currently considered to be experimental.

Scheme: Default is free text, URL, URN, ISBN

DC.Coverage The spatial and/or temporal characteristics of the resource. Formal specification of coverage is currently under development. Users and developers should understand that use of this element is currently considered to be experimental

Sub-Element: DC.Coverage.PeriodName, DC.Coverage.PlaceName, DC.Coverage.t, DC.Coverage.x, DC.Coverage.y, DC.Coverage.z, DC.Coverage.Polygon, DC.Coverage.Line, DC.Coverage.3d

DC.Rights A link to a copyright notice, to a rights-management statement, or to a service that would provide information about terms of access to the resource. Formal specification of rights is currently under development. Users and developers should understand that use of this element is currently considered to be experimental.

Scheme: Default is free text, URL, URN

5.3 Die formale Erschließung

Unter den Kategorien zur Beschreibung einer Ressource gibt es neben bibliographischen Kategorien (Titel, Autor, Schlagwörter, Beschreibung etc.) und statistischen sowie Datumskategorien auch formale Kategorien, von denen der Formalschlüssel eine zentrale Rolle spielt. Dieser *Formal Key* ermöglicht das Blättern („Browse“-Funktion) im Source Type Catalog und steuert die Erzeugung der entsprechenden Ergebnislisten, dabei wird eventuell auch noch der Länderschlüssel benutzt. Die Abkürzungen des Länder- und Sprachcodes werden automatisch in die Angaben über das Land, in dem der beschriebene Server liegt und die Sprache bzw. Sprachen, in denen der Text vorliegt, übersetzt.

5.3.1 Der *Formal Key* und der *Source Type Catalog*

Nach dem derzeitigen Stand des Internet ist es für ein Fachinformationssystem eher ungewöhnlich, daß eine zweidimensionale „Browse“-Funktion angeboten wird. Die SSG-FI-Guides erlauben es, sowohl nach thematischen als auch nach formalen Gesichtspunkten zu blättern. Bei den meisten fachspezifischen Gateways im Internet wird zwar auch nach den verschiedenen Themenschwerpunkten wie Journals, Newsgroups etc. unterschieden, dies bedeutet aber nicht, daß eine Ressource auch in beiden Komplexen auftritt. Dies ist in den SSG-FI-Guides anders: Dort wird eine Ressource sowohl thematisch als auch nach formalen Kriterien, also nach Typ des Dokuments, jeweils mindestens einmal katalogisiert. Die Schwierigkeit, einen formalen Katalog zu erstellen, der nicht nur fachübergreifend, sondern auch für elektronische Ressourcen im Internet oder auf CD-ROM wie auch für gedruckte Materialien gilt, liegt auf der Hand. So kann die Kategorie Datenbank bei den Geisteswissenschaftlern etwas anderes bedeuten als bei den Naturwissenschaftlern. Einige Kategorien sind nur bei Online-Ressourcen sinnvoll, wie z.B. Museum, Bibliothek, Universität etc., andere dagegen beziehen sich hauptsächlich auf den gedruckten oder CD-ROM-Bereich, wie z.B. Current Bibliographic and Abstracting Series.

Trotz allem kann der Formalschlüssel als fachübergreifend und für alle Ressourcen geeignet betrachtet werden; er umfaßt über 40 Dokumenttypen. Der „Formal Key“ ist ein Code für Dokumenttypen („Source Type“), der auf der Verschlüsselung der Formschlagwörter aus Anlage 6 der RSWK (2. Auflage 1991)

basiert. Über den „Source Type Catalog“ des SSG-FI-Systems kann zu einem Ressourcentyp die Liste der vorliegenden Informationen eingesehen werden, bei den Infodateien werden die „Formal Keys“ als SSG.Formal-Metadaten, die Source Types im sichtbaren Bereich als Klassifikationen eingetragen.

Die einzelnen Arten sind in 4 verschiedene Blöcke zusammengefaßt:

- Information Providers/Servers: hier finden sich die verschiedenen Informationsanbieter, die bevorzugt als Online-Quellen formal kategorisiert werden,
- Factual Reference Works: die verschiedenen Sachauskunftsmittel,
- Bibliographic Sources: Literatúrauskunftsmittel,
- Primary Information Sources: primäre Informationsquellen.

No.	Source Type	Key	Description/ Examples
1.	Servers/ Information Providers		
1.1	Libraries	lb	Homepage von Bibliotheken
1.2	Archives	ar	Homepage von Archiven
1.3	Museums	mu	Homepage von Museen
1.41	University Departments	sf1	Institute, Fachbereiche
1.42	Research Institutes	sf2	Wissenschaftliche Forschungsinstitute
1.51	Organizations and Societies	so1	Fachgesellschaften, Fachverbände etc.
1.52	Official Governmental Server	so2	Landesämter, Surveys (Regierungsstellen)
1.6	Commercial Provider	sc	Verlage oder andere kommerzielle Anbieter
1.7	Communication Forums	dg	discussion oder news groups, mailing lists
1.8	Other Servers/ Information Providers	etc-info	Private Homepages etc.
2.	Factual Reference Works		
2.1	Encyclopedia and Biographic Lexicons	ej	Fachlexika
2.2	Dictionaries, Thesauri	we	Fachbezogene Übersetzungswerke
2.3	Lists of Abbreviations	ak	Abkürzungswerke
2.4	Gazetteers	oe	Ortsverzeichnisse
2.5	Addresses and Personnel Directories	ay	Adressenlisten von Universitäten, Organisationen oder Mitgliedern, E-Mail-Verzeichnisse
2.6	Data Collections	d6	Primärdatenbanken (Fakten, Statistik)
2.7	Compendiums of Laws	gs	Rechtsprechungen etc.

2.8	Norms and Standards	tm	Normen, ISO-/DIN-Vorschriften etc.
2.9	Manuals	a2	Technische oder methodische Anleitungen
2.10	(Current) Research Projects	fp	Forschungsprojekte und -programme
2.11	Other Factual Reference Work	etc-fact	Residualkategorie
3.	Bibliographic Sources		
3.1	Reference Books and Virtual Libraries	ff	Internet Gateways, Fachinformationsführer etc.
3.11	Special Subject Virtual Libraries	is	Spezielle Abhandlungen zu einem begrenzten Themenkomplex mit Linkliste etc.
3.12	Author-Specific Sites	au	Einem Autor oder einer Autorin gewidmete Sites.
3.2	Current Bibliographic and Abstracting Series	blz6	Große u. laufend ergänzte Bibliographien oder Rezensionen, Bücher, Auszüge etc.
3.3	Completed Bibliographies of Special Subjects	bl	Abgeschlossene Bibliographien
3.4	Bibliographies by Specific Document Type		Bibliographische Sammlungen von:
3.41	Lists of Periodicals/Journals	vcz6	Zeitschriften
3.42	Lists of Conference Proceedings	vckn	Konferenzen, Tagungen
3.43	Lists of Government Reports	vcbd	Regierungsveröffentlichungen
3.44	Lists of Research Reports	vcfd	Forschungsberichten
3.45	Lists of Theses and Dissertations	vcho	Dissertationen
3.46	Lists of Map Collections	vcka	Adressen von Kartensammlungen
3.47	Lists of Standard Specifications	vctm	DIN-/ISO-Normen etc.
3.48	Lists of Translation Indexes	vcue	Übersetzungsnachweisen
3.49	Other Lists of Specific Document Types	vcan	andere, z.B. Software, Datenbanken
3.5	Subject-Specific Library Catalogs	kb	Sammlung von fachspezifischen Bibliothekskatalogen
3.6	Other Bibliographic Sources	etc-bib	Residualkategorie

4.	Primary Information Sources		
4.1	Teaching Materials	le	Lehrbücher, Lehrmaterial, Vorlesungen
4.2	Handbooks	hb	Handbücher
4.3	Journals:	z6	Zeitungen und Zeitschriften:
4.31	with Free Full Articles	z61	mit kostenfreiem Volltextangebot
4.32	with Selected Articles	z62	mit teilweisem Volltextangebot
4.33	with Abstracts only	z63	mit Zusammenfassungen
4.34	with Table of Contents only	z64	nur mit Inhaltsverzeichnis
4.35	in Early Development Stage	z65	im Aufbaustadium
4.4	Series of Preprints, Reports, and Working Papers	fd	Sammlungen von (noch) nicht veröffentlichten Artikeln
4.5	Congresses	kn	Konferenzen, Tagungen, Symposien
4.6	Source Materials	q6	Quellenmaterial
4.61	Primary Text	q61	Primärquellen
4.62	Secondary Texts	q62	Sekundärliteratur
4.7	Maps	ka	Kartenmaterial
4.8	(Collections of) Images, Photos	bb	Photosammlungen
4.9	Moving Pictures	mp	Bewegte Bilder: Animationen, Videos, Filme
4.10	(Virtual) Exhibits	at	Virtuelle Ausstellungen, Kataloge etc.
4.11	Software and Expert Systems	do	fachspezifische Software, Expertensysteme (kopierbar oder zur Online-Nutzung)
4.12	Other Primary Information Sources	etc-prim	Residualkategorie

5.3.2 Ländercodes nach ISO 3166⁴⁶

Im ISO Standard 3166 sind Abkürzungen für Länder definiert, die jeweils aus zwei Großbuchstaben bestehen. Sie finden in HTML auch als Toplevel-Domain Verwendung.

Im SSG-FI-System werden zwei in ISO 3166 nicht vorgesehene Ländercodes benutzt, die z.B. für internationale Fachgesellschaften oder europaweite Projekte gebraucht werden: **II = International** und **EU = European Union**.

⁴⁶ Quelle: <http://www.chemie.fu-berlin.de/adressen/isocodes.html>

Ländercodes nach Ländern sortiert

Afghanistan	AF	Central African Republic	CF	French Southern Territories	TF
Albania	AL	Chad	TD	Gabon	GA
Algeria	DZ	Chile	CL	Gambia	GM
American Samoa	AS	China	CN	Georgia	GE
Andorra	AD	Christmas Island	CX	Germany	DE
Angola	AO	Cocos (Keeling) Islands	CC	Ghana	GH
Anguilla	AI	Colombia	CO	Gibraltar	GI
Antarctica	AQ	Comoros	KM	Greece	GR
Antigua And Barbuda	AG	Congo	CG	Greenland	GL
Argentina	AR	Cook Islands	CK	Grenada	GD
Armenia	AM	Costa Rica	CR	Guadeloupe	GP
Aruba	AW	Cote d'Ivoire	CI	Guam	GU
Australia	AU	Croatia (local name: Hrvatska)	HR	Guatemala	GT
Austria	AT	Cuba	CU	Guinea	GN
Azerbaijan	AZ	Cyprus	CY	Guinea-Bissau	GW
Bahamas	BS	Czech Republic	CZ	Guyana	GY
Bahrain	BH	Denmark	DK	Haiti	HT
Bangladesh	BD	Djibouti	DJ	Heard And Mc Donald Islands	HM
Barbados	BB	Dominica	DM	Honduras	HN
Belarus	BY	Dominican Republic	DO	Hong Kong	HK
Belgium	BE	East Timor	TP	Hungary	HU
Belize	BZ	Ecuador	EC	Iceland	IS
Benin	BJ	Egypt	EG	India	IN
Bermuda	BM	El Salvador	SV	Indonesia	ID
Bhutan	BT	Equatorial Guinea	GQ	International (SSG-FI)	II
Bolivia	BO	Eritrea	ER	Iran (Islamic Republic Of)	IR
Bosnia And Herzegovina	BA	Estonia	EE	Iraq	IQ
Botswana	BW	Ethiopia	ET	Ireland	IE
Bouvet Island	BV	European Union (SSG-FI)	EU	Israel	IL
Brazil	BR	Falkland Islands (Malvinas)	FK	Italy	IT
British Indian Ocean Territory	IO	Faeroe Islands	FO	Jamaica	JM
Brunei Darussalam	BN	Fiji	FJ	Japan	JP
Bulgaria	BG	Finland	FI	Jordan	JO
Burkina Faso	BF	France	FR	Kazakhstan	KZ
Burundi	BI	France, Metropolitan	FX	Kenya	KE
Cambodia	KH	French Guiana	GF	Kiribati	KI
Cameroon	CM	French Polynesia	PF	Korea, Democratic People's Republic Of	KP
Canada	CA				
Cape Verde	CV				
Cayman Islands	KY				

Korea, Republic Of	KR	New Caledonia	NC	Solomon Islands	SB
Kuwait	KW	New Zealand	NZ	Somalia	SO
Kyrgyzstan	KG	Nicaragua	NI	South Africa	ZA
Lao People's Democratic Republic	LA	Niger	NE	South Georgia And The South Sandwich Islands	GS
Latvia	LV	Nigeria	NG	Spain	ES
Lebanon	LB	Niue	NU	Sri Lanka	LK
Lesotho	LS	Norfolk Island	NF	St. Helena	SH
Liberia	LR	Northern Mariana Islands	MP	St. Pierre And Miquelon	PM
Libyan Arab Jamahiriya	LY	Norway	NO	Sudan	SD
Liechtenstein	LI	Oman	OM	Suriname	SR
Lithuania	LT	Pakistan	PK	Svalbard And Jan Mayen Islands	SJ
Luxembourg	LU	Palau	PW	Swaziland	SZ
Macau	MO	Panama	PA	Sweden	SE
Macedonia, The Former Yugoslav Republic Of	MK	Papua New Guinea	PG	Switzerland	CH
Madagascar	MG	Paraguay	PY	Syrian Arab Republic	SY
Malawi	MW	Peru	PE	Taiwan, Province Of China	TW
Malaysia	MY	Philippines	PH	Tajikistan	TJ
Maldives	MV	Pitcairn	PN	Tanzania, United Republic Of	TZ
Mali	ML	Poland	PL	Thailand	TH
Malta	MT	Portugal	PT	Togo	TG
Marshall Islands	MH	Puerto Rico	PR	Tokelau	TK
Martinique	MQ	Qatar	QA	Tonga	TO
Mauritania	MR	Reunion	RE	Trinidad And Tobago	TT
Mauritius	MU	Romania	RO	Tunisia	TN
Mayotte	YT	Russian Federation	RU	Turkey	TR
Mexico	MX	Rwanda	RW	Turkmenistan	TM
Micronesia, Federated States Of	FM	Saint Kitts And Nevis	KN	Turks And Caicos Islands	TC
Moldova, Republic Of	MD	Saint Lucia	LC	Tuvalu	TV
Monaco	MC	Saint Vincent And The Grenadines	VC	Uganda	UG
Mongolia	MN	Samoa	WS	Ukraine	UA
Montserrat	MS	San Marino	SM	United Arab Emirates	AE
Morocco	MA	Sao Tome And Principe	ST	United Kingdom (UK)	GB
Mozambique	MZ	Saudi Arabia	SA	United States	US
Myanmar	MM	Senegal	SN	United States Minor Outlying Islands	UM
Namibia	NA	Seychelles	SC	Uruguay	UY
Nauru	NR	Sierra Leone	SL		
Nepal	NP	Singapore	SG		
Netherlands	NL	Slovakia (Slovak Republic)	SK		
Netherlands Antilles	AN	Slovenia	SI		

Uzbekistan	UZ	Virgin Islands (British)	VG	Yugoslavia	YU
Vanuatu	VU	Virgin Islands (U.S.)	VI	Zaire	ZR
Vatican City State (Holy See)	VA	Wallis And Futuna Islands	WF	Zambia	ZM
Venezuela	VE	Western Sahara	EH	Zimbabwe	ZW
Viet Nam	VN	Yemen	YE		

Ländercodes nach Codes sortiert:

AD	Andorra	BZ	Belize	FI	Finland
AE	United Arab Emirates	CA	Canada	FJ	Fiji
AF	Afghanistan	CC	Cocos (Keeling) Islands	FK	Falkland Islands (Malvinas)
AG	Antigua And Barbuda	CF	Central African Republic	FM	Micronesia, Federated States Of
AI	Anguilla	CG	Congo	FO	Faroe Islands
AL	Albania	CH	Switzerland	FR	France
AM	Armenia	CI	Cote d'Ivoire	FX	France, Metropolitan
AN	Netherlands Antilles	CK	Cook Islands	GA	Gabon
AO	Angola	CL	Chile	GB	United Kingdom (UK)
AQ	Antarctica	CM	Cameroon	GD	Grenada
AR	Argentina	CN	China	GE	Georgia
AS	American Samoa	CO	Colombia	GF	French Guiana
AT	Austria	CR	Costa Rica	GH	Ghana
AU	Australia	CU	Cuba	GI	Gibraltar
AW	Aruba	CV	Cape Verde	GL	Greenland
AZ	Azerbaijan	CX	Christmas Island	GM	Gambia
BA	Bosnia And Herzegovina	CY	Cyprus	GN	Guinea
BB	Barbados	CZ	Czech Republic	GP	Guadeloupe
BD	Bangladesh	DE	Germany	GQ	Equatorial Guinea
BE	Belgium	DJ	Djibouti	GR	Greece
BF	Burkina Faso	DK	Denmark	GS	South Georgia And The South Sandwich Islands
BG	Bulgaria	DM	Dominica	GT	Guatemala
BH	Bahrain	DO	Dominican Republic	GU	Guam
BI	Burundi	DZ	Algeria	GW	Guinea-Bissau
BJ	Benin	EC	Ecuador	GY	Guyana
BM	Bermuda	EE	Estonia	HK	Hong Kong
BN	Brunei Darussalam	EG	Egypt	HM	Heard And Mc Donald Islands
BO	Bolivia	EH	Western Sahara	HN	Honduras
BR	Brazil	ER	Eritrea	HR	Croatia (local name: Hrvatska)
BS	Bahamas	ES	Spain		
BT	Bhutan	ET	Ethiopia		
BV	Bouvet Island	EU	European Union (SSG-FI)		
BW	Botswana				
BY	Belarus				

HT	Haiti	MC	Monaco		Miquelon
HU	Hungary	MD	Moldova, Republic Of	PN	Pitcairn
ID	Indonesia	MG	Madagascar	PR	Puerto Rico
IE	Ireland	MH	Marshall Islands	PT	Portugal
II	International (SSG-FI)	MK	Macedonia, The Former Yugoslav Republic Of	PW	Palau
IL	Israel			PY	Paraguay
IN	India	ML	Mali	QA	Qatar
IO	British Indian Ocean Territory	MM	Myanmar	RE	Reunion
IQ	Iraq	MN	Mongolia	RO	Romania
IR	Iran (Islamic Republic Of)	MO	Macau	RU	Russian Federation
IS	Iceland	MP	Northern Mariana Islands	RW	Rwanda
IT	Italy	MQ	Martinique	SA	Saudi Arabia
JM	Jamaica	MR	Mauritania	SB	Solomon Islands
JO	Jordan	MS	Montserrat	SC	Seychelles
JP	Japan	MT	Malta	SD	Sudan
KE	Kenya	MU	Mauritius	SE	Sweden
KG	Kyrgyzstan	MV	Maldives	SG	Singapore
KH	Cambodia	MW	Malawi	SH	St. Helena
KI	Kiribati	MX	Mexico	SI	Slovenia
KM	Comoros	MY	Malaysia	SJ	Svalbard And Jan Mayen Islands
KN	Saint Kitts And Nevis	MZ	Mozambique	SK	Slovakia (Slovak Republic)
KP	Korea, Democratic People's Republic Of	NA	Namibia	SL	Sierra Leone
KR	Korea, Republic Of	NC	New Caledonia	SM	San Marino
KW	Kuwait	NE	Niger	SN	Senegal
KY	Cayman Islands	NF	Norfolk Island	SO	Somalia
KZ	Kazakhstan	NG	Nigeria	SR	Suriname
LA	Lao People's Democratic Republic	NI	Nicaragua	ST	Sao Tome And Principe
LB	Lebanon	NL	Netherlands	SV	El Salvador
LC	Saint Lucia	NO	Norway	SY	Syrian Arab Republic
LI	Liechtenstein	NP	Nepal	SZ	Swaziland
LK	Sri Lanka	NR	Nauru	TC	Turks And Caicos Islands
LR	Liberia	NU	Niue	TD	Chad
LS	Lesotho	NZ	New Zealand	TF	French Southern Territories
LT	Lithuania	OM	Oman	TG	Togo
LU	Luxembourg	PA	Panama	TH	Thailand
LV	Latvia	PE	Peru	TJ	Tajikistan
LY	Libyan Arab Jamahiriya	PF	French Polynesia	TK	Tokelau
MA	Morocco	PG	Papua New Guinea	TM	Turkmenistan
		PH	Philippines	TN	Tunisia
		PK	Pakistan		
		PL	Poland		
		PM	St. Pierre And		

TO	Tonga	US	United States	WF	Wallis And Futuna
TP	East Timor	UY	Uruguay		Islands
TR	Turkey	UZ	Uzbekistan	WS	Samoa
TT	Trinidad And Tobago	VA	Vatican City State	YE	Yemen
TV	Tuvalu		(Holy See)	YT	Mayotte
TW	Taiwan, Province Of	VC	Saint Vincent And The	YU	Yugoslavia
	China		Grenadines		
TZ	Tanzania, United	VE	Venezuela	ZA	South Africa
	Republic Of	VG	Virgin Islands (British)	ZM	Zambia
UA	Ukraine	VI	Virgin Islands (U.S.)	ZR	Zaire
UG	Uganda	VN	Viet Nam	ZW	Zimbabwe
UM	United States Minor	VU	Vanuatu		
	Outlying Islands				

Außer den Ländercodes werden im Internet noch folgende Toplevel-Domains verwendet:

ARPA	Teil des ARPANET. Aus diesem militärisch ausgerichteten Netz ist das Internet entstanden. Dieser Name wird nicht mehr verwandt, funktioniert aber noch.
COM	Firmen, kommerzielle Organisationen
EDU	Forschungseinrichtungen, Universitäten (nur USA und Kanada)
GOV	Regierung und Regierungsstellen (nur USA)
INT	Internationale Organisationen
MIL	US-Militär
NATO	inzwischen durch .INT ersetzt
NET	Netzwerk-Administration und Provider
ORG	Sonstige nicht-kommerzielle Organisationen

In den meisten Fällen wird außerhalb der USA eine Toplevel-Domain mit ISO-Code gewählt, nur in Großbritannien wird „UK“ statt „GB“ benutzt. Weitere 6 Toplevel-Domains sind „im Gespräch“, aber noch nicht offiziell:

FIRM	Firmen, kommerzielle Organisationen
STORE	Firmen, die Waren zum Verkauf anbieten
WEB	Anbieter, die sich auf das WWW spezialisiert haben
ARTS	Anbieter, die sich auf Kultur im weitesten Sinne spezialisiert haben
REC	Anbieter, die sich auf Erholung und Unterhaltung im weitesten Sinne spezialisiert haben
INFO	Informationsdienste

5.3.3 Sprachcodes nach ISO 639-1⁴⁷

Im ISO Standard 639-1 sind Abkürzungen für Sprachen definiert, die jeweils aus zwei Kleinbuchstaben bestehen. Sie finden in HTML auch beim LANG-Attribut Verwendung.

Sprachcodes nach Ländern sortiert:

(Afan) Oromo	om	Fiji	fj	Kirundi	rn	Samoan	sm
Abkhazian	ab	Finnish	fi	Korean	ko	Sangro	sg
Afar	aa	French	fr	Kurdish	ku	Sanskrit	sa
Afrikaans	af	Frisian	fy	Laothian	lo	Scots Gaelic	gd
Albanian	sq	Galician	gl	Latin	la	Serbian	sr
Amharic	am	Georgian	ka	Latvian, Lettish	lv	Serbo-Croatian	sh
Arabic	ar	German	de	Lingala	ln	Sesotho	st
Armenian	hy	Greek	el	Lithuanian	lt	Setswana	tn
Assamese	as	Greenlandic	kl	Macedonian	mk	Shona	sn
Aymara	ay	Guarani	gn	Malagasy	mg	Sindhi	sd
Azerbaijani	az	Gujarati	gu	Malay	ms	Singhalese	si
Bashkir	ba	Hausa	ha	Malayalam	ml	Siswati	ss
Basque	eu	Hebrew (former iw)	he	Maltese	mt	Slovak	sk
Bengali	bn	Hindi	hi	Maori	mi	Slovenian	sl
Bhutani	dz	Hungarian	hu	Marathi	mr	Somali	so
Bihari	bh	Icelandic	is	Moldavian	mo	Spanish	es
Bislama	bi	Indonesian (former in)	id	Mongolian	mn	Sudanese	su
Breton	br	Interlingua	ia	Nauru	na	Swahili	sw
Bulgarian	bg	Interlingue	ie	Nepali	ne	Swedish	sv
Burmese	my	Inupiak	ik	Norwegian	no	Tagalog	tl
Byelorussian	be	Inuktitut (Eskimo)	iu	Occitan	oc	Tajik	tg
Cambodian	km	Irish	ga	Oriya	or	Tamil	ta
Catalan	ca	Italian	it	Pashto, Pushto	ps	Tatar	tt
Chinese	zh	Japanese	ja	Persian	fa	Tegulu	te
Corsican	co	Javanese	jw	Polish	pl	Thai	th
Croatian	hr	Kannada	kn	Portuguese	pt	Tibetan	bo
Czech	cs	Kashmiri	ks	Punjabi	pa	Tigrinya	ti
Danish	da	Kazakh	kk	Quechua	qu	Tonga	to
Dutch	nl	Kinyarwanda	rw	Rhaeto- Romance	rm	Tsonga	ts
English	en	Kirghiz	ky	Romanian	ro	Turkish	tr
Esperanto	eo			Russian	ru	Turkmen	tk
Estonian	et					Twi	tw
Faeroese	fo					Uigur	ug
						Ukrainian	uk
						Urdu	ur

47 Quellen: <http://medoc.informatik.tu-muenchen.de/HTML3/anhang/node1.html> mit Aktualisierung von 1989 nach <http://ds.internic.net/rfc/rfc1766.txt>

Uzbek	uz	Welch	cy	Yiddish (former ji)	yi	Zhuang	za
Vietnamese	vi	Wolof	wo	Yoruba	yo	Zulu	zu
Volapuk	vo	Xhosa	xh				

Sprachcodes nach Ländern sortiert:

aa	Afar	gl	Galician	mg	Malagasy	so	Somali
ab	Abkhazian	gn	Guarani	mi	Maori	sq	Albanian
af	Afrikaans	gu	Gujarati	mk	Macedonian	sr	Serbian
am	Amharic	ha	Hausa	ml	Malayalam	ss	Siswati
ar	Arabic	hi	Hindi	mn	Mongolian	st	Sesotho
as	Assamese	he	Hebrew	mo	Moldavian	su	Sudanese
ay	Aymara	hr	Croatian	mr	Marathi	sv	Swedish
az	Azerbaijani	hu	Hungarian	ms	Malay	sw	Swahili
ba	Bashkir	hy	Armenian	mt	Maltese	ta	Tamil
be	Byelorussian	ia	Interlingua	my	Burmese	te	Tegulu
bg	Bulgarian	id	Indonesian	na	Nauru	tg	Tajik
bh	Bihari	ie	Interlingue	ne	Nepali	th	Thai
bi	Bislama	ik	Inupiak	nl	Dutch	ti	Tigrinya
bn	Bengali	in	former Indonesian	no	Norwegian	tk	Turkmen
bo	Tibetan	is	Icelandic	oc	Occitan	tl	Tagalog
br	Breton	it	Italian	om	(Afan) Oromo	tn	Setswana
ca	Catalan	iu	Inuktitut (Eskimo)	or	Oriya	to	Tonga
co	Corsican	iw	former Hebrew	pa	Punjabi	tr	Turkish
cs	Czech	ja	Japanese	pl	Polish	ts	Tsonga
cy	Welch	ji	former Yiddish	ps	Pashto, Pushto	tt	Tatar
da	Danish	jw	Javanese	pt	Portuguese	tw	Twi
de	German	ka	Georgian	qu	Quechua	ug	Uigur
dz	Bhutani	kk	Kazakh	rm	Rhaeto- Romance	uk	Ukrainian
el	Greek	kl	Greenlandic	rn	Kirundi	ur	Urdu
en	English	km	Cambodian	ro	Romanian	uz	Uzbek
eo	Esperanto	kn	Kannada	ru	Russian	vi	Vietnamese
es	Spanish	ko	Korean	rw	Kinyarwanda	vo	Volapuk
et	Estonian	ks	Kashmiri	sa	Sanskrit	wo	Wolof
eu	Basque	ku	Kurdish	sd	Sindhi	xh	Xhosa
fa	Persian	ky	Kirghiz	sg	Sangro	yi	Yiddish
fi	Finnish	la	Latin	sh	Serbo-Croatian	yo	Yoruba
fj	Fiji	ln	Lingala	si	Singhalese	za	Zhuang
fo	Faeroese	lo	Laothian	sk	Slovak	zh	Chinese
fr	French	lt	Lithuanian	sl	Slovenian	zu	Zulu
fy	Frisian	lv	Latvian, Lettish	sm	Samoan		
ga	Irish			sn	Shona		
gd	Scots Gaelic						

5.4 Das Grundkonzept der sachlichen Erschließung

Die sachliche Erschließung soll die Informationsquelle in den wissenschaftlichen Kontext einordnen. Ein kurzer Text, Deskriptoren und Schlagwörter geben eine ausführliche inhaltliche Beschreibung.

Zunächst wird die Ressource von dem Fachreferenten oder -wissenschaftler durchgesehen. Stehen vom Herausgeber oder Autor einer Informationsquelle verfaßte Abstracts oder vergebene Schlagwörter zur Verfügung, sollte diese Fremdleistung übernommen werden, in den übrigen Fällen müssen Texte und Schlagwörter in Eigenleistung erstellt werden.

Um eine aussagekräftige Einordnung der Informationsquelle in den Kontext der jeweiligen Wissenschaftsdisziplin – und damit auch in den jeweiligen Subject Catalog des SSG-FI – zu ermöglichen, wird zusätzlich eine Sacherschließung nach anerkannten fachspezifischen Klassifikationen bzw. bibliothekarischen Universalklassifikationen und/oder lokal verwendeten Bibliotheksklassifikationen vergeben. Die Auswahl der zur Anwendung kommenden Klassifikationssysteme erfolgt fachspezifisch nach ihrer allgemeinen Akzeptanz in den betreffenden Wissenschaften. Sinnvollerweise sollte bei der Auswahl der jeweiligen Klassifikationssysteme auch der mögliche Anteil verfügbarer Fremdleistungen berücksichtigt werden. Generell können mehrere Klassifikationssysteme parallel angeboten werden.

5.4.1 Die Fachklassifikation

Als Hauptklassifikation ist für die einzelnen SSG-Fächer eine Klassifikationssystem zu wählen, das die betreffende Disziplin hinreichend hierarchisch erschließt. Die Notationsstruktur bildet die Grundlage zur Erzeugung der Ergebnislisten nach der fachlichen Gliederung im Subject Catalog. Der wichtigste Aspekt für diese Gliederung ist die Übersichtlichkeit für den Nutzer, d.h. sie darf nicht zu viele Einträge enthalten, um zu lange Listen zu vermeiden. Die hierarchische Gliederung sollte deshalb nicht zu tief gehen. Die Notation des betreffenden Klassifikationssystems wird in das Feld 41 und die zugeordnete Klassenbezeichnung in das Feld 40 der Allegro-Datenbank eingetragen. Bei Bedarf ist ein Mehrfacheintrag möglich, so daß die Anzeige einer Informationsquelle mit mehreren fachlichen Schwerpunkten unter den jeweiligen fachlichen Systemstellen möglich wird. Notation und Klassenbezeichnung werden im Datensatz angezeigt, aber lediglich die Klassenbezeichnung zur Strukturierung der fachlichen Gliederung im Subject Catalog benutzt.

Wird hierfür ein Klassifikationssystem ausgewählt, das keine internationale Verbreitung findet bzw. dessen Klassenbezeichnungen nicht in englischer Sprache vorliegen, sind die Klassenbezeichnungen nach einem normierten fachspezifischen Wortschatz (z.B. in den Geowissenschaften dem Multilingual Thesaurus of Geosciences) zu übersetzen.

5.4.2 Basisklassifikation und Dewey Decimal Classification

Neben der Hauptklassifikation können noch weitere Klassifikationssysteme verwendet werden. Unter dem Aspekt der Übernahme von Fremddaten (besonders für konventionelle Auskunftsmittel) werden in den verschiedenen Guides die Notationen der Basisklassifikation des GBV (BK)⁴⁸ und des international weit verbreiteten Klassifikationssystems der Dewey Decimal Classification (DDC) vergeben. Dazu werden die jeweiligen Notationen in die Felder 42 (DDC) und 43 (BK) der Allegro-Datenbank eingetragen. Sie erscheinen im Metadata-Template ohne Klassenbezeichnung. Diese Notationen dienen als zusätzliches Abfrageinstrument bei einer sachlichen Recherche in der lokalen Suchmaschine.

Die Basisklassifikation dient als klassifikatorische Komponente für die kooperative Sacherschließung im GBV und soll eine Grobrasterung aller Titel ermöglichen. Diese grobklassifikatorische Komponente ist fachbezogen und dient der orientierenden Fragestellung. Sie basiert auf einer überarbeiteten Version der *Nederlandse Basisclassificatie* und besteht aus 48 Hauptklassen, verteilt auf 5 Großgruppen. Alle Hauptklassen haben eine zweistellige Nummer, die weitere Unterteilung innerhalb einer Hauptklasse wird in der Notation durch eine zweite Gruppe von zwei Ziffern ausgedrückt. Beide Zifferngruppen werden durch einen Punkt getrennt, Beispiel: 38.00.

Die Dewey Decimal Classification ist ein weltweit anerkanntes amerikanisches System. Die Notationen bestehen aus Ziffern. Die Untergliederung der Sachgruppen erfolgt stets in Zehnergruppen. Der gesamte Wissensbereich ist in 10 Hauptgruppen aufgeteilt, die wiederum durch Hinzufügen einer zweiten Ziffer in 10 Abteilungen zweiter Ordnung zerlegt werden. Jede dieser 10 Abteilungen wird wieder in weitere 10 Abteilungen zerlegt usw. Dadurch kann jeder Begriff bis in die feinsten Einzelheiten zergliedert werden. Sie stellt somit eine feinklassifikatorische Komponente dar. Um die Übersichtlichkeit zu erhöhen, wird nach jeder 3. Ziffer ein Punkt gesetzt, Beispiel: 531.7. Die DDC hat vor allem in den Bereichen Naturwissenschaften und Technik internationale Verbreitung gefunden. Sie wird als Ordnungsschema in vielen Bibliographien, Referateblättern und Dokumentationsdiensten verwendet.⁴⁹

48 Vgl. die im GBV (Gemeinsamer Bibliotheksverbund der Länder Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Sachsen-Anhalt, Schleswig-Holstein und Thüringen) angewandte *Basisklassifikation für den Bibliotheksverbund Niedersachsen/ Sachsen-Anhalt/ Thüringen*, bearb. durch die Facharbeitsgruppe Sacherschließung, 1995

49 Vgl. neuerdings hierzu: *Klassifikationen für wissenschaftliche Bibliotheken. Analysen, Empfehlungen, Modelle* (=dbi-materialien, Bd. 175), Berlin 1998

5.4.3 Verbale sachliche Erschließung durch Schlagworte

Zur sachlichen Feinerschließung wird eine verbale Komponente in Form von Schlagwörtern angeboten, die in das Feld 19 eingetragen werden. Die verbale Komponente ist sachbezogen und dient der speziellen Fragestellung. Die Suchanfrage über die natürliche Sprache dürfte die gebräuchlichste Form der Benutzer sein, sich dem Informationsangebot zu nähern. Die Erfahrung hat aber gezeigt, daß nur ein normierter Wortschatz eine komplette Ergebnismenge garantiert, deshalb empfiehlt es sich, auf einen international verbreiteten Fachthesaurus zurückzugreifen. Natürlich müssen für die Recherche entsprechende Hilfefunktionen angeboten werden.

5.5 Die Auswahl- und Bewertungskriterien im Projekt SSG-FI

In der Auswahl von Ressourcen für einen Katalog liegt bereits eine erste (positive) Bewertung (vgl. Kap. 3.4). Im Rahmen von SSG-FI wurde eine Entscheidung für eine *zusätzliche* Bewertung der erfaßten Ressourcen getroffen (vgl. Kapitel 2.3), weil nicht alle katalogisierten Ressourcen in jeder Hinsicht und durchgängig qualitativ hochwertig sind. In einzelnen Fällen werden auch Informationsquellen katalogisiert und erschlossen, die bei einer strikten Auslegung der Kriterien nicht berücksichtigt würden, um den Fachinformationsführer für die jeweiligen Fachwissenschaftler so „rund“ wie möglich zu gestalten.

Das jeweilige SSG-FI Team evaluiert nicht die Forschungsinfrastruktur und ihre Institutionen, sondern Fachinformationen. Dies bedeutet, daß nicht die fachwissenschaftliche Bedeutung von Organisationen und Individuen *an sich* bewertet werden, sondern das inhaltliche Angebot der beschriebenen Ressource(n). Es können also durchaus Internet-Ressourcen erfaßt werden, die auf einer relativ unbekanntem und/oder „unbedeutenden“ Universität aufliegen, aber sehr wohl exzellente Fachinformationen liefern. Auch ist es keineswegs immer so, daß z.B. renommierte Fachgesellschaften über ihre Server hochwertige Fachinformationen bereitstellen.

5.5.1 Die einzelnen Kategorien der Bewertung





Das Team des SSG-FI Projekts hat sich dafür entschieden, jede katalogisierte Ressource nach verschiedenen Kriterien zu evaluieren. Damit soll unterschiedlichen Anforderungen, die fachwissenschaftliche Nutzer an elektronische Ressourcen stellen können, im einzelnen Rechnung getragen werden. Die Bewertungskategorien (die im folgenden weiter erläutert werden) sind

- Content Inhaltliche Relevanz,
- Clarity Übersichtlichkeit und Layout,
- Index Indizierung und
- Links Verweise auf weitere Ressourcen.

Dazu kommen als indirekte Kriterien:

- Level Zielgruppe der Ressource und
- Backlinks Anzahl der Verweise auf diese Ressource von anderen Quellen.

Für die ersten vier Aspekte sieht SSG-FI grundsätzlich folgende Bewertungsstufen vor:

	fair
	good
	excellent
	none (nur in Kategorie „Links“ möglich)

Content

Die inhaltliche Bedeutung der Ressource, ihr Informationsgehalt („Content“), ist die erste und vielleicht für die meisten Nutzer wichtigste Dimension der Bewertung und entscheidet in erster Linie über die Aufnahme der Ressource in den Guide. Alle anderen Aspekte, wie Übersichtlichkeit oder Indizierung/Suchbarkeit können ein Defizit in diesem Bereich nicht wirklich kompensieren. Unter „Content“ wird der Informationsgehalt des thematischen Site verstanden. Das bedeutet, SSG-FI versucht eine Aussage darüber zu treffen, wie solide und umfassend das gewählte Thema (bzw. der Gegenstand) abgehandelt ist.

Wichtig ist: *SSG-FI bewertet nicht die Relevanz des behandelten Themas für die jeweilige Disziplin, sondern lediglich, wie gut die Bereitsteller der Informationen ihrer (selbstgestellten) Aufgabe nachkommen.* Das heißt, es geht nicht darum, welchen Stellenwert einzelne Themen im Rahmen einer Fachwissenschaft haben oder ob eine bestimmte Thematik „zu speziell“ sein könnte. Wenn eine Aufnahme in den Katalog von SSG-FI erfolgt, dann geschieht die Evaluation des inhaltlichen Informationswerts nur im Rahmen dieser - wie auch immer speziellen - Thematik. Ein Beispiel: Ob „Religion, Society and Culture in Newfoundland and Labrador“ zum Gesamtverständnis der Geschichte und Politik des anglo-amerikanischen Kulturraumes zentral sind, mag dahingestellt sein. Wenn aber eine Ressource dieses Thema erschöpfend, kompetent und akkurat behandelt, dann erhält sie auch eine entsprechende Bewertung.

Die Angabe zu „Content“ ist auch abhängig von dem Typ der Ressource. Bei einer bibliographischen Datenbank ist der Inhalt abhängig von Anzahl und Qualität der Datensätze, bei einem Text- oder Preprint-Archiv eher von der Qualität und Aktualität der Texte.

Es ist zudem unmöglich, einen *lückenlosen* Überblick über jede beschriebene Ressource zu erhalten (vgl. dazu Abschnitt 3.4.). Vielmehr geht es darum, einen möglichst guten *Eindruck* von einer Ressource zu bekommen.

Aspekte, die dabei beachtet werden, sind:

- Reichweite (scope); erschöpfende Behandlung des Gegenstands
- Autorität (authority) der Bereitsteller der Informationen
- Validität (validity)
- Genauigkeit (accuracy)
- Aktualität (currency)
- Niveau (level)

All diese Kriterien fließen in die Evaluation des Informationsgehalts einer thematischen Site durch SSG-FI mit ein. Weil es sich aber um ein Zusammenspiel dieser Aspekte handelt, lassen sich die Bewertungsstufen von SSG-FI nicht in ein simples Schema fassen: Eine Aussage nach dem Muster „wenn, ... dann ...“ ist mithin schwierig, wenn nicht gar unmöglich. Das bedeutet: *Jede Bewertung im Feld „Content“ ist das Ergebnis einer Einzelfallabwägung aller genannten Aspekte für die beschriebene Ressource.* Die folgende Schematisierung ist also nur eine sehr grobe Richtlinie.

★/fair:	„ordentliche“ Behandlung des Themas
★★/good	qualitativ überdurchschnittliche Abhandlung
★★★/excellent	„erschöpfende“ Behandlung des Themas, außergewöhnliche Qualität

Clarity

Unter „Clarity“ versteht SSG-FI nicht inhaltliche Klarheit, sondern die Übersichtlichkeit einer Ressource, das Layout und die technische Ausgestaltung. Es gibt eine separate Bewertung, weil es sich um eine vom Inhalt zumindest theoretisch entkoppelte Dimension handelt. So ist vorstellbar, daß eine Ressource hervorragende Primärinformationen bietet, diese aber nur sehr schwer auffindbar sind. Unter dem Aspekt „Clarity“ geht es dabei weniger um Auffindbarkeit im technischen Sinne (vgl. dazu unten „Index“), sondern um gestalterische Umsetzung: Kommt das Informationsangebot einer intuitiven Nutzung entgegen?

Wenn auch Erwartungen und Verhalten der Nutzer unterschiedlich sind, so ist doch naheliegend, daß die meisten Fachwissenschaftler ein primäres Interesse an Fachinformationen haben. Gleichwohl dürfte ein zumindest sekundäres Interesse an einer „mundgerechten“ Aufbereitung dieser Informationen bestehen. Die Bewertung im Feld „Clarity“ soll Auskunft darüber geben, inwieweit das gelungen ist. Es geht also bei dieser Dimension der Evaluation um Aspekte, die in der Kommunikationswissenschaft unter dem Stichwort „web-usability“, d.h. (Be-) Nutzbarkeit des Informationsangebots diskutiert werden.⁵⁰

50 Vgl. Elisabeth Klaus, Ulrike Röttger, *Medium, Organisation, Nutzung: Bedingungen erfolgreicher Öffentlichkeitsarbeit im Internet*, in Irene Neverla (Hrsg.): *Das Netz-*

Prinzipiell favorisiert SSG-FI - wie die meisten verwandten Projekte - eine „simple is best“-Position, d.h. idealerweise sollten fachwissenschaftlich relevante Informationen nur so stark technisch und graphisch aufbereitet sein, wie zu ihrem Verständnis oder ihrer Nutzung sinnvoll ist. Als Faustregel gilt, daß ein Zuviel an *Style* die inhaltliche Substanz nicht in den Hintergrund drängen sollte. Anhaltspunkte sind:

- Übersichtlichkeit (Farben, Schriften)
- keine Überfrachtung mit Graphiken oder technischen Features (Java Applets, Skripte)
- Eignung für verschiedene - auch rein textbasierte - Browser
- Vorhandensein von ALT-Tags etc.
- Einsatz von Java etc. in vertretbarem Rahmen

Noch stärker als im Feld „Content“ birgt die Bewertung der Übersichtlichkeit allerdings die Gefahr der Subjektivität des Urteilenden; denn was dem einen Nutzer als hilfreich erscheinen mag, kann für den anderen unsinnige Ablenkung vom Hauptsächlichen sein.

★/fair: übersichtliche Gestaltung

★★/good sehr übersichtliche Gestaltung; klare Gliederung

★★★/excellent optimales Layout: einer intuitiven Nutzung entgegenkommend

Index

Mit einer Bewertung der Indizierung einer Ressource wird eine Angabe darüber gemacht, wie gut ihre Inhalte strukturiert, auffindbar und/oder suchbar sind. Eine Internet-Ressource kann beispielsweise inhaltlich sehr gut und relevant sein und diese Inhalte auch für Benutzer in ansprechender und scheinbar übersichtlicher Form zur Verfügung stellen. Gleichwohl kann es sein, daß ein Nutzer dennoch keinen Überblick darüber gewinnt, welche Informationen „in den Tiefen“ dieser Ressource schlummern. Im Internet heißt das etwa, daß in Unterverzeichnissen relevante Informationen gelagert sein könnten, ohne daß diese direkt suchbar wären. Um ein Bild zu bemühen: Eine Ressource kann eine Goldmine sein, bei der allerdings das Gold nicht gleich dem Benutzer entgegen schimmert. Durch die Bewertung dieses Aspekts ist es also dem Nutzer möglich zu sehen, wieviel Zeit und Arbeit er investieren muß, bis sich ihm das volle Potential der Ressource erschließt.

Die Idee der Indizierung ist natürlich nicht internetspezifisch. Aus dem gedruckten Bereich sind Inhaltsübersichten und Register (etwa Personen-, Sach- oder Ortsregister) bereits bekannt. Sie erlauben auch dort den gezielte Zugriff

auf Informationen. Ohne ein solches (oder mehrere) Register ist die Ressource vergleichsweise weniger wertvoll, weil der Zeitaufwand im Verhältnis zum Informationsgewinn erheblich steigt.

Wie im gedruckten Bereich schwankt auch im Online-Bereich die Qualität der Indizierung. Es finden sich einerseits Ressourcen ohne jegliche Übersicht und andererseits solche mit Inhaltsverzeichnis oder Register, vielleicht auch einem Site Index oder einer Site Map. Darüber hinaus bietet sich im Bereich elektronischer Ressourcen allerdings eine erweiterte Möglichkeit der Suche, die in dieser Form bei gedruckten Materialien nicht besteht: die lokale Suchmaschine. Mit einer solchen kann typischerweise nach bestimmten Begriffen und/oder Schlagworten gesucht werden, was die Auffindbarkeit von Informationen selbst gegenüber einem guten Register noch deutlich erhöhen kann. Eine solche lokale Suchoption stellt - weil prinzipiell technisch immer möglich - momentan nach Maßstäben von SSG-FI die beste Form einer Indizierung dar. Allerdings sind solche Kriterien bei der Bewertung dieser Kategorie nicht bei allen Dokumententypen maßgeblich entscheidend. Eine umfangreiche Datensammlung z.B. sollte, um sehr gut indiziert zu sein, über eine Suchfunktion verfügen. Bei Sites die allgemeine Information anbieten, wie z.B. Fachgesellschaften, muß dies nicht zwingend sein. Gute Sprungmarken, Back-Funktionen etc. sind in diesem Beispiel eher das Bewertungsmaß.

Es ergeben sich somit die folgenden Bewertungsstufen für den Bereich „Index“:

- ★/fair: keinerlei Indizierung
- ★★/good Register („Site Index“, „Site Map“)
- ★★★/excellent lokale Suchmaschine

Links

Unter „Links“ werden Verweise auf weitere Ressourcen verstanden. Damit wird eine Angabe darüber gemacht, inwieweit diese Ressource zugleich Zugang zum Gegenstand und eine Übersicht über weitere, thematisch verwandte Informationsquellen beinhaltet. Zwar gehört eine Bibliographie oder Aufstellung der genutzten Materialien zu jeder wissenschaftlichen Arbeit dazu. Doch werden dabei typischerweise nur jene Materialien aufgenommen, die für die eigene Arbeit herangezogen wurden; es wird normalerweise nicht der Versuch unternommen, einen Überblick über die gesamte Forschungs- oder Quellenlage in diesem Bereich zu geben (etwa in Form eines bibliographischen Essays).

Weil das Internet aber *Hyperlinks* geradezu fordert und sie typischerweise „dazu gehören“, wird im SSG-FI dieser Aspekt auch bei jenen Ressourcen bewertet, bei denen Verweise in Form von Links von der Natur der Informationen her zunächst verzichtbar erscheinen. Ein Beispiel: eine im Internet verfügbare fachwissenschaftliche Zeitschrift bietet an sich bereits Fachinformationen, ein Verweis auf thematisch ähnliche Informationsangebote im Internet wird eigen-

lich nicht als typische Serviceleistung einer Fachzeitschrift erwartet. Gleichwohl lassen sich aber auch in gedruckten Zeitschriften oft Hinweise auf weitere Fachinformationen finden (in der Form von Anzeigen oder Kurzrezensionen). Deshalb wird auch hier das (Nicht-)Vorhandensein solcher Verweise mitbewertet. Denn das „Einlinken“ weiterer Sites ist oft - wenn auch nicht immer - ein Indiz dafür, wieweit sich eine Ressource als „Einfallstor“ (Subject Gateway) zu einem bestimmten Thema eignet.

Zu den Faktoren, die in die Bewertung nach dem Kriterium „Links“ Eingang finden, zählen:

- Anzahl der Verweise
- „Güte“ der genannten Quellen
- Vorhandensein von Kommentaren (Annotationen)
- Anordnung (sinnvolle Strukturierung)
- Aktualität (Funktionalität der Verweise)

Daraus ergeben sich die folgenden Bewertungsstufen in der Kategorie „Links“:

—/none	keine Hyperlinks
★/fair:	einige wenige, typischerweise nicht annotierte Links
★★/good	etliche z.T. geordnete und/oder annotierte Links
★★★/excellent	viele, gut strukturierte und/oder annotierte Links

Level

Mit der Kategorie „Level“ wird der Versuch unternommen, die *Zielgruppe* einer Informationsquelle anzugeben bzw. zu bestimmen. Ausgehend von der obigen Überlegung, daß im Internet keine etablierte Verlags- und Veröffentlichungsstruktur besteht, ist es vielfach für Nutzer schwer einzuschätzen, ob eine bestimmte Ressource Informationen auf dem gewünschten oder erforderlichen Niveau bereitstellt.

SSG-FI hat vier Ausprägungen für diese Kategorie definiert:

popular:	populärwissenschaftlich, d.h. ohne fachliche Vorkenntnisse verständlich
undergraduate:	mit wissenschaftlichem Anspruch, aber eher auf Grundstudiums-Niveau
graduate:	mit wissenschaftlichem Anspruch, auf Hauptstudiums-Niveau
professional:	gerichtet an und geeignet für Fachwissenschaftler in diesem Bereich

Die Angabe einer oder mehrerer Zielgruppe(n) birgt allerdings einige Fragen. Zum einen ist zu klären, wie vorzugehen ist, wenn Verfasser, Herausgeber oder Verleger einer Ressource selbst eine Angabe hierzu machen. Eine solche Vor-

gabe ist aus gedruckten Medien nicht in derselben Form bekannt. Zwar sind Lehrbücher zumeist als solche deklariert, doch findet sich bei „normalen“ Monographien typischerweise keine bibliographische Angabe über die Leserschaft, für die das Werk gedacht oder geeignet ist. Im Bereich gedruckter Informationen erfährt man das in aller Regel durch etablierte Rezensionsorgane, die allerdings - wie bereits mehrfach betont - in dieser Form im Internet (noch) nicht vorhanden sind.

Eine weitere Frage ist, wie man Angaben zur Zielgruppe für die Benutzer unmißverständlich gestalten kann. So ist es einerseits möglich, die Zielgruppe möglichst eng zu definieren; andererseits kann die Angabe etwas breiter gefaßt werden, um alle Benutzergruppen einzuschließen, für welche die Ressource geeignet ist. SSG-FI hat sich für einen Mittelweg entschieden. Damit stellte sich zwangsläufig die Frage, wie die Abgrenzung dann erfolgen sollte („von oben“ oder „von unten“) - mit jeweils unterschiedlichen Konsequenzen. Eine Abgrenzung „von oben“ (nach dem Motto „gerade noch für *undergraduates* verständlich“) würde immer die Angabe *professional* beinhalten müssen. Nun mag eine Ressource für einen Experten *verständlich* sein; ob sie aber das im Regelfall weitaus differenzierte oder speziellere Erkenntnisinteresse bedient, ist fraglich. Als Konsequenz daraus ergab sich für SSG-FI als Aufgabe die möglichst genaue Einschränkung der Zielgruppe.

Backlinks

SSG-FI ermittelt weiterhin sog. „backlinks“ für alle katalogisierten Internet-Ressourcen. Darunter ist die Anzahl von Internet-Ressourcen zu verstehen, die auf die beschriebene Ressource verweisen. Es werden allerdings nur diejenigen Verweise gezählt, die sich *auf anderen Servern* befinden. „Backlinks“ soll also nicht Auskunft darüber geben, wie gut eine Ressource in ihren internen Verknüpfungen ist, sondern darüber, inwieweit die „Internet Community“ sie (qua Hyperlink) empfiehlt. Die Kategorie ist also einem „Citation Index“ im Bereich gedruckter Werke durchaus vergleichbar.

Die Anzahl der Backlinks ermittelt SSG-FI nach einem einheitlichen Schema mit Hilfe der Suchfunktion von AltaVista (<http://www.altavista.com>). Sie wird als „advanced query“ mit der folgenden Syntax gestartet:

Struktur: `link:URL der Ressource ohne Angabe des Protokolls AND NOT host:Servername`

Beispiel: `link:www.sub.uni-goettingen.de/ssgfi AND NOT host:www.sub.uni-goettingen.de`

Die Anzahl der Treffer wird im Regelfall auf- bzw. abgerundet.

Eine Zusammenfassung aller Bewertungskriterien gibt Tab. 5.1.

Kategorie	excellent ★★★	good ★★	fair ★
Content	„erschöpfende“ Behandlung des Themas; außergewöhnliche Qualität	qualitativ überdurchschnittliche Abhandlung	„ordentliche“ Behandlung des Themas
Clarity	optimales Layout: einer intuitiven Nutzung entgegenkommend	sehr übersichtliche Gestaltung; klare Gliederung	übersichtliche Gestaltung
Index	lokale Suchmaschine	Register („Site Index“)	kein Index
Links	viele, gut strukturierte und/oder annotierte Links	etliche (ggf.) geordnete und/oder annotierte Links	wenige, typischerweise nicht annotierte Links (Eingabe „-“ / none, wenn keine Hyperlinks)

Tab. 5.1: Evaluationsschema SSG-FI

6 Die technische Implementierung des SSG-Fachinformationssystems

Von Thomas Fischer

Die SSG-FI-Daten werden in einer Datenbank gesammelt und über Internet allgemein zugänglich gemacht. Das technische System des SSG-FI-Projektes besteht aus zwei Bereichen: dem Bereich der Datensammlung und -verwaltung und dem Bereich der Datenpräsentation. Während die Verwaltung der Daten mit einer Allegro-Datenbank geschieht (Kap. 6.1), wird zur Präsentation ein Web-Server benötigt, der die produzierten HTML-Dateien allgemein zugänglich macht (Kap. 6.2). Das ist prinzipiell mit einem Server unter einem beliebigen Betriebssystem möglich, an der SUB Göttingen wird dazu ein LINUX-Server benutzt. Für den Export aus der Allegro-Datenbank in den Internet-Bereich ist wiederum Allegro zuständig (Kap.6.3).

Für den Aufbau eines Fachinformationsführers nach dem SSG-FI-Muster werden Dateien und Dateistrukturen benötigt, die im folgenden erläutert werden. Zuerst werden die Dateien beschrieben, die für den Aufbau der Allegro-Datenbank notwendig sind, dann folgt die Beschreibung der Serverstruktur und zuletzt die der Dateien, die für den Export der Allegro-Daten benutzt werden, da in diesen Bezug auf die Serverstruktur genommen wird.

6.1 Datensammlung und -verwaltung: das Allegro-System

In der Beschreibung des Allegro-Systems des SSG-FI-Projektes wird vorausgesetzt, daß ein funktionierendes *Allegro 15²-System* vorliegt, das heißt, daß zumindest die Programme ACP, PRESTO, SRCH, IMPORT, INDEX, QRIX, ASORT und UPDATE sowie die zugehörigen Dateien in dem *Programmordner*⁵¹ (= Programmpfad) installiert sind. In diesem Ordner wird einerseits der *SSGFI-Ordner* eingerichtet, der die allgemeinen Dateien und Programme („Skripte“) des SSG-FI-Systems enthält, und andererseits der *Datenordner* (= Datenbankpfad), in dem die für ein Gebiet spezifischen Dateien liegen. Diese Einrichtung ist so gestaltet, daß das vorhandene Allegro-System nicht verändert werden muß: Alle SSG-FI-Funktionen sind in diesen beiden Ordnern gekapselt und beeinflussen andere Allegro-Nutzungen nicht. Die Einrichtung erfolgt typischerweise für ein spezielles (Sondersammel-) Gebiet, der Name des Datenordners sollte auf dieses Gebiet verweisen. Da er aber auch in verschiedenen DOS-

51 Im SSG-FI-Projekt stellte sich heraus, daß das DOS-System mit langen Pfadnamen Schwierigkeiten hat, wie sie insbesondere bei gemeinsamer Nutzung über ein Netz entstehen. Die Pfade werden als Parameter mit übergeben, deren Länge unter DOS begrenzt ist. Deshalb wurde mit der Funktion „Netzlaufwerk verbinden“ der Programmpfad als Laufwerk „F:“ eingerichtet.

Dateinamen auftaucht, empfiehlt es sich, eine Kurzform zu wählen; in Göttingen wird zum Beispiel Geo für die Geowissenschaften und Math für die Mathematik benutzt. Im folgenden wird dafür XYZ eingesetzt, damit klar wird, welche Dateien wie anzupassen sind, der Datenordner heißt dann SSG-XYZ.

Die beschriebene Einrichtung erlaubt es, gleichzeitig verschiedene Gebiete effektiv zu verwalten. Wenn nur ein einziges Gebiet bearbeitet und nur eine Allegro-Datenbank aufgebaut werden soll, so können der SSGFI-Ordner und der Datenordner identisch sein, wenn die zugehörigen Verweise entsprechend angepasst werden.

Eine veränderliche Größe ist die benutzte fachliche Klassifikation, die möglichst einem internationalen Standard entsprechen und umfangreich genug sein sollte, um die verschiedenen Informationstypen erfassen zu können. In Göttingen wird dazu im Mathematikbereich die *Mathematics Subject Classification* von 1991 benutzt (MSC 91), für den geowissenschaftlichen Bereich und den Anglo-Amerikanischen Kulturraum die *Göttinger Online Klassifikation* (GOK). Im folgenden steht das Kürzel KAT für die Bezeichnung einer generischen Primärklassifikation, die jeweils an die lokalen und fachlichen Bedingungen angepasst werden muß. Im SSG-FI-System sind außerdem die Klassifikation nach Dewey (*Dewey Decimal Classification, DDC*) und die *Basisklassifikation* (BK) des Gemeinsamen Bibliotheksverbundes (GBV) vorgesehen; beide könnten auch als Primärklassifikation eingesetzt werden, wenn die benutzten Dateien entsprechend angepasst werden. Die Primärklassifikation ist von zentraler Bedeutung, da sie über den Subject Catalog zum Einstieg in den SSG-FI-Bereich genutzt wird.

Als Zeichensatz wird vom SSG-FI-System der MS-DOS-Zeichensatz verwandt, der für die Internet-Darstellung in den entsprechenden HTML-Code umgesetzt wird und als Latin-1 von allen Browsern auf allen Computersystemen dargestellt werden kann. Wenn viele Informationen aus Ländern vorliegen, deren Buchstaben mit Latin-1 nicht erfaßt werden, müßte allerdings sowohl die Allegro-Datenbank als auch die Umsetzung verändert werden. Außerdem führt diese Einstellung dazu, daß bei den benutzten allgemeinen Allegro-Dateien, insbesondere der Umwandlungsdatei S1.ASP (von SRCH benutzt) auf die Konsistenz der benutzten Zeichensätze zu achten ist: da das SSG-FI-System den MS-DOS-Zeichensatz benutzt, müssen diese Dateien darauf umgestellt werden.

6.1.1 Die Dateien des Datenordners SSG-XYZ

Die Allegro-Dateien sind Textdateien mit dem MS-DOS-Zeichensatz. Sie können mit jedem Texteditor bearbeitet werden, nur bei Sonderzeichen (Umlaute etc.) kann leicht Unvorhergesehenes passieren, das hat aber keine Auswirkungen auf das Funktionieren des Systems, sondern nur auf die Anzeige. Die wich-

tigsten Dateien sind die *Konfigurationsdatei*, die die Kategorien der Datenbank festlegt, und die *Indexdatei*, die die Indizierung steuert. Wie im MS-DOS-System üblich, wird der Typ der Datei durch eine dreibuchstabile Endung festgelegt, bei der nicht zwischen Groß- und Kleinbuchstaben unterschieden wird. Der Name der Dateien ist frei wählbar (im Rahmen der 8.3-Konvention: bis zu acht Buchstaben, keine Sonder- oder Leerzeichen), im SSG-FI Projekt sind Namen gebräuchlich, die auf das bearbeitete Gebiet verweisen. Nach Allegro-Konvention gilt der *erste Buchstabe des Namens* der Konfigurationsdatei als Kennbuchstabe und tritt bei allen zugehörigen *Dateityp-Endungen* als erster Buchstabe auf⁵². Dieser Buchstabe ist bei der SSG-FI Konfiguration **G**. Bevor die Dateien im einzelnen kurz beschrieben werden, werden sie kurz aufgelistet:

die Konfigurationsdatei	G XYZ.CFG
die Indexdatei	XYZ.GPI
das Aufrufprogramm	XYZDAT.BAT

Für spezielle Anpassungen gibt es eventuell noch die folgenden Dateien:

die spezielle Anzeigetabelle	ANZEIGE.GPT
die spezielle Zwischenteiltabelle	Z-SPEZ.GPT

Mit diesen Dateien wird die Datenbank XYZ von Allegro aufgebaut und verwaltet. Dazu legt Allegro die folgenden Dateien an:

die Datenbankdatei	XYZ_1.GLD
die Satztable	XYZ.TBL
die Indexdatei	XYZ.GDX
das Kurztitelregister	XYZ.STL
die Logdatei	XYZ.LOG

Diese werden nicht einzeln angesprochen, sondern von Allegro als eine Einheit behandelt. Zentral ist die Datenbankdatei XYZ_1.GLD, in der die eingetragenen Datensätze gespeichert sind, aus dieser können Index, Satztable und Kurztitelregister jederzeit (und in Sekundenschnelle) wieder neu erzeugt werden. Nur wenn XYZ_1.GLD versehentlich zerstört oder gelöscht wurde, kann aus der *Logdatei* und einer *Sicherungsversion von XYZ_1.GLD* (die vom Benutzer angelegt werden muß!) die Datenbank wieder regeneriert werden.

Die Einträge eines Datensatzes werden in der üblichen Anzeige entweder mit Namen oder Nummer der Kategorie untereinander angegeben und können so

⁵² Das führt dazu, daß die Dateitypen, die unter Windows angezeigt werden, normalerweise so nicht erkannt werden und einzeln eingerichtet werden müssen. In Kapitel 6.1.2 wird eine Liste der benutzten Typen gegeben.

auch bearbeitet werden. Zum schnellen Auffinden einzelner Einträge sind die *Register* da: In jedem Register werden die Einträge einer oder mehrerer Kategorien gesammelt und sortiert angezeigt. Dadurch kann leicht überblickt werden, ob es zu einem Wert (z.B. einer URL) schon Einträge gibt oder ob die Einträge in einem Bereich formal richtig sind.

Die Register erlauben auch einen anderen Blick auf die Datenbank: Das gewählte Register bestimmt die Reihenfolge der Dokumente beim Durchblättern der Datenbank, so daß z.B. durch Wahl des Registers 9 (Länder) alle Dokumente aus Deutschland hintereinander angezeigt werden können. Zusätzlich werden einige Register zur Übersetzung von Abkürzungen und Schlüsselwörtern benutzt, um Sprachen und Länder mit Namen statt ISO-Code anzuzeigen. In der im SSG-FI-Projekt benutzten Konfiguration (XYZ.GPI, s.u.) werden die Register wie folgt belegt (das kann aber an die eigenen Bedürfnisse angepaßt werden):

Nr.	Inhalt
1	Titel
2	Formalschlüssel
3	Klassifikation
4	URL
5	Veröffentlichungsdatum
6	Erfassungsdatum; Dekodierung Formal- und Themenschlüssel
7	Status, Niveau, Kommentar; Dekodierung Länder- und Sprachcode
8	Sprache
9	Land
10	Nummer
11	Typ, Zeitrahmen

Die Einstellungsdateien selbst sind mit umfangreichen Kommentaren versehen, die die Funktion und Bedeutung einzelner Einträge erklären. Die Dateien müssen für den jeweiligen Zweck angepaßt werden, die entsprechenden Stellen sind mit „Anpassen:“ markiert und mit einer Erläuterung versehen. Die einzelnen Dateien haben die folgende Bedeutung:

GXYZ.CFG:

In der Konfigurationsdatei werden die Kategorien des SSG-FI-Systems festgelegt: Welche zulässig sind, wie sie heißen, wie sie abgefragt werden und welche als notwendig gelten. Die Datei besteht zunächst aus einem einleitenden Abschnitt, in dem einige Allegro-Grundeinstellungen festgelegt werden, der Liste der erlaubten Kategorien und einigen Konfigurationsbefehlen. Schließlich folgt die recht umfangreiche Abfrageliste, die durchgegangen

wird, wenn ein neuer Datensatz aufgenommen wird. Hier legt ein „!“ an der ersten Stelle nach der Kategorienummer eine zwingende Eingabe fest, außerdem können Standardvorgaben eingetragen werden, z.B. „en“ für Sprache und „US“ für Land. Am Anfang steht eine Abfrage nach dem Typ der Informationsquelle, dadurch können einige Einstellungen festgelegt werden. Nach diesem speziellen Teil springt das Programm in den allgemeinen Teil der Liste. Diese Liste kann und sollte den Bedürfnissen des jeweiligen Bereiches angepaßt werden.

Anpassen ist die Bezeichnung der primären Klassifikation KAT, eventuell können auch Voreinstellungen anders gesetzt werden.

XYZ.GPI:

Die Indexparameterdatei legt fest, welche Register wie angelegt werden: Der Name der Überschriften der elf möglichen Register und ihr Inhalt wird hier angegeben. Die Datei kann im Prinzip nach Belieben verändert werden, aber die Exportprogramme (s.u.) gehen davon aus, daß in *Register 7* die Übersetzung des Länder- und des Sprachcodes steht und in *Register 6* die des Formal- und des Themenschlüssels. Die Datei enthält außerdem einen „Hilfsabschnitt zur Validierung“, der dazu dient, beim Einfügen von Text in die Kategorien 19 und 20 der Allegro-Datenbank etwaige auftretende Zeilenwechsel abzufangen.

Anpassungen sind nicht notwendig.

ANZEIGE.GPT

Diese Tabelle enthält Zwischenteile, die von der Anzeige D-1.GPR benutzt werden; derzeit stehen darin nur die Namen der fachspezifischen Kategorien.

Z-SPEZ.GPT

Diese Tabelle enthält Zwischenteile, die für den Export der Datenbank benutzt werden. In der derzeitigen Einstellung sind es folgende:

- der Namen des Fachprojektes (z.B. GeoGuide),
- die Namen der fachspezifischen Kategorien (z.B. GOKverbal),
- die Hintergrundfarbe der HTML-Dateien,
- die fachspezifischen Präfixe der HTML-Dateien.

XYZDAT.BAT

Dies ist das Programm, das Allegro für die jeweilige Datenbank startet. In dieser Datei werden die verschiedenen Parameter (Konfiguration, Datenbankname, Daten- und Programmordner etc.) gesetzt und das Allegro-Cockpit aufgerufen.

Von Allegro werden bei verschiedenen Aktionen weitere Dateien in diesem Ordner eingerichtet: CCC.BAT, CCC.ALT, PROTOK, PROTOQ. Dabei steht in CCC.BAT der aktuelle Programmaufruf, in CCC.ALT werden alle Programmauf-

rufe gespeichert, und die anderen Dateien sind Allegro-übliche Protokolldateien der Indizierung. Außerdem liegt hier noch die Phrasendatei Phrase.apf.

6.1.2 Die Dateien des Ordners SSGFI

Während der Datenordner die Datenbank(-dateien), die Konfigurationsdateien und die Batchdatei *XYZdat.bat* zum Aufruf der Datenbank enthält, liegen alle weiteren benutzten Dateien und Programme im Ordner **SSGFI**. Damit bleibt das System überschaubar und die allgemeinen Dateien des SSG-FI-Systems sind von den speziellen Dateien des jeweiligen Faches getrennt. Für diese Vereinheitlichung werden zwei Mechanismen genutzt, die bei einer Anpassung des Systems an andere Bereiche beachtet werden müssen:

1. Verschiedene Konstanten (=Zwischenteile) werden in speziellen Dateien (Z-SPEZ.GPT, ANZEIGE.GPT, s.o.) angegeben, damit nur diese geändert werden müssen.
2. In den einzelnen Export-Parameterdateien (s.u., vor allem 6.2) steht der Name der Datenbank in der Variablen <fb> zur Verfügung. Ein Vergleich mit dem Inhalt dieser Variablen, z.B.

```
#u fb +G c"Geo",
```

erlaubt eine Verzweigung des Programmes (im Beispiel Sprung nach #-G) und so die spezielle Bearbeitung fachspezifischer Anforderungen.

Die SSGFI-Dateien dienen verschiedenen Funktionen:

Anzeige und Voreinstellungen

D-1.GPR:

Die Bildschirmanzeigedatei enthält die Namen der SSG-FI-Kategorien und kann individuell angepaßt werden. Wenn mehrere Datenbanken verwaltet werden sollen und die Bezeichnung D-1.GPR nicht mit anderen Allegro-Benutzern in Konflikt gerät, kann diese Datei auch im Programmordner liegen, der die Allegro-Programme enthält. Für die spezifischen Anzeigen einzelner Datenbanken wird die Datei ANZEIGE.GPT (s.o.) benutzt.

MCP.OPT:

Voreinstellungen für die von Allegro benutzten Programme, Festlegung der Struktur des SSG-FI-Systems, Zugriff auf SSG-FI-Routinen

MCP.PRE:

Speicherung der letzten Allegro-Einstellung, bei Mehrfachnutzung eher hinderlich

Zwischenteiltabellen

WINDOS.GPT:

Umsetzung vom Windows-Zeichensatz in den DOS-Zeichensatz

DOSWIN.GPT:

Umsetzung vom DOS-Zeichensatz in den Windows-Zeichensatz

P-HTML.GPT:

Umsetzung von Sonderzeichen des DOS-Zeichensatz in HTML (z.B. ä → ä)

Z-ALLG.GPT:

Zwischenteile für den HTML-Export des SSG-FI-Systems

Exportprogramme

Diese Programme bestehen aus einer Batchdatei und einer oder mehreren zuzusätzlichen Dateien (normalerweise Produktionsdateien vom Typ *.GPR). Beispiele:

INFODATA.BAT mit P-DATA.GPR:

Exportiert die Datenbank im HTML-Format, Ergebnis: Infodata.htm

INDEXDAT.BAT mit P-IND.GPR:

Exportiert eine Version der Sätze der Datenbank zur Indizierung, Ergebnis: Indexdat.htm

MINITEMP.BAT mit S-MINI.GPR und P-MINI.GPR:

Export der Verteilerdokumente als Mini-Templates, Ergebnis: Minitemp.htm

LINKLIST.BAT mit S-LINK.GPR und P-LINK.GPR:

Export der Verteilerdokumente als Listen, Ergebnis: Linklist.htm

GESAMT.BAT:

Benutzt S-MINI.GPR, P-MINI.GPR, P-DATA.GPR und P-IND.GPR um einen vollständigen Export für den HTML-Bereich zu erstellen: Verteiler-, Daten- und Indexdokumente, Ergebnis: Infodata.htm, Indexdat.htm, Minitemp.htm

Hilfsprogramme**IMPORT.BAT mit IMPORT1.GPR und IMPORT2.GPR:**

Importiert Datensätze in Standardform

TEXT.BAT mit I-TEXT.GPR:

exportiert einen in I-TEXT.GPR definierten Auszug der Datenbank als Textdatei

NUMMER.BAT mit I-NR.GPR:

Numeriert die Datenbank neu

PRINT.BAT mit S-PRINT.GPR und P-PRINT.GPR („Export für den Druck“):

Produziert eine geordnete Textdatei für eine Druckversion der Datenbank

Die unter „Anzeige und Voreinstellungen“ und „Hilfsprogramme“ stehenden Dateien dienen der allgemeinen Arbeit an und mit der Datenbank und werden im folgenden kurz besprochen. Die Zwischentabellen und Exportprogramme

werden für die Präsentation der Datenbank im Internet benötigt und in Kapitel 6.3 ausführlicher beschrieben.

Anzeige und Voreinstellungen

Das Allegro-System wird mit der Batchdatei XYZDAT.BAT gestartet, man ruft damit das Programm **Cockpit** (ACP) auf. Von hier aus werden die Datenbank und Routinen aufgerufen und dazu wird die Datei MCP.OPT eingelesen. Hierin werden die Routinen definiert und die Parameter gesetzt, die nicht direkt in der Batchdatei bestimmt werden. In der Voreinstellungsdatei MCP.OPT werden auch das Programm- und Datenverzeichnis, die Datenbank und die Konfigurationsdatei festgelegt. Dabei werden die Endungen der Dateien nicht angegeben, sondern von Allegro automatisch erschlossen. Darum dürfen diese auch nicht verändert werden und der Anfangsbuchstabe der Konfigurationsdatei muß **G** bleiben (bzw. sonst alles ersetzt werden). Wichtig ist, daß der Gesamtpfad des Datenverzeichnisses nicht zu viele Buchstaben hat, da das von Allegro als MS-DOS-Programm nicht verwaltet werden kann.

Außerdem enthält die MCP.OPT-Datei Voreinstellungen, Definition und Parameter für Programme und Routinen, die aus Allegro aufgerufen werden können, z.B. den Texteditor und einige SSG-FI-Routinen (s.u.). Automatisch legt Allegro die Datei MCP.PRE an, in der die Grundeinstellungen gespeichert werden. Diese Datei kann gelöscht werden, wenn mit den Voreinstellungen etwas nicht klappt, z. B. der Allegro-Befehl „Zurücksetzen der Optionen“ nicht funktioniert. Die MCP.OPT-Datei enthält auch Verweise auf einige Allegro-Routinen, die für SSG-FI nicht benötigt werden und je nach Installation verfügbar sind oder nicht.

In der MCP.OPT-Datei werden auch die verschiedenen SSG-FI Routinen definiert, die direkt aus dem Cockpit aufgerufen werden können. Diese sind unterteilt in „SSG-Export-Routinen“ und „SSG-Sonder-Routinen“. Die ersten dienen der Präsentation der Datenbank im Internet (vgl. Kap. 6.3), die anderen speziellen Exportabläufen und der allgemeinen Allegro-Arbeit.

Von Bedeutung sind noch einige Dateien, die allgemein von Allegro genutzt werden können:

die Buchstabenersatztable I.GPT

(für Indizierung mit XYZ.GPI vorgesehen, aber nicht genutzt) ersetzt alle Zeichen durch ASCII-Zeichen mit den Nummer 1-127, und zwar in der Haupttabelle (p) durch Großbuchstaben und in der Alternativtabelle (q) durch Kleinbuchstaben. Wenn dies für die Indexdatei eingesetzt wird, kann der Index einheitlich groß bzw. klein geschrieben werden, dann ist die Registersuche nicht mehr „Case sensitive“.

die Stoppworttable SWL1.GPT

(für XYZ.GPI vorgesehen, aber nicht genutzt) enthält eine Liste von Stopwörtern in verschiedenen Sprachen, die bei der Indizierung ausgelassen

werden sollen. Da derzeit keine Einzelwörter indiziert werden, ist sie überflüssig, sie sollte aber aktiviert werden, wenn z.B. die Beschreibung (Kategorie 20) in einem Register indiziert wird.

S1.ASP, S2.ASP

sind alternative Suchtabellen, die auf dem MS-DOS-Zeichensatz beruhen. Das Programm SRCH für die Volltextsuche benutzt die Datei S1.ASP, um Groß- in Kleinbuchstaben umzusetzen und Umlaute aufzulösen. Wenn für die Suche eine andere Einstellung gewünscht wird, kann die Datei durch eine andere ersetzt werden, die Groß/Kleinschreibung unverändert läßt oder vielleicht auch gar nichts ersetzt; benutzt wird immer die Datei mit Namen S1.ASP.

Hilfsprogramme

Für den Import von Standard-Textdateien, für Textexporte und die Neunummerierung der Datenbank stehen spezielle Routinen zur Verfügung, die im Allegro-Cockpit im Untermenü des Eintrags „2 SSG-Sonder-Routinen“ im dem Menü „Routinen“ zur Verfügung stehen.

Import der HTML-Eingabe:

wandelt Datensätze, die über das Internet in einem Formular eingegeben wurden, in Allegro-Grunddatei um und importiert diese dann in die Datenbank. Dabei werden Datensätze mit einer schon vorhandenen URL zurückgewiesen und in einer Datei „Doppelt.txt“ gespeichert. Die Einstellung geht davon aus, daß die Datei mit den neuen Datensätzen Allegro.txt heißt und in dem Datenordner liegt. Benutzt IMPORT.BAT mit IMPORT1.GPR und IMPORT2.GPR.

Numerieren:

führt in mehreren Schritten eine Neunummerierung der Datenbank durch. Diese Funktion dient dazu, die Dateinummern für die Verweise bei den HTML-Dateien eindeutig und fortlaufend zu machen, sie funktioniert nur richtig, wenn die Datenbank nicht gleichzeitig von jemand benutzt wird, da sonst zu verändernde und zu löschende Dateien gesperrt sind. Außerdem ist es sinnvoll, die Datenbank vor der Numerierung zu entlüften. Durch Anpassung der Exportdatei I-NR.GPR kann die Datenbank auch auf andere Art manipuliert werden. Der Prozeß läuft wie folgt ab: Zunächst wird die Datenbank exportiert und dabei neu nummeriert, die alte Dateinummer wird gelöscht und in Kategorie 99 gespeichert, die aktuelle Satznummer wird zur neuen Dateinummer. Danach wird die alte Datenbank in XYZ_1.g2d umbenannt, alle Register gelöscht und der Export neu als Datenbank indiziert. Da bei diesem Vorgang die bestehenden Identifikationsnummern verändert werden, wird vor Ablauf des Programms gefragt, ob die Nummern wirklich gelöscht werden sollen. Außerdem sind als zusätzliche Sicherungen die Speicherung der

alten Nummer in #99 und die Sicherung der alten Datenbank als XYZ_1.g2g eingebaut. Benutzt NUMMER.BAT mit I-NR.GPR.

Export als Text:

fragt nach einem Namen und speichert einen Auszug der Datenbank als Textdatei (im Windows-Zeichensatz) unter diesem Namen mit Endung txt. Die zu exportierenden Kategorien werden in der Datei I-TEXT.GPR angegeben und können leicht geändert werden. Die einzelnen Einträge werden zeilenweise ausgegeben und die Datensätze durch eine Leerzeile von einander getrennt. Dieses Format kann durch Bearbeitung der Datei I-text.gpr verändert werden. Benutzt TEXT.BAT mit I-TEXT.GPR.

Export für den Druck:

erstellt eine Datei zur Druckvorbereitung, in der die Datensätze formal und thematisch geordnet sind. Dazu müssen im Register 5 für die Formalschlüssel Konkordanzen der Form „vcz6=1m“ eingerichtet werden, die die Ordnung des Exports steuern. Dabei bedeutet:

- **=1 1. Block (allgemein),
- **=2 2. Block (spezielle Themen),
- **=0 kein Export.

Die Datensätze im 2. Block werden thematisch geordnet und entsprechend ihrer Themen zusammengefaßt. Der Buchstabe hinter 2 in „**=2m“ steuert dabei die formale Sortierung. Benutzt PRINT.BAT mit S-PRINT.GPR und P-PRINT.GPR.

Für diesen formatierten Ausdruck kann dann noch mit MS Word automatisch ein Titel- und Schlagwortindex aufgebaut werden. Dazu wird eine Konkordanzdatei mit dem Format

<auftretender Ausdruck> Tabulator <Indexeintrag>

benötigt. Deren Erzeugung dienen die folgenden Hilfsprogramme.

Titelexport für Index:

erstellt eine geordnete Liste aller Titel in der Form „Titel Tabulator [Titel]“, wobei bei der zweiten Version des Titels Artikel am Anfang entfernt sind. Benutzt TITEL.BAT mit TITEL.GPR.

Schlagwortexport für Index:

erstellt eine geordnete Liste aller Schlagwörter in der Form „Schlagwort Tabulator Schlagwort“. Benutzt KEYWORD.BAT mit KEYWORD1.GPR und KEYWORD2.GPR

Benutzte Datentypen

Wie oben schon bemerkt, werden von Allegro verschiedene Datentypen verwandt, deren Endung von der benutzten Konfiguration abhängig ist. Zur Übersicht werden diese Datentypen hier noch einmal zusammengestellt, so daß sie unter Windows als Dateitypen eingerichtet werden können. Zur Bearbeitung

kann ein beliebiges Textverarbeitungsprogramm zugeordnet werden, z.B. WordPad oder der MS-DOS-Editor. Die folgenden Endungen beziehen sich somit auf die Allegro-Konfigurationen mit dem Kennbuchstaben **G**. Groß- und Kleinschreibung werden bei den Endungen nicht beachtet.

Datenbank-Definition:	cfg	Allegro Konfiguration
	gpi	Allegro Indexparameter
	opt	Allegro Voreinstellungen
Datenbankdateien:	gld	Allegro Datenbank
	glg	Allegro Grunddatei
	stl	Allegro Kurztitel
	tbl	Allegro Satztable
	gdx	Allegro Indexdatei
Produktions- und Hilfedateien:	gpr	Allegro Produktionsdatei
	gim	Allegro Importdatei
	gpt	Allegro Tabelle
	aph	Allegro Phrasendatei (Textbausteine)
	pre	Allegro Einstellungen
Allgemeine Dateien:	bat	Batchdatei (Stapelverarbeitung)
	asp	Allegro Suchtable
	log	Allegro Logdatei

6.1.3 Eingabe in die Allegro-Datenbank

Mit der beschriebenen Einrichtung können nun Daten in das Allegro-System eingegeben werden. Der gewünschte SSG-FI-Bereich wird mit einer Batchdatei XYZDAT.BAT gestartet und die Allegro-Umgebung nach den Vorgaben dieser Batchdatei und der Voreinstellungsdatei MCP.OPT eingerichtet. Die Bedienung des Allegro-Datenbanksystems kann hier nicht beschrieben werden, dazu sei auf das Allegro-Systemhandbuch⁵³ oder die Einführungen in Allegro wie die „*allegro-Ouvertüre*“⁵⁴ verwiesen.

Zur Eingabe gibt es prinzipiell drei Möglichkeiten:

- Mit „I“ wird ein neuer Datensatz angelegt, indem die Abfrageliste abgearbeitet wird,
- mit „C“ wird eine Kopie eines Datensatzes angelegt, die dann zu einem neuen Datensatz umgearbeitet wird,
- mit dem Import-Programm werden Datensätze als formatierter Text in die Datenbank importiert.

53 Bernhard Eversberg: *allegro-C*, Systemhandbuch, Version 15. Braunschweig 1998

54 *allegro*-Entwicklungsabteilung: *allegro-Ouvertüre*, Partitur für Einsteiger. Braunschweig 1996

Die dritte Möglichkeit wird bei der Eingabe mit dem HTML-Formular besprochen (vgl. Kap. 6.4.2).

Die zweite Möglichkeit kann benutzt werden, wenn Varianten eines Datensatzes angelegt werden, z.B. ein Server in mehrere thematische Sites unterteilt wird oder weitgehend ähnliche Datensätze aufgenommen werden sollen. Außerdem gibt es in den Datenbanken einen Satz „Kopiervorlage“, der mit dieser Methode kopiert werden kann. Dieser Datensatz ist bis auf den Titel und einen Stoppvermerk leer, so daß hier formatierter Text *eingeklebt* werden kann: Wenn der Text aus Zeilen der Form

„Kategorienummer <Leerzeichen> Kategorietext <Zeilenwechsel>“

besteht, kann er so direkt in die Datenbank eingebunden werden.

Die Standardeingabe aber ist die mit „I“ eingeleitete Neuerfassung mit der Abfrageliste. Die Abfrageliste ist in der Konfigurationsdatei GXYZ.CFG enthalten und somit an die einzelne Fächer angepaßt.

6.2 Datenpräsentation: der SSG-FI-Server

6.2.1 Die Dateistruktur des Servers

Der so dargestellten Datenstruktur entsprechen die Dateistrukturen auf dem Server, die von der Datenbank (halb-) automatisch aufgefüllt werden. Abb. 6.1 gibt einen Überblick über die Dateistruktur auf dem Server (Ordner sind fett und Dateien normal dargestellt)

Für die Dateinamen gelten die folgenden Konventionen:

1. Der Server, auf dem die HTML-Dokumente liegen, kann normalerweise ein HTML-Dokument automatisch aufrufen, wenn ein *Ordner* statt einer *Datei* angefordert wird. Dieses Dokument heißt auf dem Göttinger Server *index.html*, und dieser Name wird auf den verschiedenen Ebenen für den Zentralverteiler benutzt. Somit liefert die HTML-Anfrage „http://www.SUB.Uni-Goettingen.de/ssgfi/math/“ automatisch das Dokument „http://www.SUB.Uni-Goettingen.de/ssgfi/math/index.html“, den Verteiler des MathGuide des SSG-FI. Bei anderer Servereinstellung sollte dieses Dokument umbenannt werden. Die Ordner, die nicht vom Internet aus direkt angesprochen werden sollen (z.B. **grafiken**, **indexdata** etc.) enthalten eine Datei *index.html*, die auf die Hauptverteiler verweist.
2. Der SSG-FI-Bereich wird auf Englisch aufgebaut. Die Endung „_en.html“ weist auf ein Dokument in *englischer* Sprache hin (en ist der ISO-639-Code für Englisch), die Endung „_de.html“ auf die *deutsche* Sprache (z.B. Hilfedateien etc.).
3. Die Verteiler-Dokumente, die durch den *Source Type Catalog* bzw. den *Subject Catalog* erreicht werden, liegen im Ordner „**formal**“ bzw. „**subject**“. Ihre Namen werden klein geschrieben und bestehen aus dem Kurznamen

der Datenbank xyz, dem Code für die formale bzw. thematische Klassifikation, gegebenenfalls dem Ländercode (nach ISO 639), dem Code für die Art der Ressource (on, off, pr) und der Endung „.en.html“. Die einzelnen Namensteile werden durch „_“ voneinander getrennt, so daß sich z.B. xyz_fk27_on_en.html oder xyz_sf1_at_on_en.html ergibt.

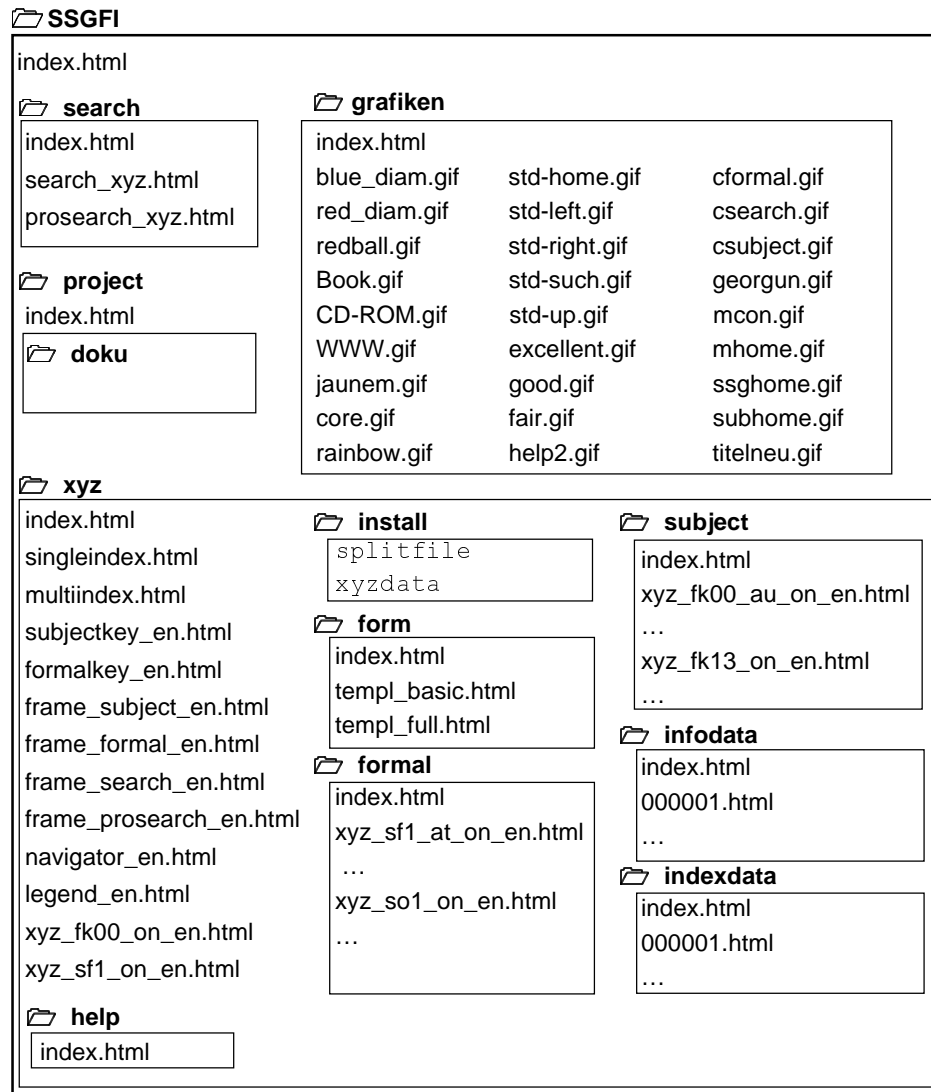


Abb. 6.1: Datenstruktur des SSG-FI-Servers

4. Die eigentlichen Informationsdateien liegen im Ordner „**infodata**“ und tragen als Namen nur die Nummer (sechsstellig) des Datensatzes in der Datenbank mit der Endung html, z.B. 000239.html.
5. Die Dateien für die Indizierung liegen im Ordner „**indexdata**“ und tragen als Namen ebenfalls die Nummer des Datensatzes in der Datenbank mit der Endung html, z.B. 000239.html.

Die Dokumente im Ordner „**grafiken**“ werden bei der Darstellung der Dokumente genutzt und müssen im Dateiverzeichnis auf der gleichen Ebene wie der Ordner „**xyz**“ liegen, ebenso der Ordner „**search**“ mit den Suchformularen. Der Ordner „**projekt**“ enthält etwaige Projektdokumentationen.

Die in den *Ordnern* „**formal**“, „**subject**“, „**infodata**“ und „**indexdata**“ enthaltenen Dokumente werden durch die Datenbank automatisch mit den angegebenen Namen erzeugt und sind als Hypertextdokumente miteinander verbunden. Im Ordner „**help**“ liegt eine fachspezifische Hilfe- und Erklärungsdatei, und der Ordner „**install**“ dient der Installation des Datenbereiches (s.u.).

6.2.2 Die Dateien des Hauptordners

Die *Dateien* im Ordner „**xyz**“ werden durch die Datenbank *nicht* verändert und müssen per Handarbeit an die gegebenen Bedingungen angepaßt werden. Sie haben die folgende Bedeutung:

index.html	Das Einstiegsdokument des xyz -Bereiches, hat diesen Namen, damit es automatisch aufgerufen wird, wenn jemand den Ordner „xyz“ öffnen will (s.o.). Das Dokument ruft bei einem framefähigen Browser das Dokument „singleindex.html“ in einem Rahmen auf, sonst zeigt es ein identisches Dokument ohne Rahmen.
singleindex.html	Das Einstiegsdokument des xyz -Bereiches für framefähige Browser mit den vier Einstiegsmöglichkeiten.
multiindex.html	Ein weiteres Einstiegsdokument, das benutzt wird, wenn die externen Links in einem zweiten Fenster angezeigt werden sollen.
frame_..._en.html	HTML-Rahmendokument, das bei dem „...“ entsprechenden Einstieg (z.B. <i>Source Type Catalog, Search</i>) aufgerufen wird und seinerseits die „...“ entsprechende Datei (z.B. <i>formalkey_en.html, search_xyz.html</i>) sowie <i>navigator_en.html</i> in einem Rahmen („Frame“) darstellt.
legend_en.html	Zeichenerklärung für die Mini-Templates, wird alternativ zu <i>navigator_en.html</i> im selben Rahmen angeboten.

navigator_en.html	Navigationshilfe für Internetbrowser, die Rahmen (Frames) unterstützen, enthält Verbindungen zu den Hauptdokumenten des SSG-FI-Seiten und muß an die lokalen Namen angepaßt werden.
formalkey_en.html	Die Liste mit den Verbindungen zu den Verteilern des Formalschlüssels, einheitlich für alle SSG-FI-Bereiche (<i>Source Type Catalog</i> ; „xyz_so1_on_en.html“ etc.).
subjectkey_en.html	Die Liste mit den Verbindungen zu den Verteilern des Themenschlüssels, ist für den jeweiligen SSG-FI-Bereich spezifisch (<i>Subject Catalog</i> ; „xyz_fk01_on_en.html“ etc.).
xyz_fk00_on_en.html	Zwei spezielle Verteiler-Dokumente, die für einen Themenschlüssel (z.B. fk00) und einen Formalschlüssel (sf1) eingeführt wurden, da die entsprechenden Listen zu groß geworden sind. Diese Kategorien werden nun nach Ländern geordnet angeboten, dazu gehören Verteiler-Dokumente mit einem Ländercode im Namen wie xyz_fk00_au_on_en.html und xyz_sf1_at_on_en.html.
xyz_sf1_on_en.html	

Die Dateien ab navigator_en.html enthalten Verweise auf die Hauptdokumente der SSG-FI-Guides, auf übergeordnete Dateien (*SSG-FI Home*, *SUB Home*) und verwandte Seiten (*Webis*) und müssen an die lokalen Wünsche und Gegebenheiten angepaßt werden (auch wenn der Namen „index.html“ der Homepage verändert wird).

Während die Anpassung der meisten Dokumente an ein neues Thema durch das Suchen und Ersetzen in Textdokumenten (Ersetze überall „xyz“ durch das gewünschte Datenbankkürzel) möglich ist, muß für das neue Themengebiet auch ein neuer Themenschlüssel erarbeitet und die bestehende Liste subjectkey_en.html entsprechend angepaßt werden. Das ist neben dem eigentlichen Erfassen der Dokumente die aufwendigste Tätigkeit bei dem Aufbau eines Internet-Fachinformationsführers nach SSG-FI-Muster.

Die Dateien in „**formal**“, „**subject**“, „**infodata**“ und „**indexdata**“ werden aus drei Dateien erstellt, die aus der Allegro Datenbank exportiert werden und typischerweise `Minitemp.htm`, `Infodata.htm` und `Indexdat.htm` heißen. Diese Dateien enthalten die vollständigen HTML-Dateien, müssen aber noch an geeigneten Stellen „auseinandergeschnitten“ werden. Diese „Schnittstellen“ sind durch „>>>>filename“ markiert, dahinter folgt der Name, den die Datei mit dem nachfolgenden Inhalt bekommt. Die Arbeit des Auseinanderschneiden kann durch ein einfaches Skript auf dem jeweiligen Server erledigt werden (erst dort, weil sonst oft Probleme mit langen Dateinamen auftreten). In Göttingen wird auf dem LINUX-Server ein awk-Skript „splitfile“ benutzt: `Minitemp.htm`, `Infodata.htm` und `Indexdat.htm` werden in das Verzeichnis

„install“ im Ordner „xyz“ kopiert, wo auch `splitfile` liegt. Dort muß dann mit einem geeigneten Programm (z.B. `xterm` oder `TkDesk`) weitergearbeitet werden. Das Verzeichnis wird (unter `xterm` z.B. mit `cd ...?/SSGFI/xyz/install`) als Arbeitsverzeichnis gewählt und durch die Eingabe „`splitfile minitemp.htm`“ z.B. die Ordner „**formal**“ und „**subject**“ mit Inhalt erzeugt. Diese liegen dann im **übergeordneten** Verzeichnis. Wenn diese Ordner und Dateien schon bestehen, werden die alten Dateien nicht ersetzt, sondern die neuen an sie angehängt, darum müssen die alten Dateien zuerst gelöscht werden. Alternativ können die neuen Ordner in einem Extraordner erzeugt werden.

6.2.3 Die Suchdateien

In dem Ordner „**Search**“ sind die zur Suche benötigten HTML-Dateien enthalten. Für jeden SSG-FI-Bereich wird je eine Datei „`search_xyz.html`“ für die einfache Suche und „`prosearch_xyz.html`“ für die komplexe Suche benötigt. Beide Dateien sind HTML-Formulare, die an ein *CGI-Script* auf dem Server geschickt werden. Dabei hat „`search_xyz.html`“ nur ein Feld und ein Auswahlnenü für die Suche in speziellen Kategorien. Die komplexe Suche „`prosearch_xyz.html`“ besitzt drei Felder für die kombinierte Suche mit booleschen Operatoren sowie ein Auswahlnenü für den Formalschlüssel und eventuell ein weiteres Menü für die thematische Suche. Die genaue Funktionsweise der Suche wird in Kapitel 4.2 und die Installation in Kapitel 6.4.1 beschrieben.

6.3 Der Export der Daten in das Internet

6.3.1 Die Export-Routinen

Für die Präsentation im Internet müssen die Daten aus der Allegro-Datenbank exportiert werden. Dazu dienen die Export-Skripte, die in den Exportparameter-Dateien enthalten sind. Sie werden am einfachsten vom Allegro-Cockpit aus als SSG-FI-Routinen aufgerufen. In Klammern wird der zugeordnete Allegro-Hinweis angegeben:

1 SSG-FI-Export-Routinen

- 1 Export der Infodateien
(Exportiert die Datenbank im HTML-Format)
- 2 Export der Indexdateien
(Exportiert die Datenbank zur Indizierung)
- 3 Export der Mini-Templates
(Export der Formal- und Themenverteiler als Mini-Templates)
- 4 Export der Verteilerdokumente
(Exportiert die Formal- und Themenverteiler als HTML-Listen)
- 5 Gesamtexport mit Mini-Templates
(Exportiert die Daten-, Verteiler- und Indexdateien)

Diese Programme werden im folgenden kurz dargestellt, zur ihrer Ausführung werden benötigt:

die Batchdateien	INFODATA.BAT, INDEXDAT.BAT, MINITEMP.BAT, LINKLIST.BAT, GESAMT.BAT
die Exportparameter-Dateien	P-DATA.GPR, P-IND.GPR, S-MINI.GPR, P-MINI.GPR, S-LINK.GPR, P-LINK.GPR
die HTML-Ersatztable	P-HTML.GPT
die Zwischenteiltabellen	Z-ALLG.GPT, Z-SPEZ.GPT

Bis auf Z-SPEZ.GPT (im Datenverzeichnis) liegen diese Dateien im Ordner SSGFI. Mit der Auswahl des entsprechenden Menüpunktes bzw. durch Eingabe seiner Nummer wird das Programm aufgerufen. Wenn mehr als eine Datenbank im Datenverzeichnis liegt, muß nach dem Aufruf des Programmes noch die Datenbank ausgewählt werden, die exportiert werden soll. Bei den Programmen 1 – 4 wird außerdem nach einem Suchbegriff gefragt, so daß nur die Datensätze exportiert werden, die diesen Begriff enthalten – das empfiehlt sich insbesondere zum Testen der Skripte. Der Suchbegriff „0“ exportiert die gesamte Datenbank. Mit dem Aufruf eines Programms wird als Ausgabe eine Textdatei mit der Endung .htm erzeugt, sie kann so direkt mit einem Browser betrachtet werden. Das Programm ruft die entsprechende Batchdatei und diese die Exportparameter-Dateien auf. Dabei wird intern automatisch das Programm- und Datenverzeichnis sowie der SSGFI-Ordner benutzt, so daß hier keine Anpassung notwendig ist. Für die Anpassung der Exportprogramme an das jeweilige SSG-FI-System müssen die verschiedenen Programmvarianten ausgewählt werden. In den Exportparameter-Dateien steht der Name der Datenbank in der Variablen <fb> zur Verfügung, so daß das Skript auf verschiedene Namen entsprechend reagieren kann. Diese Verzweigungen sind mit dem Kommentar „Test: ...“ versehen, in dem entsprechenden Vergleich (z.B. c"Geo") muß der eigene Datenbankname eingetragen, also „Geo“ durch das eigene Datenbankkürzel ersetzt werden, wenn die zugehörige Verzweigung durchgeführt werden soll.

6.3.2 Spezielle Datensätze

Die Exportprogramme greifen auf einige spezielle Datensätze zurück, die als Konkordanzlisten für verschiedene Abkürzungen benutzt werden. Wenn nicht anders beschrieben, funktioniert das so, daß in einem bestimmten Register (i.A. Register 6 oder 7) die Einträge der Umsetzungstabelle stehen müssen. Um diese Registerinträge zu erzeugen, wird ein Datensatz angelegt (typischerweise mit einem Titel „Referenzliste ...“, in dem die Umsetzungstabelle in eine geeignete Kategorie geschrieben wird. Dabei werden die einzelnen Einträge

jeweils durch Semikolon und Leerzeichen getrennt. „Geeignet“ heißt, daß die benutzte Kategorie bei Semikola getrennt in dem gewünschten Register indiziert wird. Dies wird in der Indexparameterdatei XYZ.gpi festgelegt. Die Wahl der Kategorie ist beliebig, sie muß nur im richtigen Register indiziert werden, da dort nach der Übersetzung dieser Codes gesucht wird. Diese Datensätze werden in diesem Kapitel beschrieben.

Ersetzung der Sprach- und Ländercodes

Listen für diese Datensätze ergeben sich aus dem Sprachcode nach ISO 639-1 (vgl. Kap. 6.2.3) und dem Ländercode nach ISO 3166 (vgl. Kap. 6.2.2). Der eine Datensatz hat den Namen „Referenzliste Sprachcode“ und enthält in Kategorie #94 Einträge der Form „aa=Afar; ab=Abkhazian; ...“, der andere heißt „Referenzliste Ländercode“ und enthält in Kategorie #94 Einträge der Form „AD=Andorra; AE=United Arab Emirates; ...“. Die Kategorie #94 muß in Register 7 indiziert werden, getrennt bei den Semikola. Die Identifikation ist eindeutig, da der Sprachcode aus Klein- und der Ländercode aus Großbuchstaben besteht.

Ersetzung der Formal- und Themenschlüssel

Der im SSG-FI-Projekt verwendete Formalschlüssel wird mit einem Datensatz „Referenzliste Formal Key“ umgesetzt. Dieser hat in Kategorie #91 Einträge der Form „a2=Manuals; ak=Lists of Abbreviations; ...“ (vgl. Kap. 6.2.1) und diese Kategorie muß in Register 6 indiziert werden.

Der fachspezifische Themenschlüssel ist im Projekt SSG-FI unterschiedlich komplex. Während im Mathematikbereich nur ein Datensatz „Referenzliste MSC“ benutzt wird, werden für die Umsetzung des jeweiligen GOK-Teiles im Bereich Geowissenschaften drei (Referenzliste Subject Key 1 - 3) und im Bereich Anglo-Amerikanischer Kulturraum (Geschichte) sechs Datensätze (AA - Referenzliste für den Subject Catalog Teil 1 - 6) benötigt. Einträge haben z.B. die Form „PF 000=World History; PFH 000=History of the discoveries (to 1800); ..“ und müssen in Register 6 indiziert werden.

Damit die Ersetzung eindeutig ist, darf keiner der SSG-FI-Formalschlüssel als Themenschlüssel auftreten.

Ein zusätzlicher Mechanismus wurde eingebaut, um verschiedene Klassen zusammenzufassen. Dazu müssen die entsprechenden Ersatzeinträge alphabetisch (nach ASCII-Wert) geordnet sein und als letzter Eintrag ein „*verz=#??“ hinzugefügt werden. Dabei steht „*“ für den Anfang der Kategorien, die hier eingeordnet werden sollen (die ersten 1 - 3 Buchstaben der Klassifikation) und „#??“ für die Nummer der Kategorie, in der die Einträge stehen. Zum Beispiel werden im Bereich Anglo-Amerikanischer Kulturraum (Literatur) die aufeinanderfolgenden GOK-Klassen IC 520-527 zu „American Literature: Literature and

Society“ zusammengefaßt. Damit z.B. die Klasse IC 523 hier eingeordnet werden kann, ist als letzter Eintrag „ICverz=#41“ vorhanden (die Einträge stehen in Kategorie #41). Wenn nun kein Eintrag „IC 523=...“ vorhanden ist, prüft das Programm der Reihe nach, ob „IC verz=...“, „ICverz=...“ oder „Iverz=...“ existiert. Da „ICverz=#41“ vorhanden ist, wird der entsprechende Datensatz aufgerufen, Kategorie #41 eingelesen und deren Einträge der Reihe nach mit IC 523 verglichen. Der letzte, der kleiner als IC 523 ist, wird als Ergebnis ausgegeben.

Konkordanzen von GOK zu anderen Klassifikationen

Im Bereich Geowissenschaften wurden weitere Konkordanzlisten erstellt, mit denen die GOK-Klassifikation in DDC, BK und eine zusätzliche reduzierte Klassifikation („Subject Main Fields“) überführt werden kann. Dazu wurden Datensätze wie zur Umsetzung des Themenschlüssels angelegt (z.B. Referenzliste für die Konkordanz GOK zu BK), nur daß bei diesen Datensätzen ein anderes Trennzeichen gewählt wurde, so daß sich z.B. „XX 000<XX.XX; QA 000<74.00; ...“ ergibt. In den Programmen wird für die Datenbank Geo die folgende Konvention benutzt:

Trennzeichen	Bedeutung
=	GOK → GOKverbal
<	GOK → DDC
>	GOK → BK
:	GOK → Subject Main Fields

6.3.3 Die Exportprogramme

Der Export der Daten wird durch Batchprogramme gestartet, die ihrerseits geeignete Allegro-Programme (srch und asort) aufrufen und für den Export die Exportparameter-Dateien (Typ GPR) benutzen. In denen ist festgelegt, welche Daten exportiert werden und welche Zusatzinformationen noch in das Exportdokument eingefügt werden sollen. Die Parameterdateien selbst sind mit Kommentaren versehen, die Detailinformationen über die jeweiligen Programmschritte geben. Hier sollen die Programme nur global beschrieben werden.

Bei allen Exportprogrammen werden diejenigen Datensätze übergangen, die in der Allegro-Kategorie 94 das Wort „stop“ enthalten. Damit wird z.B. verhindert, daß auch die Konkordanzlisten für die verschiedenen Abkürzungen exportiert werden.

Alle Exportprogramme benutzen die allgemeine Textbausteintabelle („Zwischenteile“) Z-Allg.gpt, die im Datenordner befindlichen speziellen Textbausteine Z-SPEZ.GPT und die Umsetzungstabelle P-HTML.GPT für die Umwandlung von Sonderzeichen in HTML-Code.

Export der Infodateien

„Export der Infodateien“ (INFODATA.BAT mit P-DATA.GPR) exportiert die Datenbank im HTML-Format, das Ergebnis ist eine Textdatei mit dem Namen „Infodata.htm“. Benutzt wird dabei die Exportparameter-Datei P-DATA.GPR.

P-DATA.GPR

In dieser Datei wird die Gestaltung der Infodateien festgelegt: der Seitenaufbau und die verwendeten Grafiken. Wenn eine andere Darstellung gewünscht wird, kann (mit nötiger Vorsicht) hier eingegriffen werden. Die Einträge der A Ilegro-Datenbank werden einerseits als (unsichtbare) Metadaten und andererseits in einer (sichtbaren) Tabelle dargestellt.

Das Programm benutzt die Umsetzungstabelle für den Formal- und Th emenschlüssel in Register 6 und die Länder- und Sprachdekodierung in Register 7 der Datenbank. Eine Anpassung für andere Informationsführer ist nicht notwendig, nur die Sekundär- und Tertiärklassifikationsbezeichnungen (DDC und BK) könnten geändert oder deaktiviert werden.

Export der Indexdateien

„Export der Indexdateien“ (INDEXDAT.BAT mit P-IND.GPR) führt fast dasselbe wie das obige Programm durch, nur einige Kategoriebezeichnungen (diejenigen, die „Schema“ oder „Role“ benutzen) werden anders aufbereitet. Außerdem wird bei Internet-Adressen die vorn stehende Formatangabe (z.B. „http://“ oder „ftp://“) entfernt, da diese Ausdrücke sonst nicht indiziert werden können. Die Einführung dieser besonderen Indexdateien ist eine etwas unglückliche Zwischenlösung, um Schwächen des benutzten Indizierungsprogramms „Swish-e“ auszugleichen. Normalerweise sollten die eigentlichen Infodateien indiziert werden. Benutzt wird dabei die Exportparameter-Datei P-IND.GPR, das Ergebnis ist eine Textdatei Indexdat.htm.

P-IND.GPR

Wie P-DATA.GPR mit den oben beschriebenen Modifikationen.

Export der Mini-Templates

„Export der Mini-Templates“ (MINITEMP.BAT mit S-MINI.GPR und P-MINI.GPR) exportiert die Verteilerdokumente als Zusammenstellung kleiner Tabellen (Mini-Templates), das Ergebnis ist eine Datei Minitemp.htm, in der alle HTML-Dateien in der oben beschriebenen Weise enthalten sind. Benutzt werden die Exportparameter-Dateien S-MINI.GPR und P-MINI.GPR.

S-MINI.GPR

Diese Datei sammelt die Informationen für die Verteilerdokumente des Source Type Catalog und des Subject Catalog, dazu werden die Informationen pro Datensatz in drei Felder angeordnet.

Das erste Feld ist das Sortierfeld. Es beginnt mit „u1“ und enthält den Formal- (Source Type) oder Themenschlüssel (Subject), eventuell das Land (für spezielle Formal- oder Themenschlüssel), dann die Art (on, off, pr) und den Namen (Title) des Site. Dabei wird der Name in Kleinschrift umgesetzt und etwaige Nichtsortierwörter (Der, Das, The ...) am Anfang werden beseitigt. Durch Sortieren anhand dieses Feldes werden die Informationen zusammengestellt.

Das zweite Feld beginnt mit „u2“ und enthält den Namen der Verteilerdatei in der SSG-FI-Standardform (xyz_fk73_on_en.html, xyz_sc_on_en.html...); das dritte Feld liefert die zusätzlichen Informationen: Dateinummer, Titel, Formal- oder Themenschlüssel, Infoquellenart, Niveau, Bewertung und URL sowie eventuell das Land.

Der Titel des Dokumentes wird in diesem Prozeß von MS-DOS-Schrift in HTML-Code umgewandelt, damit Umlaute und andere Sonderzeichen im Internet richtig angezeigt werden; dazu wird auf die Tabelle P-HTML.GPT zurückgegriffen. Außerdem werden einige Zwischenteile (für den Dateinamen) aus der Tabelle Z-SPEZ.GPT verwandt. Das Resultat dieses Exportes wird als uuu.glg im Datenordner gespeichert.

Dieses Skript erfordert keine Anpassung für andere Informationsführer, besitzt aber einige Einstellungen, die geändert werden können. Dabei prüft das Skript den benutzten Datenbanknamen in der Variablen <fb> und führt entsprechende Verzweigungen durch.

Zum einen können ausgewählte Schlüssel nach Ländern getrennt aufgelistet werden, das wird im MathGuide z.B. für den Bereich allgemeine Mathematik und im GeoGuide für die geowissenschaftlichen Fachbereiche etc. genutzt.

Zum anderen kann der Themenschlüssel modifiziert werden, z.B. um verschiedene Themen zu einem Themenbereich zusammenzufassen. Wenn der Themenschlüssel eine hierarchische Struktur aufweist, kann es sinnvoll sein, die Hierarchie ab einer bestimmten Tiefe abzuschneiden (im MathGuide werden nur die ersten zwei Zeichen benutzt), das bedeutet dann, daß im „ Subject Guide“ verschiedene Unterthemen zu einem Bereich zusammengefaßt werden. Es könnten auch die ersten Zeichen vor einem Trennzeichen (z.B. „-“) eingesetzt werden. Für komplexere Zusammenfassungen werden Konkordanzlisten benutzt, die den Themenschlüssel umsetzen (z.B. für den AAC-LitGuide). In jedem Fall wird die verbleibende Zeichenfolge zur Identifikation des Themas benutzt und muß so in Register 6 aufzufinden sein.

P-MINI.GPR

Die Datei *uuu.glg*, die mit *S-MINI.GPR* erzeugt wird, wird als nächstes sortiert, die sortierte Datei heißt dann *sss.glg*, und *uuu.glg* wird gelöscht. Aus der so sortierten Datei *sss.glg* erstellt *P-MINI.GPR* die HTML-Dokumente für die Verteiler nach Formal- und Themenschlüssel (Mini-Templates). Dazu werden zu jedem Dateinamen (Feld 2 in *sss.glg*) die zugehörigen Dokumente aufgelistet. Die Liste wird mit einer Überschrift versehen, die sich aus dem Namen des SSG-FI-Bereiches, dem Formal- bzw. Themenschlüssel, der Infoquellenart und gegebenenfalls dem Land oder – bei Autorensseiten – dem Anfangsbuchstaben ergibt. Die Einträge in der Tabelle sind in HTML-Zeichen (genauer: Latin-1) umgesetzte Titel des Dokumentes, der Formal- oder Themenschlüssel (der jeweils andere steht in der Überschrift des Mini-Template), die Schlagwörter, die Internetadresse (URL, mit der entsprechenden Internetquelle verbunden) bzw. entsprechende Identifikatoren für CD-ROMs oder Bücher, das Niveau und die Bewertung (fair - good – excellent entsprechend 1 – 3 Sternen). Zudem gibt es über das Dublin Core Icon eine Verbindung zu dem vollen Metadaten-Satz, der Infodatei. Das Programm benutzt (intensiv) die Zwischentabelle *Z-ALLG.GPT*, die Umsetzungstabellen für die Formal- und Themenschlüssel in Register 6 und eventuell die Länderdekodierung in Register 7 der Datenbank.

Export der Verteilerdokumente

„Export der Verteilerdokumente“ (*LINKLIST.BAT* mit *S-LINK.GPR* und *P-LINK.GPR*) wirkt wie *Minitemp.bat*, nur sind die exportierten Listen kleiner und weniger informativ als die Mini-Templates. Dies sind die ursprünglichen Verteilerdokumente des SSG-FI-Projektes, die jetzt nicht mehr benutzt werden. Das Ergebnis ist eine Datei *Linklist.htm*, benutzt werden die Exportparameter-Dateien *S-Mini.gpr* und *P-Mini.gpr*.

S-LINK.GPR

S-LINK.GPR wirkt genau wie *S-MINI.GPR* mit dem einzigen Unterschied, daß die Informationen im dritten Datenfeld nur Dateinummer, Titel, Infoquellenart, Land, Formal- oder Themenschlüssel und URL sind.

P-LINK.GPR

P-LINK.GPR wirkt genau wie *P-MINI.GPR* mit dem Unterschied, daß nur Tabellen der betreffenden Dokumente erstellt werden, die pro Datensatz als Eintrag nur den in HTML-Zeichen (genauer: Latin-1) umgesetzten Titel des Dokumentes (mit der entsprechenden Internetquelle verbunden) und eine zur zugehörigen Infodatei (ausführliches Metadaten-Template) führende Taste enthält.

Gesamlexport mit Mini-Templates

Gesamlexport mit Mini-Templates (GESAMT.BAT mit S-MINI.GPR, P-MINI.GPR, P-DATA.GPR und P-IND.GPR) wirkt wie die Programme 1, 2 und 3 zusammen, erstellt also einen vollständigen Satz der für die Internet-Darstellung erforderlichen Dokumente.

Für den Export benutzte Tabellen

Die HTML-Ersatztable P-HTML.GPT

Die HTML-Ersatztable P-HTML.GPT dient zwei verschiedenen Zwecken und besteht aus zwei Teilen. Der erste Zweck ist die Herstellung von HTML-Dateien, insbesondere die Umsetzung von MS-DOS Schrift in HTML-Schrift. Dazu werden einige Zwischenteile angegeben, die zur Zeichenformatierung in HTML genutzt werden können, z.B. für Fett und Kursiv, hoch- und tiefgestellte Schrift etc., und eine Umsetzungstabelle vom MS-DOS-Zeichensatz in HTML (also z.B. von ü in „ü“). Die Zeilen dieser Tabelle haben alle die folgende Struktur: „p“, Leerzeichen, MS-DOS-Zeichen, Leerzeichen, HTML-Ersatz in Anführungszeichen. Wenn eine andere Umsetzung gewünscht wird, kann diese Tabelle bearbeitet werden. Benutzt wird sie bei den Exportskripten in den Zeilen, die mit „#“ anfangen: der so aufgerufene Ausdruck (z.B. #10, der Titel eines Dokumentes) wird anhand dieser Tabelle umgesetzt.

Der zweite Teil der Datei P-HTML.GPT dient der Umsetzung von Schrift in einen reduzierten Zeichensatz, der im wesentlichen aus Kleinbuchstaben besteht. Dazu dient eine zweite Tabelle, deren Zeilen die Struktur „q“, Leerzeichen, MS-DOS-Zeichen, Leerzeichen, Ersatzzeichen in Anführungszeichen, haben. Diese Tabelle wird bei den Exportskripten in den Zeilen benutzt, die mit „!“ anfangen: der so aufgerufene Ausdruck (z.B. !10, der Titel eines Dokumentes) wird anhand dieser Tabelle umgesetzt. Das wird einerseits für die Sortierung von Datensätzen benutzt (damit nach dem ASCII-Wert eines Wortes sortiert werden kann) und andererseits für die Festlegung von Dateinamen, die im SSG-FI-System durchgängig mit Kleinbuchstaben geschrieben werden.

Bei beiden Tabellen bedeutet eine „1“ in der dritten Spalte (ohne Anführungszeichen), daß das Zeichen in der mittleren Spalte ersatzlos entfällt. Solche Zeichen tauchen also im Export nirgends auf und könnten durch andere Zeichen belegt werden - darauf beruht im wesentlichen der Einsatz des OSTWEST.FON in Allegro. Das Problem ist die Umsetzung weiterer Zeichen in HTML-Code: Hier gibt es zwar weitere Zeichensätze (z.B. Latin-2 für die Darstellung osteuropäischer Sprachen), aber keinen flexiblen Formalismus, der das Umschalten des Zeichensatzes innerhalb eines Dokumentes zuläßt. Dazu muß vielmehr dem ganzen HTML-Dokument im Header ein spezielles Meta-Tag mitgegeben werden, der „charset“ Parameter in einem „Content-Type“-Feld:


```
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=EUC-JP">.
```

Wenn dieser Mechanismus konsequent genutzt werden soll (z.B. für ein Sondersammelgebiet, das sich mit entsprechenden Ländern befaßt), so müssen spezielle Verfahren entwickelt werden, die ein Infodokument in einem anderen als dem Standardzeichensatz Latin-1 erstellen (einfaches Verfahren) und die die Titel dieser Dokumente zusammen mit anderen Titeln in Standardschrift in den Verteilerlisten darstellen (kompliziertes Verfahren). Falls daran Bedarf bestehen sollte, so sind die Projektmitarbeiter des SSG-FI Göttingen zur Unterstützung bei dieser Entwicklung behilflich.

Die Tabelle der Zwischenteile Z-ALLG.GPT

In der Tabelle Z-ALLG.GPT sind die allgemeinen Zwischenteile enthalten, die beim Export genutzt werden. Dies sind zum einen Teile des HTML-Code, die hier zur Abkürzung zusammengefaßt sind, und zum anderen variable Namen, die den verschiedenen SSG-FI-Bereichen gemeinsam sind. Wenn ein neuer Bereich aufgebaut wird, muß der Verweis auf den Göttinger Server und die SUB angepaßt werden. Zudem stehen hier die Namen der benutzten Grafiken und Ordner.

Die Tabelle der Zwischenteile Z-SPEZ.GPT

Diese Tabelle enthält Zwischenteile, die für den Export einzelner Datenbanken benutzt werden. Dies sind

- der Namen des Fachprojektes (z.B. GeoGuide),
- die Namen der fachspezifischen Kategorien (z.B. GOKverbal),
- die Hintergrundfarbe der HTML-Dateien,
- die fachspezifischen Präfixe der HTML-Dateien.

6.3.4 Einrichtung des Internetservers

Der SSG-FI-Bereich wird auf dem Server entsprechend der Darstellung in Kapitel 6.2 eingerichtet. Dazu muß die entsprechende gepackte Datei an den gewünschten Ort auf dem Server kopiert und entpackt werden:

Unter *Windows*: Die Daten aus server.zip mit der Option „use folder names“ extrahieren.

Unter *LINUX*: Die Daten aus server.tar.z mit „tar -v -z -x -f server.tar.z“ extrahieren.

Beide Optionen richten einen Ordner „**ssgfi**“ mit der benötigten Datenstruktur ein. Die einzelnen Dateien müssen nun entsprechend dem eigenen Projekt modifiziert werden (die mitgelieferten Dateien sind dem MathGuide entnommen), wobei für die Verbindungen des Themenschlüssels die Namenskonventionen zu beachten sind.

Für den Export der Datenbank müssen die Skripte wie oben beschrieben angepasst werden. Die notwendigen Anpassungen sind noch einmal tabellarisch zusammengefasst:

Datei	zu ändern
Z-ALLG.GPT	alle Bezüge auf die SUB
ANZEIGE.GPT	Bezeichnungen der Kategorien
Z-SPEZ.GPT	alle Bezüge auf <ul style="list-style-type: none"> – die Datenbank xyz – den Informationsführer XYZGuide – die Primärklassifikation KAT/KATverbal evtl. die Hintergrundfarbe
P-DATA.GPR, P-IND.GPR S-MINI.GPR, P-MINI.GPR S-LINK.GPR, P-LINK.GPR	Bei „Test“ (#ufb ... c"XYZ"): Prüfen welche Variante gewünscht wird

Dann wird ein Gesamtexport der Datenbank durchgeführt (im Allegro-Cockpit: Menü „1 SSG-Export-Routinen“, dann Untermenü „5 Gesamtexport mit Mini-Templates“). Die damit erzeugten Dateien Infodata.htm, Indexdat.htm und Minitemp.htm werden in den Ordner „**install**“ auf dem Server kopiert und dort das awk-Skript „splitfile“ benutzt, um die einzelnen Ordner „**formal**“, „**subject**“, „**infodata**“ und „**indexdata**“ aufzubauen, z.B. baut „splitfile Infodata.htm“ den Ordner „**infodata**“ und seinen Inhalt auf. Diese Vorgänge werden durch das Shell-Skript „xyzdata“ automatisiert: der Aufruf „xyzdata“ löscht zunächst etwaige Ordner „**formal**“, „**subject**“, „**infodata**“ und „**indexdata**“, erzeugt sie dann neu unter Benutzung von „splitfile“ und kopiert zusätzlich die Datei „index.html“ aus dem Ordner „**grafiken**“ in die neu erzeugten Ordner. Damit stehen die Daten auf dem Server zur Verfügung.

In dem Ordner ssgfi/xyz/install steht eine MS-Word-Datei „linktest.dot“, mit deren Hilfe die Verbindungen in den Verteilerdokumente in „**xyz**“ (z.B. formal_key_en.html) getestet werden können: Falls keine entsprechenden Dokumente vorhanden sind, werden die zugehörigen Links deaktiviert und auf Wunsch die ganze Zeile verborgen.

6.4 Zusätzliche Einrichtungen: Suche und Formulareingabe

Für die Einrichtung der Such- und Eingabefunktionen müssen auf dem Server einige CGI-Scripte angelegt werden. CGI (Common Gateway Interface) ermöglicht es, über das Internet ausgefüllte Formulare an den Server (in diesem Falle des SSG-FI) zu schicken, die dort dann weiter bearbeitet werden. Die zugehörigen Programme oder Scripte⁵⁵ ermöglichen es, die Eingabe weiter zu bearbeiten und entsprechende Aktionen auszulösen, also z.B. die Suche nach einem Ausdruck oder das Abspeichern in einer Datei. Damit das funktioniert, muß der eingesetzte Computer diese Programme verstehen, der Server also über einen *Perl*-Interpreter verfügen und etwaige *Shell-Scripte* an das vorliegende Betriebssystem angepaßt werden. Darüber hinaus müssen die Scripte in einem Bereich des Servers abgelegt werden, wo ihre Ausführung erlaubt wird. Im SSG-FI-Bereich ist dafür ein Verzeichnis *cgi-bin* vorgesehen.

6.4.1 Die technische Implementierung der Suche

Die Suche auf dem SSG-FI-Server ist von dem Allegro-System völlig unabhängig und beruht auf der Indizierung der HTML-Infodateien (tatsächlich spezieller Indexdateien) der einzelnen SSG-FI-Bereiche. Der Index wird von dem Programm *swish-e* erzeugt und ist je nach Umfang des entsprechenden Bereichs 0,5 - 2 MB groß. Zum Vergleich: die indizierten Dateien haben insgesamt einen Umfang von 1,5 - 5,7 MB. Dieser Index wird für jeden Bereich täglich automatisch neu erstellt, kann aber bei Bedarf auch manuell neu erzeugt werden, das dauert bei den umfangreicheren Guides etwa fünf Minuten.

Der so erstellte Index wird benutzt, um Anfragen zu beantworten, die mit einem der Suchformulare an den Server geschickt werden. Dazu wird diese Anfrage durch ein CGI-Script aufbereitet (*suche.cgi*). Die Installation dieser Programme und die Funktionsweise der Anfragebearbeitung wird in diesem Kapitel beschrieben.

Das Indexprogramm *swish-e*

Das Indexprogramm *swish-e* ist eine Weiterentwicklung von *swish*. *swish* (*Simple Web Indexing System for Humans*) und wurde von Kevin Hughes (khughes@worldpoint.com) entwickelt, ist aber mit dem Verschwinden des Servers der Enterprise Integration Technologies Corp. (<http://www.eit.com>) nicht mehr erreichbar. Im Internet sind noch Kopien der Dokumentation vorhanden,

⁵⁵ Im SSG-FI Projekt wird zur Programmierung im CGI-Bereich entweder *Perl* oder das LINUX-Betriebssystem (*shell-scripts*) eingesetzt. Diese werden nicht kompiliert, sind im Prinzip lesbar und werden im Gegensatz zu (z.B. C- oder PASCAL-) Programmen als *Scripte* bezeichnet.

z.B. bei <http://botes2.tesre.bo.cnr.it/Manuals/swish.1.1.3/swish.html>. *swish-e* (das e steht für *enhanced*) wird am Berkeley Digital Library SunSITE (<http://sunsite.berkeley.edu/SWISH-E/>) gepflegt, an dieser Web-Adresse findet sich auch eine kurze Dokumentation des Programmes, die hier nicht wiederholt werden soll. Die Besonderheit von *swish-e*, spezielle HTML-Tags zu indizieren, ermöglicht die Suche in den SSG-FI-Kategorien.

Wie *swish* ist *swish-e* kostenlos und kann von dem Server des Berkeley Digital Library heruntergeladen werden. Man bekommt den Quellcode mit den benutzten Bibliotheken und hat so prinzipiell die Möglichkeit, das Programm den eigenen Bedürfnissen anzupassen - das erfordert allerdings eine intensive Beschäftigung mit dem nicht sehr umfangreich kommentierten Quellcode in C.

Zur Installation muß das Programm kompiliert (derzeit in Version 1.2 für UNIX-Verwandte und Windows NT verfügbar) und auf dem Server eingerichtet werden. Die Kompilierung erfolgt einfach durch das Aufrufen einer „makefile“ Datei (in das Verzeichnis mit den *swish-e*-Dateien wechseln und „make“ eingeben), benötigt wird natürlich ein C-Kompiler. Durch die Konfigurationsdatei `config.h` können dabei die Grundeinstellungen von *swish-e* festgelegt werden, ohne daß in den Quellcode eingegriffen werden müßte. Umgekehrt muß für andere Einstellungen *swish-e* neu kompiliert werden.

Die Datei `config.h` enthält umfangreiche Kommentare und ist somit weitgehend selbsterklärend, hier sollen nur die für SSG-FI wichtigsten Einstellungen erwähnt werden.

Die folgenden Einstellungen können in `config.h` vorgenommen werden:

- Entfernung zu häufig vorkommender Wörter
- Höchstzahl der ausgegebenen Fundstellen
- die minimale und die maximale Wortlänge
- die Sonderzeichenkonvertierung
- die Buchstaben, die in Wörtern vorkommen dürfen, die am Anfang bzw. am Ende stehen dürfen, sowie die, die am Anfang bzw. am Ende entfernt werden
- ob Wörter in HTML-Tags indiziert werden

Für den SSG-FI-Bereich sind die Einstellungen so gewählt, daß eigentlich alle Wörter mit mindestens zwei Buchstaben indiziert werden, also auch z.B. die Bezeichnungen der Kategorien, die in jeder Datei vorkommen. Bei der Sonderzeichenkonvertierung werden die Buchstaben mit Umlaut, Akzent o.ä. auf ihren Grundbuchstaben zurückgeführt, also ä zu a usw. Damit können auch Benutzer anderer Tastaturen die Suche einfach durchführen.

Da der Doppelpunkt bei den erlaubten Buchstaben ausgenommen werden muß, gibt es die oben erwähnten Probleme mit den URLs, ansonsten werden z.B. Satzzeichen und Klammern am Anfang und Ende von Wörtern entfernt.

Um die Suche in den HTML-kodierten Metadaten zu ermöglichen, muß die Indizierung von Ausdrücken in HTML-Tags erlaubt werden.

Wenn das Programm kompiliert ist, können weitere Einstellung in der Konfigurationsdatei vorgenommen werden. Deren Namen ist im Prinzip beliebig, er wird als Parameter beim Aufruf von swish-e durchgereicht. Im SSG-FI-Projekt heißen diese Dateien swish-xyz.conf, wobei xyz für die Kurzbezeichnung des Faches steht.

Folgende Einstellungen werden in swish-xyz.conf vorgenommen:

- Angabe der zu indizierenden Verzeichnisse
- der Name der Indexdatei
- die Namen der zu indizierenden Meta-Tags
- die Typen und Namen der zu indizierenden bzw. zu übergehenden Dateien
- Ersetzungen für die Pfade der indizierten Dateien
- weitere Begrenzung der Indizierung zu häufiger Wörter und Einbindung einer Stoppwortliste

Die Angabe der zu indizierenden Verzeichnisse und der inhaltsleeren Dateien (z.B. .gif oder .cgi) ist selbstverständlich, ebenso die Frage der Stoppwörter. Nur die Liste der zulässigen Meta-Tags ist ungewöhnlich: hier werden die Namen der im SSG-FI-Bereich benutzten Dublin-Core- und SSG-FI-Kategorien angegeben. Insbesondere muß für andere Kategoriensysteme diese Datei angepasst werden.

Hier liegt auch der Grund, warum es Probleme mit Kategorien gibt, die sich nur im Schema unterscheiden, wie z.B. die Schlagwörter (Dc.Subject, kein Schema) und die thematische Klassifikation (DC.Subject, Scheme="GOK"). Bei der Indizierung wird nur der Name der Kategorie mit deren Inhalt indiziert, das Schema wird nicht beachtet. In SSG-FI wurden daher einige Pseudokategorien wie DC.Subject.GOK oder DC.Identifier.ISSN eingeführt, die indiziert werden und die Suche nach dem Schema „GOK“ bzw. „ISSN“ ermöglichen. Zusammen mit den erwähnten URL-Problemen war dies der Grund, eigene Indexdateien zu erzeugen, die für die Bildung des Index benutzt werden, während die Fundstellen durch eine in swish-xyz.conf eingestellte Namensersetzung auf die richtigen Infodateien zeigen.

Nach der Installation wird das Programm aufgerufen durch

```
Programmpfad/swish -c Konfigurationspfad/ swish-xyz.conf.
```

Die Bearbeitung der Suchanfrage: das CGI-Script *suche.cgi*

Um die Suche komfortabel zu gestalten, wurden die in Kapitel 4.2 erwähnten einfachen und komplexen Suchformulare gestaltet. Diese setzen die Benutzerangabe in eine Anfrage an das Suchsystem um, die an das CGI-Script we-

tergereicht wird. Dieses bereitet die Eingabe vor, leitet die Suchanfrage an *swish-e* weiter und bereitet das Ergebnis nach.

Im einzelnen führt *suche.cgi* die folgenden Schritte aus:

- die Anfrage wird eingelesen und in ihre Bestandteile aufgespalten,
- etwaige Sonderzeichen werden übersetzt,
- die Suchanfrage wird zusammengesetzt,
- die Suchanfrage wird an *swish-e* zur Bearbeitung geschickt,
- das Ergebnis wird nach den einzelnen benutzten Indizes aufgeteilt und zu einer HTML-Liste verarbeitet,
- diese Liste wird zusammen mit einem HTML-Kopf und bereichsspezifischen Angaben an den/die BenutzerIn zurückgeschickt.

Das Programm ist in Perl geschrieben und somit relativ lesbar. Etwaige Änderungen sind leicht durchführbar, da z.B. der Ausgabekopf und -fuß direkt in Klartext eingegeben wird und Perl keine Kompilierung erfordert. Das Programm prüft den SSG-FI-Bereich und setzt die Eigenschaften der Antwort entsprechend (z.B. Farbe und Namen), so daß es von allen Bereichen benutzt werden kann. Außerdem verarbeitet es sowohl die einfachen als auch die komplexen Anfragen.

6.4.2 Die technische Implementierung der Dateneingabe über Internet

Dateneingabe per HTML-Formular

Für die Benachrichtigung und die Dateneingabe über Internet wurden für jeden SSG-FI-Bereich je zwei Formulare erstellt, *templ_basic.html* und *templ_full.html*. Mit „*templ_basic.html*“ können die SSG-FI-Betreuer auf neue Sites, die in das System aufgenommen werden sollten, auf Fehler oder mögliche Verbesserungen hingewiesen werden. Es gibt in diesem Formular nur die Felder „Title“, „Description“ und „URL“, die an das SSG-FI-Team geschickt werden. Außerdem ist eine Absenderangabe möglich. Bei dem Formular des MathGuide können auch Autoren, Schlagwörter und die Fachklassifikation MSC eingetragen werden.

Das Formular „*templ_full.html*“ dient hingegen der vollständigen Eingabe eines SSG-FI-Datensatzes und hat für jede der SSG-FI-Kategorien ein Eingabefeld bzw. eine Eingabemöglichkeit. Dabei werden die kodierten Angaben wie Sprache oder Land automatisch erzeugt (Auswahl von „Germany“ erzeugt einen Eintrag „DE“), und der Formalschlüssel kann aus einer Liste der möglichen Typen ausgewählt werden. Mit diesem Formular können über das Internet mit jedem beliebigen Computer vollständige Eingaben erzeugt werden, und die Bearbeitung ist darüber hinaus etwas komfortabler als mit Allegro unter MS-DOS. Trotzdem ist die Eingabe so umfangreich und zeitaufwendig, daß nur wenige sich bereit finden werden, diese Mühe auf sich zu nehmen. Für die Zusammen-

arbeit mit anderen Partnern ist dies aber eine ausgezeichnete Möglichkeit, die Daten über das Internet an das SSG-FI-Projekt zu schicken.

Die mit den beiden Formularen an das SSG-FI-Projekt geschickten Beiträge werden von einem CGI-Script (s.u.) bearbeitet und dann in zwei Dateien abgelegt, also nicht automatisch in die Allegro-Datenbank eingebunden. Einfache Benachrichtigungen werden bearbeitet und dann gegebenenfalls der Datenbank hinzugefügt.

Bei vollständigen Datensätzen gibt es die Möglichkeit, sie alle auf einmal mit dem Programm „Import der HTML-Eingaben“ (siehe Kapitel 6.1.2) in die Datenbank aufzunehmen. Dabei werden Datensätze mit schon vorhandenen URLs in einer Datei „Doppelt.txt“ abgelegt und nicht in die Datenbank geschrieben. Andererseits ist das Format der von den Eingabefeldern erzeugten Textdateien so gewählt, daß einzelne Datensätze direkt durch „Einkleben“ (copy & paste Funktion) in die Datenbank eingefügt werden können.

Die Bearbeitung der Benachrichtigung: das CGI-Script *ssgfi.cgi*

Die mittels HTML-Formular verschickten Benachrichtigungen müssen bearbeitet werden, um sie in normalen Text zu überführen: Die Zeilen müssen wieder getrennt, Leerzeichen eingefügt und Sonderzeichen umgesetzt werden. Danach muß festgestellt werden, für welchen SSG-FI-Bereich die Meldung bestimmt ist und ob es sich um eine Benachrichtigung oder um einen Datensatz handelt. Der Inhalt der Nachricht wird dann entsprechend abgelegt. Schließlich wird eine kurze Mitteilung an den Absender der Meldung zurückgeschickt, daß seine Nachricht angekommen ist und vom SSG-FI-Team bearbeitet wird.

Diese Aufgaben werden von einem CGI-Script *ssgfi.cgi* erledigt, das die Eingaben für alle SSG-FI-Bereiche bearbeitet. Das Programm ist in Perl geschrieben und somit relativ verständlich. Neben den oben erwähnten Aufgaben bearbeitet es die Ausgabe des Textes so, daß die Zeilen das Format

„Kategorienummer <Leerzeichen> Kategorietext <Zeilenwechsel>“
haben und so direkt in die Allegro-Datenbank „eingeklebt“ werden können.

7 Inhaltliche und technische Verwaltung des SSG-Fachinformationssystems

Von Thomas Fischer, Karsten Hülsemann und Heike Neuroth

Ziel beim Aufbau der SSG-FI-Guides ist es, den wissenschaftlichen Nutzern des Internet eine effektive Übersicht über die fachlich relevanten Ressourcen zu geben. Dazu müssen im Projekt diese Ressourcen zunächst einmal *gefunden* (Kap. 7.1) und dann beschrieben werden. Die so erstellte Sammlung von Datensätzen muß *gepflegt* werden: Die URLs müssen stimmen (Kap. 7.2.1), die Beschreibungen aktuell (Kap. 7.2.2) und die wichtigsten Sites – und nur diese – erfaßt sein (Kap. 7.2.3).

Dabei stößt man wieder auf die schon erwähnten Strukturprobleme des Internet (vgl. Kap. 2.4): Die Suche nach den Quellen weist auf das Fehlen eines fachlichen Informationssystems hin, die Notwendigkeit der Pflege auf die Instabilität der URLs und allgemeiner die Variabilität der beschriebenen Sites, die durch kein integriertes Informationsinstrument (z.B. Versionsnummern) aufgefangen wird. So geht die Arbeit an den SSG-FI-Guides weit über die übliche bibliothekarische Arbeit der Erfassung bestimmter Informationsressourcen hinaus: Nicht die einmalige Erfassung, sondern die permanente Überprüfung und Pflege erfordern die Hauptarbeit bei den Subject Gateways des SSG-FI-Projektes.

Da die SSG-FI-Guides aber nicht zum Selbstzweck erstellt werden, ist es interessant zu ermitteln, ob die Adressaten der Guides denn erreicht werden. Das geht natürlich nicht, aber die Analysen der Zugriffsstatistiken des SSG-FI-Servers ermöglichen doch einige Aussagen dazu, welche Dokumente wie oft aufgerufen wurden, und damit auch darüber, ob Konzeption und Struktur der Guides angenommen werden (Kap. 7.3).

7.1 Recherchestrategien beim Aufbau der SSG-FI-Guides

Beim Aufbau eines jeden Subject Gateway stellt sich die Frage, wie dieser praktisch erfolgen soll. Das heißt, beim Suchen stellen sich die Fragen des „wie“ und des „wo“: Wo und wie sucht der „Internet Librarian“ die Ressourcen, die er/sie katalogisiert? Einmal mehr gilt, daß die traditionellen Hilfsmittel (Bibliographien der Nationalbibliotheken etc.) im Online-Bereich typischerweise noch nicht entwickelt sind.

Wenn auch die Recherchestrategien in den einzelnen SSG-Fachinformationsführern sich auf Grund der teilweise sehr heterogenen „Forschungslandschaft“ und Infrastruktur unterscheiden⁵⁶, so gibt es durchaus Gemeinsamkeiten im

56 Vgl. dazu in Teil 2 die entsprechenden Kapitel 3

Vorgehen. Eine besteht darin, daß in allen SSG-FI Systemen ein Strategie-Mix verwandt wird, in dem verschiedene Recherchestrategien nebeneinander bzw. in Kombination zu finden sind.

7.1.1 Systematische Ansätze

Eine erste Suchstrategie besteht in dem Versuch, eine „systematische“ Erschließung von Internet-Ressourcen zu betreiben. „Systematisch“ soll in diesem Zusammenhang weder implizieren, daß alternative Suchstrategien „unsystematisch“ sind, noch daß mit einem solchen „systematischen“ Ansatz der gesamte Informationsraum Internet lückenlos erschlossen werden kann. Vielmehr ist darunter zu verstehen, daß nach bestimmten formalen Kriterien *alle* Quellen eines bestimmten Typus „abgearbeitet“ und - soweit einschlägig - katalogisiert werden.

Da sind zunächst die Server derjenigen Institutionen und Organisationen, hinter denen sich berechtigterweise relevante Fachinformationen vermuten lassen. Darunter fallen die - den Fachwissenschaftlern meist namentlich bekannten - Bibliotheken, Archive, Museen, Fachgesellschaften, Forschungsinstitute etc., also jene Server, die im ersten Block des Source Type Catalog als „Information Providers“ aufgeführt sind (vgl. Kap. 5.3.1). Ein Beispiel aus der Arbeit am Anglo-American History Guide soll das daraus abgeleitete Vorgehen verdeutlichen:

Es ist plausibel anzunehmen, daß fachwissenschaftlich relevante Materialien zu anglo-amerikanischer Geschichte auch und vor allem an jenen Einrichtungen zu finden sind, an denen Forschung und/oder Lehre in diesem Bereich stattfindet: den Universitätsinstituten und Forschungszentren im anglo-amerikanischen Raum. Folgerichtig bestand eine „systematische“ Suchstrategie darin, die Internetseiten der Departments of History in den entsprechenden sechs Staaten (Großbritannien, Irland, Kanada, USA, Australien und Neuseeland) der Reihe nach aufzurufen und auszuwerten. Das konnte unter Nutzung bestehender Linklisten erfolgen, wie sie im Netz in großer Heterogenität (Qualität und Vollständigkeit betreffend) vorliegen. Idealerweise werden solche Linklisten von den zuständigen Fachgesellschaften bereitgestellt.⁵⁷

Im Ergebnis dieses systematischen Zugriffs entstand zum einen ein Verzeichnis dieser Institute, das bereits an sich für Fachwissenschaftler informativer als die bestehenden Linklisten ist.⁵⁸ Denn hier sind nicht bloße Verweise (Hyper-

57 Vgl. etwa die Übersicht der American Historical Association (AHA) (URL <http://chnm.gmu.edu/aha/pubs/directory/index.html>)

58 Vgl. den Eintrag „University Departments“ im Source Type Catalog des Anglo-American History Guide (URL http://www.SUB.Uni-Goettingen.de/ssgfi/aac-hist/aac-hist_sf1_on_en.html)

links) zu finden, sondern ein ganzer Satz von Metadaten, der das Informationsangebot des jeweiligen Department of History nach einem einheitlichen Schema beschreibt und bewertet.

Zum anderen dient das Aufrufen und Auswerten solcher Seiten als „Einfallstor“ für das Auffinden weiterer relevanter Online-Materialien. Nicht nur haben etliche Institute eigene Linklisten zusammengestellt, die dann bisweilen - um am Beispiel zu bleiben - auf historisch relevante Primärquellen verweisen, sondern es finden sich auch Informationen zu aktueller Forschung bzw. Forschungsergebnisse, die auf diese Weise veröffentlicht werden.⁵⁹

Ein anderer systematischer Ansatz der Erschließung von Ressourcen besteht darin, die für den jeweiligen Bereich einschlägigen Mailing-Listen zu nutzen, um sich z.B. über Neuerscheinungen im Internet auf dem Laufenden zu halten. Das setzt allerdings einerseits Grundkenntnisse der Informationsstrukturen voraus und ist andererseits für eine gezielte Suche nach spezifischen Informationen wenig hilfreich. Im Rahmen eines langfristigen *tracking* von Fachinformationen im Internet hingegen, wie SSG-FI es betreibt, ist ein solches Vorgehen durchaus erfolgversprechend.

7.1.2 Nutzung vorhandener Suchverzeichnisse

Es steht außer Frage, daß auch beim Aufbau der SSG-FI-Guides auf Arbeiten und Vorarbeiten von anderer Stelle zurückgegriffen wurde. Wie bereits oben gesagt, wird von vielen Seiten daran gearbeitet, eine Übersicht über das Informationsangebot im Internet zu erstellen. Einige Projekte können dabei auf eine - für Internet-Verhältnisse - lange Laufzeit zurückblicken, und auch verschiedene „Projekte“ von Einzelpersonen sind inzwischen sehr umfangreich und „etabliert“, d.h. sie werden immer wieder als Einstiegspunkte genannt. Die meisten dieser Sites haben allerdings den Charakter von reinen Linklisten. Das heißt, man erhält typischerweise nur Verweise auf weitere Ressourcen in der Form sog. Hotlinks (aktiver Hyperlinks). Diese sind manchmal annotiert, oft aber ohne jede Kommentierung, d.h. die Nutzer wissen nur selten, welcher Art und Güte die dahinter liegende Ressource ist.

Ein weiteres Problem besteht darin, daß der Pflegeaufwand für Linklisten aller Art angesichts der stetigen Veränderung im Internet sehr hoch und offenbar von vielen „Privatprojekten“ nicht zu leisten ist, so daß selten alle Links auf (noch) verfügbare Ressourcen verweisen. Gleichwohl sind solche Listen oft wertvolle Einstiege für die weitere Suche. Neben diesen Aktivitäten Einzelner finden sich Linklisten auch in institutionalisierter Form. Beispielsweise sei hier verwiesen auf „Horus' Web Links to History Resources“ an der University of California, Riverside. Obwohl sehr umfangreich, strukturiert und suchbar, bietet

59 Zu Details und den aus diesem Vorgehen gewonnenen Erfahrungen vgl. Teil 2.

Horus lediglich Links und somit keine Vorabinformationen über die dort katalogisierten Ressourcen. Zudem zeigt sich an diesem Beispiel auch die Pflegebedürftigkeit eines solchen Systems.

Als sehr gut funktionierendes Unterfangen kann hingegen das „Internet Scout Project“ am Computer Sciences Department, University of Wisconsin-Madison gelten, das in vielerlei Hinsicht wegweisend ist. Dieses bietet sowohl mit dem wöchentlich erscheinenden allgemeinen „Scout Report“ wie den 14-täglichen, themenspezifischen Reports (Social Sciences, Business & Economics, Science & Engineering) etwa zwanzig neue Ressourcen an, die in Kurzform beschrieben werden. Die im Scout Report erfaßten Ressourcen sind inzwischen über „Signpost“ suchbar.⁶⁰ Das Team des Internet Scout Project ist dabei bemüht, in jedem Scout Report eine Balance zu finden zwischen „neuen“, d.h. neu zur Verfügung gestellten, und „neu entdeckten“, d.h. bereits seit längerem verfügbaren, jedoch als wichtig erachteten, Ressourcen.⁶¹

Dabei verwendet das SSG-FI Team eine Methode, die im amerikanischen Sprachgebrauch als „citation pearl growing“ bezeichnet wird. Mit Blick auf die spezifische Anwendung dieses Suchkonzepts spricht Jack Solock, der frühere Herausgeber des Scout Report, von „Site-ation Pearl Growing“.⁶² Darunter ist zu verstehen, daß solche - von Experten erarbeiteten - Linklisten oder Kataloge die Basis darstellen, auf der die weitere Suchstrategie aufsetzt. Passender vielleicht noch als das Bild von der Perlenzucht ist das eines Schneeballsystems: das Auffinden einer ersten relevanten Quelle führt zu weiteren, so daß einzelne Themengebiete z.T. sehr schnell zu erschließen sind. Typisch für SSG-FI ist dabei eigentlich das „Vor-sich-Herschieben“ einer ganzen Reihe von Bookmarks auf Sites, die auf diese Weise entdeckt wurden. Diese werden dann der Reihe nach abgearbeitet. Ein solches Vorgehen ist Fachwissenschaftlern ebenfalls aus dem Bereich gedruckter Ressourcen bekannt. Ein gutes aktuelles Werk wird immer die bisherige Literatur berücksichtigen und aufführen, so daß von hier aus eine weitere Literaturrecherche eingeleitet werden kann.

7.1.3 Nutzung kommerzieller Suchmaschinen

Grundsätzlich ist auch der Einsatz verfügbarer Suchmaschinen vorstellbar. Die großen kommerziellen Suchmaschinen (z.B. AltaVista und Lycos) oder kombi-

60 Zur Systematik des Signpost und Kooperation zwischen SSG-FI und Scout siehe Kapitel 8.

61 Vgl. Jack Solock, „Anatomy of a Scout Report: Resource Discovery in the Information Age, or How We Do It“; *InterNIC News*, March 1997 (URL <http://wwwscout.cs.wisc.edu/scout/toolkit/enduser/archive/1997/euc-9703.html>)

62 Vgl. Jack Solock, „Site-ation Pearl Growing“, *InterNIC News*, June 1996 (URL <http://wwwscout.cs.wisc.edu/scout/toolkit/enduser/archive/1996/euc-9606.html>)

nierten Suchmaschinen und Kataloge (wie etwa Yahoo) haben allesamt Weise auf relevante Ressourcen. Die von den Suchdiensten wie Yahoo oder AltaVista bereitgestellten Listen sind für wissenschaftliche Zwecke jedoch unzureichend und können nur einen ersten Einstieg, vergleichbar den oben erwähnten Verzeichnissen, liefern. Umgekehrt ist beim Einsatz der Suchmaschinen selbst zu beachten, daß Anfragen typischerweise *zu viele* Treffer liefern. Solock spricht hier von „false drops“; das sind Treffer, die zwar die Suchkriterien formal erfüllen, aber gleichwohl für die Suche nicht relevant sind. Die „precision“, d.h. das Verhältnis von relevanten Treffern zur Anzahl der Treffer insgesamt, ist dabei oft vergleichsweise gering.⁶³ Das Dilemma ist hier also für die Recherchen des SSG-FI Teams dasselbe wie für Fachwissenschaftler (was ja auch den Ausgangspunkt für die Notwendigkeit eines SSG-FI begründete).

Nun kann das Suchergebnis im Prinzip qualitativ besser ausfallen, wenn die Suche etwas differenzierter erfolgt, was Kenntnisse in der Syntax der Suchanfragen der verschiedenen Anbieter erfordert. Eine große Rolle spielt weiterhin auch die Fachsprache der jeweiligen Disziplinen. So dürfte die Terminologie, die in der Reinen Mathematik verwandt wird, erheblich weiter vom Alltagssprachgebrauch entfernt sein als die, mit der historisch relevante Inhalte beschrieben werden. Entsprechend liefert eine Suchanfrage nach „Geomagnetismus“ qualitativ andere Ergebnisse als eine nach „romantic literature“. Auf der anderen Seite eröffnet aber die Tatsache, daß Geschichte und Literatur oft als „general interest“ gesehen werden, auch Vorteile. Regelmäßige Kolumnen, wie etwa die „Yahoo - Picks of the Week“, liefern so bisweilen auch Beschreibungen von Ressourcen, die eben nicht nur von generellem Interesse für an Geschichte oder Literatur interessierten Zeitgenossen sind, sondern auch durchaus relevant für den Historiker oder Literaturwissenschaftler sein können.

7.2 Pflege und Administration des Systems

Die Verwaltung der Daten des SSG-FI-Bestandes auf einer Allegro-Datenbank erlaubt es, das Format der HTML-Darstellung der Daten an die sich ändernden Standards des Darstellungsformats und der Metadaten anzupassen. So werden die Daten derzeit im HTML-Format 4.0 (Transitional) und die Metadaten in einer erweiterten Form des Dublin-Core-Standards („qualified“) angezeigt. Die Datenbanklösung erlaubt diese Anpassungen, ohne die Daten selbst zu verändern, nur die Exportskripte werden entsprechend modifiziert. Diese Veränderungen sind jetzt schon mehrfach erfolgt (z.B. bei HTML von 2.0 über 3.2 zu 4.0), ohne daß bei der Benutzung erkennbare Unterschiede auftreten. Ebenso können Veränderungen des Layouts vorgenommen werden, um den (vermuteten) Bedürfnissen der NutzerInnen besser gerecht zu werden. Sowohl die

63 Vgl. ebenda

Darstellung der Daten selbst als auch vor allem die Verteilerdokumente sind umgestaltet worden, um die Informationen des Systems möglich effizient verfügbar zu machen.

Während diese Anpassungen, die sich aus der Entwicklung der Standards und den Überlegungen innerhalb der Projektarbeitsgruppe ergeben, relativ leicht zu überschauen und dann konsequent durchzuführen sind, sind die Veränderungen, die sich aus den Daten selbst ergeben, viel schwerer zu überblicken. Vor allem die Veränderungen der thematischen Sites können nicht automatisch erfaßt werden, um die Daten an den neuen Stand anzupassen.

7.2.1 Prüfung der Gültigkeit von URLs

Die wichtigste Änderung, die sich bei einem thematischen Site ergeben kann, ist der Wechsel der Internetadresse. Dieses allgemein im Internet zu beobachtende Problem wird für den SSG-FI-Bereich etwas gemildert:

- Während die Adressen einfacher Dokumente sich öfter ändern oder sie auch ganz aus dem Internet verschwinden, sind die Adressen thematischer Sites relativ stabil.
- Oft stehen Organisationen wie Fachgesellschaften, Fachbereiche oder Verlage hinter einem Server, die sich um Konsistenz bemühen.
- Bei einem Adressenwechsel so großer Bereiche wird normalerweise eine Information gegeben, wie dieser Bereich jetzt zu erreichen ist.
- Oft wird eine automatische Umleitung von der alten zur neuen Adresse vorgenommen, so daß bei der Benutzung einer Adresse entweder eine kurze Meldung über die Adressenänderung angezeigt wird oder der Wechsel ganz unbemerkt vor sich geht.

Da eine falsche Internetadresse bei der Benutzung die unangenehmste Erfahrung ist, muß im SSG-FI-Projekt hier für besondere Verlässlichkeit gesorgt werden. Die Adressenprüfung kann nur automatisch mit einem Programm erfolgen, da in jedem der SSG-FI-Bereiche etwa doppelt so viele URLs wie beschriebene Sites auftreten (wegen Spiegelungen, Verweisen auf übergeordnete Server etc.). Für diese Prüfung wurden im Projekt verschiedene Programme eingesetzt, insbesondere der *Astra Site Manager* und das kostenlose Programm *Xenu's Link Sleuth*.

Der *Astra Site Manager* ist ein ziemlich umfangreiches Programm, das nicht nur die Gültigkeit von Internet-Adressen prüft, sondern den ganzen Serverbereich analysiert und eine graphische Darstellung sowie Listen der Dokumente ausgibt. Für die Prüfung der Adressen ist die Listendarstellung praktischer, da die graphische Darstellung nur durch ständige Anpassung der Vergrößerung („Ein- und Auszoomen“) die umfangreichen SSG-FI-Sites überhaupt zeigen kann. In der Listenansicht können die nicht interessierenden Adressen ausgeblendet

und so eine Liste der nicht erreichten Verbindungen angezeigt werden. Dabei gibt es verschiedene Meldungen:

- Access denied: Der Server hat eine Prüfung durch das Programm nicht zugelassen (das wird bei einigen Datenbanken gemacht, die ihre Daten gegen Einsammeln schützen).
- Inaccessible: Der Server konnte nicht erreicht werden (der Server ist abgeschaltet oder überlastet oder die Netzverbindung konnte nicht in der vorgesehenen Zeit aufgebaut werden).
- Not found: Die Datei oder der Server existieren nicht unter der angegebenen Adresse.
- Unread: Die Datei wurde nicht gelesen, entweder weil sie nicht geprüft werden sollte oder ein nicht unterstütztes Protokoll (z.B. mailto, gopher, ftp) vorliegt.

Leider wird weder bei dieser Ansicht noch in dem Report, der im HTML-Format erzeugt wird, zu den fehlerhaften Web-Adressen das HTML-Dokument angegeben, in dem diese URL vorkommt.

Xenu's Link Sleuth ist ein wesentlich kleineres Programm, das auf graphische Darstellungen verzichtet und nur die zu einer Site gehörigen Links prüft. Dazu wird eine Liste aller Links erzeugt, von denen wahlweise nur die fehlerhaften angezeigt werden. Nach der Prüfung besteht die Möglichkeit, in weiteren Durchgängen die noch nicht bestätigten Links erneut zu testen. Die Fehlermeldungen sind hier genauer als bei dem *Astra Site Manager*, es werden die Fehlermeldungen des angesprochenen Servers angegeben und in Worte übersetzt. Außerdem werden auch die Protokolle ftp und gopher geprüft.

In dem Report, den *Xenu* erzeugt, werden zu einer fehlenden Verbindung alle Dokumente aufgelistet, in denen diese benutzt wird. So kann dann durch einen Mausklick geprüft werden, wo der Fehler auftritt und der Eintrag in der Datenbank entsprechend bearbeitet werden. Wenn diese Liste abgearbeitet wird, können die schon bearbeiteten Einträge allerdings nur im HTML-Quellcode gelöscht werden, so daß im Allgemeinen mit Allegro, dem Browser und einem HTML-Editor gleichzeitig gearbeitet wird - etwas umständlich, aber effektiv.

Die Prüfung der Internet-Adressen erfolgt nicht automatisch, sondern muß von den für einen Guide Verantwortlichen regelmäßig angestoßen und der entsprechende Report abgearbeitet werden.

7.2.2 Prüfung thematischer Sites auf Veränderungen

Funktionierende Verbindungen sind unabdingbar, aber nicht ausreichend, um die Guides des SSG-FI zu einer zentralen wissenschaftlichen Informationsquelle zu machen. Im Gegensatz zu gedruckter Literatur befinden sich die in

den SSG-FI-Guides erfaßten Sites in ständiger Veränderung. Diese kann in einem beständigen Wachstum der Einträge bestehen, aber auch in dem Aufbau immer neuer Teilbereiche bis zu drastischen Umgestaltungen des gesamten Servers. Im Internet sind für solche Veränderungen keine automatisch abfragbaren Informationen vorgesehen (z.B. könnten thematische Sites mit einer Versionsnummer versehen werden). So bleibt nur die Möglichkeit, die in den Guides erfaßten Sites regelmäßig auf Veränderungen zu überprüfen. Solche Sites werden dann für einen Monat mit dem Symbol „Revisited!“ gekennzeichnet, um auf die Aktualisierung aufmerksam zu machen. Ziel ist es, alle erfaßten Sites mindestens halbjährlich zu überprüfen, das konnte im Rahmen des bestehenden Projekts allerdings nicht ganz erreicht werden.

Eine Alternative zu diesem Vorgehen würde sich ergeben, wenn die Verwalter der erfaßten Sites sich selbst an den zuständigen SSG-FI-Guide wenden würden, um auf wesentliche Veränderungen ihrer Site hinzuweisen. Dazu müßten die Guides sich als wesentliche (oder zentrale) fachliche Ansprechpartner in der wissenschaftlichen Gemeinschaft herausbilden, ein Ziel, das mit der geplanten internationalen Kooperation sicher gefördert werden kann (vgl. Kap. 8). Vorgesehen sind dafür die einfachen und die komplexen Benachrichtigungsformulare, die alle Guides anbieten, derzeit aber noch wenig genutzt werden. Ebenso wird für jeden Guide eine E-Mail-Adresse angegeben. Ungeklärt und noch nicht hinreichend geprüft ist die Frage, ob es eine realistische Möglichkeit gibt, mit den Webverantwortlichen der erfaßten Sites eine regelmäßige Kommunikation aufzubauen, um solche Benachrichtigungen zu erhalten.

Eine dritte Möglichkeit wäre, daß sich innerhalb des Internet ein System der Benachrichtigung oder Ankündigung herausbilden würde, in dem neue Versionen von Sites oder Dateien angezeigt werden. Das ist derzeit aber noch nicht in Sicht, obwohl die entsprechenden Defizite gesehen werden.

7.2.3 Vollständigkeit der thematischen Guides

Der dritte große Bereich der Pflege der SSG-FI-Guides - neben der Richtigkeit und Aktualität der erfaßten Einträge - ist deren Vollständigkeit: Sind für den entsprechenden fachlichen Bereich alle relevanten Sites erfaßt? Diese Frage ist noch schwerer zu beantworten als die Frage der Aktualität. Hier kann nur erneut auf das Fehlen entsprechender Internetstrukturen verwiesen werden sowie auf die Ausführungen in den Kapiteln 7 „Recherchestrategien beim Aufbau der SSG-FI-Guides“ und 8 „Modelle kooperativer Nutzung des SSG-Fachinformationssystems“ verwiesen werden. Die dort besprochenen Methoden bieten auch für die Systempflege die derzeit einzigen Möglichkeiten, neue Sites zu finden und sie in das SSG-FI-System zu integrieren.

Allerdings sollte nicht vergessen werden, daß im SSG-FI System Qualität vor Quantität steht: Es ist nicht Ziel, alle Informationen zu erfassen, sondern nur die

hochwertigen. Dies bedeutet aber, daß die vorhandenen Datensätze immer überprüft werden müssen und gegebenenfalls zugunsten von Datensätzen, die bessere Sites beschreiben, aufgegeben werden. In der Datenbankverwaltung ist die Kategorie #94 (vgl. Kap. 5.2) dafür vorgesehen, Datensätze, die zwar aufgenommen wurden, aber nicht mehr für die Präsentation im Netz vorgesehen sind, durch ein Stoppwort von dem Exportvorgang auszuschließen. Die Datensätze gehen auf diese Weise nicht verloren und können zu einem späteren Zeitpunkt wieder in den Guide integriert werden.

Außerdem sind dem Umfang der Guides Grenzen gesetzt. Werden die Ergebnislisten (Mini-Templates) im Source Type und Subject Catalog zu groß, so verzögern sich die Ladezeiten dieser Seiten erheblich. Als ungefähre Grenze kann gelten, daß jede Kategorie nicht viel mehr als 100 Einträge (ca. 100 kB) enthalten sollte. Da der Source Type Catalog im Gegensatz zum Subject Catalog aus wesentlich weniger Kategorien besteht, ist dort die Grenze sehr schnell erreicht. Auch aus diesem Grund wurden einige Kategorien des 1. Blocks im Source Type Catalog (Information Providers/Servers, vgl. Kap. 5.3.1) weiter länderspezifisch unterteilt. Dies verbesserte nicht nur die Übersichtlichkeit der Ergebnislisten, sondern es können so pro Land auch ca. 100 Sites aufgenommen werden. Auch die Kategorie Journals im GeoGuide mußte weiter spezifiziert werden, als über 100 Journals aufgenommen worden waren (Free Full Text Articles Available, With Selected Articles etc., vgl. Teil 2: GeoGuide). Allerdings wird sich nicht jede Kategorie sinnvoll weiter unterteilen lassen, sobald die Ergebnislisten zu umfangreich werden sollten.

7.3 Statistische Analysen

Die Internetverbindungen eines Servers werden im allgemeinen registriert, dazu werden sogenannten Logdateien angelegt. Auch der SSG-FI-Server führt solche Logdateien, und zwar vier verschiedene: den Zugriffslog (access log), den Verweislog (referrer log), den Browserlog (agent log) und den Fehlerlog (error log). Diese Logdateien stehen zur statistischen Auswertung zur Verfügung. Die verschiedenen Dateien dienen verschiedenen Zwecken.

Die *Fehlerlogdatei* wird ausgewertet, um Zugriffsprobleme zu beheben. So stellte sich zum Beispiel heraus, daß alte Dokumente noch nachgefragt wurden, die im Zuge einer Umstellung der Dateistruktur entfernt worden waren. Für diese Dokumente wurden nun Hinweisdokumente eingesetzt, die bei Aufruf Informationen zur geänderten Struktur des Servers liefern.

Die *Browserlogdatei* wird benutzt, um eine Übersicht darüber zu bekommen, mit welchen Browsern auf den SSG-FI-Bereich zugegriffen wird⁶⁴. In der Rei-

64 Daß einige Browser dabei über ihren Namen „lügen“ ist in diesem Zusammenhang nicht wichtig.

henfolge zunehmender Komplexität können HTML-Dateien *Text*, *Bilder*, *Tabellen*, *Rahmen*, *Java*, *JavaScript* und *Stylesheets* enthalten, aber nicht alle Browser sind in der Lage, diese zu verarbeiten. Bei dem Design des SSG-FI-Servers wurde versucht, die Dateien übersichtlich und nicht zu komplex zu gestalten, trotzdem werden Tabellen, Bilder, Rahmen und Stylesheets benutzt. Die Analyse des Browserlogs zeigt nun, daß nur von einer verschwindend kleinen Minderheit Browser benutzt werden, die diese Strukturen gar nicht darstellen können - nur die Stylesheet-Verarbeitung ist nicht immer gewährleistet, das führt aber nur zu ästhetischen Problemen. Nicht rahmenfähige Browser werden automatisch zu rahmenfreien Dokumenten geführt – sie würden sonst nämlich gar nichts sehen.

Die *Verweislogdatei* gibt Auskunft darüber, welche Wege die Benutzer in dem SSG-FI-Bereich einschlagen, insbesondere von welcher Datei sie den Einstieg in SSG-FI gefunden haben. Es zeigt sich, daß viele Zugriffe von einer der allgemeinen Suchmaschinen ausgehen, und zwar eher im Bereich des „Anglo-American History Guide“, wo sich viele Schlagwörter von allgemeinem Interesse finden, als bei dem Math- oder GeoGuide, wo die Fachsprache in stärkerem Maße vorherrscht.

Schließlich ist natürlich die *Zugriffslogdatei* von Interesse, in der abgespeichert wird, welche Dateien von welchen Nutzern angefordert werden. Dies ist bei näherer Betrachtung aber komplizierter, als es zunächst erscheint.

Auf der Dateiseite stehen dabei die Bilddateien im Vordergrund, weil jedes HTML-Dokument des SSG-FI-Servers mehrere Bilder als Tasten oder zur Seitengestaltung enthält, dies sind GIF-Dateien im Umfang von etwa 1 - 2 kB. Außer diesen Bilddateien werden fast ausschließlich HTML-Dateien nachgefragt – die Dateien, die die Informationen der verschiedenen Guides beinhalten. Ein kleiner Prozentsatz der Dateien schließlich sind Anfragen an die Suchmaschine, also Aufrufe des CGI-Scripts *suche.cgi*. Von dieser Seite gesehen läßt sich also sagen: wenn man die Nutzung des SSG-FI-Dienstes analysieren will, so sind nur die HTML-Dateien und die Suchanfragen zu beachten, die Bilder sind nicht von Belang.

Auf der anderen Seite stehen die Benutzer, die diese Seiten anfragen. Diese können zunächst grob in drei Gruppen eingeteilt werden:

- Das Team des SSG-FI-Projekts selbst greift natürlich ständig auf die eigenen Seiten zu,
- automatische Indizierungsmaschinen erfassen den SSG-FI-Bereich für die großen und kleinen Suchmaschinen des Internet,
- die eigentlichen Nutzer, die aus allgemeinem oder fachlichem Interesse den SSG-FI-Bereich aufsuchen.

Für die Nutzeranalyse kommt nur die letzte Gruppe in Betracht, so daß die ersten beiden aus den Logdateien ausgefiltert werden. Das geht mit der ersten Gruppe recht vollständig, aber die zweite Gruppe ist nicht einfach zu erkennen, eine Liste bekannter Namen wird aber gepflegt und bei der Analyse berücksichtigt, d.h. herausgefiltert. Leider werden mit der ersten Gruppe auch die Benutzer der öffentlichen Computer der SUB ausgefiltert, die eigentlich zur dritten Gruppe gehören.

So bleiben dann die „echten“ Nutzer übrig – aber hier setzen die Probleme erst richtig ein. Diese sind so schwerwiegend, daß ein amerikanischer Netzadministrator sogar meint, daß „Nutzerstatistiken schlechter als gar keine Statistiken“⁶⁵ seien. Das liegt vor allem daran, daß für den Internetbereich ein komplexes System von Zwischenspeichern genutzt wird, damit die Dateien nicht immer um den Erdball herum transportiert werden müssen. Das fängt bei dem Cache des Browsers an, führt über Proxy-Server und regionale und nationale Speicher zu den Problemen der Spiegelung der Dateien an verschiedenen Orten.

So ist bei der Analyse nicht festzustellen, welche Dokumente wie oft nachgefragt werden, sondern nur, daß einige Dokumente von *vermutlich privaten Nutzern* nachgefragt werden. Trotzdem soll eine statistische Analyse versucht werden, da nur mit ihr ein – wenn auch vorläufiger und unvollständiger – Eindruck gewonnen werden kann, in welchem Maße die SSG-FI-Guides von der wissenschaftlichen Gemeinschaft als Informationsmedium angenommen werden.

7.3.1 Generelles

Die Auswertung der Logdateien für das SSG-FI Projekt erfolgte für den Zeitraum vom 16.12.1997 bis zum 13.11.1998 durch das Programm *analog 3.0*, das Programm ist von Stephen R. E. Turner entwickelt worden und im Internet kostenlos erhältlich⁶⁶. Das Programm wertete alle Anfragen an die SSG-FI Seiten während des genannten Zeitraumes aus. Dabei werden in der Konfigurationsdatei verschiedene Filteroptionen eingebaut, so sind eigene Anfragen von SUB-Rechnern aus (*SUB-Filter*), Anfragen von automatischen Indiziermaschinen (*Robot-Filter*) und die vorkommenden Grafikformate (*Grafik-Filter*: gif, jpg etc.) herausgefiltert worden.

Es gibt im SSG-FI-Projekt ca. 4200 verschiedene Dateien, die sich auf HTML-Dateien und Grafikdateien (Format: gif) verteilen. Bei den meisten Dateien handelt es sich um die Info-Datensätze, die von der Zahl der Datensätze in den

65 Jeff Goldberg, *Why web usage statistics are (worse than) meaningless*. URL <http://www.cranfield.ac.uk/stats/>

66 Statistical Laboratory, Cambridge (URL <http://www.statslab.cam.ac.uk/~sret1/>, E-Mail: sret1@cam.ac.uk)

Datenbanken der jeweiligen Guides abhängt (Stand Mitte November 98: AAC-History Guide: 750, AAC-Literature Guide: 450, GeoGuide: 1100; MathGuide: 800).

Anfragen von Rechnern der SUB Göttingen an die SSG-FI Seiten sind herausgefiltert worden (SUB-Filter: .sub.uni-goettingen.de, 134.76.161.*, 134.76.162.*, 134.76.163.*). Damit fallen die Anfragen, die vom SSG-FI Team und anderen SUB-Mitarbeitern an die Projektseiten gestellt werden, aus der Statistik heraus, leider aber auch diejenigen, die von den Benutzern der öffentlichen Rechner in der SUB gestellt wurden. Insbesondere werden damit die automatisierten Anfragen der im SSG-FI Projekt genutzten URL-Checker entfernt (*Astra-Site Manager*, *Xenu*, vgl. Kap. 7.2.1). So kann aus dem Host-Bericht genau abgelesen werden, von welchen Rechnern des SSG-FI Teams aus wieviel Anfragen insgesamt erfolgten (z.T. über 30.000) und an welchen Tagen sich diese konzentrieren. Anfang November gab es eine deutlichen Spitze, die darauf zurückzuführen ist, daß das URL-Prüfprogramm *Xenu* eingeführt wurde und mit dem bis dato verwendeten Programm *Astra Site Manager* ausführlich verglichen wurde. Dies bedeutet, daß mehrmals in diesem Zeitraum von verschiedenen SSG-FI Rechnern aus für die verschiedenen Guides Anfragen an das komplette SSG-FI System erfolgten, um externe Links zu überprüfen.

Suchanfragen von Robots laufen automatisch ab, sie erreichen beim SSG-FI-System mehreren tausend Anfragen und dauern einige Zeit. Robot-Anfragen fallen in der Regel auf Grund der hohen Seitenanfrage und der längeren Abfragezeit in der Auswertung auf und können gezielt herausgefiltert werden.

Folgende Robot-Anfragen sind herausgefiltert worden (Robot-Filter):

171-80-233.ipt.aol.com	j2002.inktomi.com
crawl3.atext.com	j2005.inktomi.com
crawl4.atext.com	j2009.inktomi.com
crawler.crl.research.digital.com	lycosidae.lycos.com
crawler2.anzwers.ozemail.net	nat-20-157.wisewire.com
crimpshrine.atext.com	netmechanic2.com
devil.schlund.de altavista.matav.net	pyrobotrys2.bio.botanik.uni-goettingen.de
firestone.alex.com	rs7.loc.gov
gerhard.bis.uni-oldenburg.de	scooter.pa-x.dec.com
green.alex.com	scooter3.av.pa-x.dec.com
heavymetal.fireball.de	search.pta.at
hera.bis.uni-oldenburg.de	soap.av.pa-x.dec.com
j12.inktomi.com	sp5.loc.gov
j16.inktomi.com	

spider.funet.fi	spider-td093.proxy.aol.com
spider.ge.man.ac.uk	spider-td112.proxy.aol.com
spider.surnet.ru	spider-to09.proxy.aol.com
spider.usm.k12.wi.us	spider-to15.proxy.aol.com
spider1.northgrum.com	spider-wc014.proxy.aol.com
spider1.wisewire.com	spider-wc074.proxy.aol.com
spider1-dx.wisewire.com	spider-wc083.proxy.aol.com
spider2.wisewire.com	templeton-bbn.infoseek.com
spider2-uu.wisewire.com	trousers.muscat.com
spider3.wisewire.com	vscooter.av.pa-x.dec.com
spider-tc014.proxy.aol.com	wilbur-bbn.infoseek.com
spider-td063.proxy.aol.com	www.netmechanic2.com

Diese Art Anfrage ist keine echte Nutzung der Guides durch z.B. Fachwissenschaftler und verfälscht deshalb die Statistik. Die Anfragen an Grafiken gehen ebenfalls nicht in die statistischen Berechnungen ein (Grafik-Filter). Da in jede HTML-Seite mehrere Grafiken eingebunden sind und somit mehrmals pro Sitzung eines Nutzers aufgerufen werden können, vervielfältigen sich einige Grafik-Dateien in der Anfragestatistik um ein beträchtliches. Folgende SSG-FI Grafiken sind am häufigsten mit Anforderung einer HTML-Seite aufgerufen worden:

Grafikdatei (gif-Format)	Anfragen insgesamt (332 Tage)	Durchschnittliche Anfrage pro Tag
/ssgfi/grafiken/std-up.gif	55500	167
/ssgfi/grafiken/std-home.gif	54292	164
/ssgfi/grafiken/std-left.gif	54168	163
/ssgfi/grafiken/std-right.gif	53613	161
/ssgfi/grafiken/std-such.gif	50395	152
/ssgfi/grafiken/www.gif	48010	145
/ssgfi/grafiken/rainbow.gif	46618	140

Tab. 7.1: Tabelle der am häufigsten geladenen Grafiken

Allein schon die in Tab. 7.1 aufgelisteten Grafiken würden die durchschnittlichen täglichen Anfragen der SSG-FI Seiten um 1092 erhöhen. Da im gesamten Projekt ca. 70 unterschiedliche Grafiken in den HTML-Seiten eingebunden sind, erklärt sich der in Abb. 7.1 hohe prozentuale Zuwachs von 460 % zu den täglichen Anfragen an das SSG-FI System, wenn bei der Auswertung kein Grafik-Filter eingebaut wird. Da bei einer Auswertung ohne Grafik-Filter ca. 78 % aller

Dateitypen nach Tab. 7.2 Grafiken zugeordnet werden können, verwundert der hohe prozentuale Zuwachs nicht.

Dateityp	Mit Grafik-Filter		Ohne Grafik-Filter	
	Anzahl	Anteil	Anzahl	Anteil
HTML	149.163	87,5 %	149.163	19 %
Verzeichnisse	21.157	12,4 %	21.157	2,7 %
Grafiken	-	-	612.811	78,2 %
nicht angezeigt	155	0,1 %	155	0,1 %

Tab. 7.2: Prozentuale Verteilung der Dateitypen mit und ohne Grafik-Filter

Abb. 7.1 faßt die Komplexität zusammen, die bei statistischen Auswertungen auftreten kann. Obwohl in den SSG-FI Seiten nicht übermäßig viele Grafiken eingebunden sind, würde sich doch ohne Grafik-Filter die tägliche Seitenanfrage um 460 % erhöhen. Werden die eigenen Anfragen nicht herausgefiltert (SUB-Filter), so erhöht sich die tägliche Seitenstatistik um 172 %. Dies ist natürlich auch von Server zu Server unterschiedlich, bei dem SSG-FI Projekt würde sich u.a. auch deshalb ein so hoher Prozentsatz ergeben, da in diesem System ein URL-Check unerlässlich ist. Die Berücksichtigung der Robot-Anfragen würde zu einem Zuwachs um 115 % führen. Da es sich dabei vermutlich ausschließlich um Textanfragen handelt, fällt der Zuwachs im Verhältnis zu den anderen Filtern geringer aus.

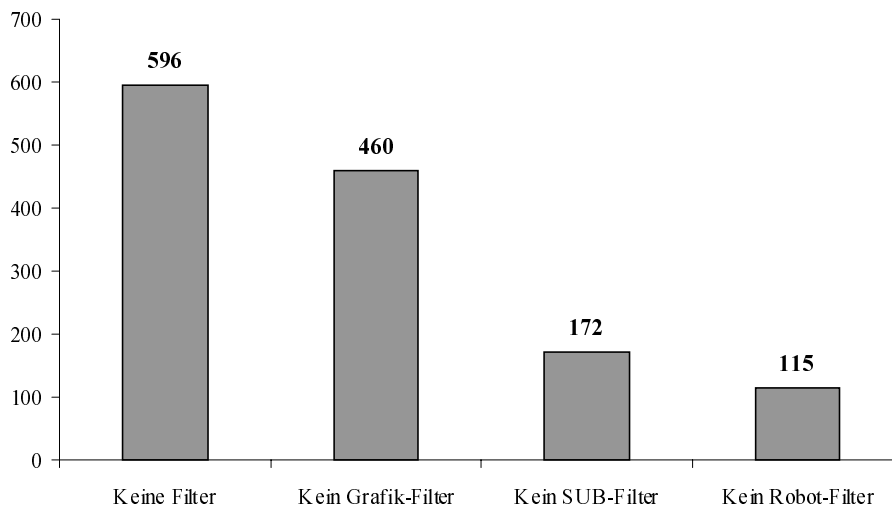


Abb. 7.1: Zunahme der täglichen Seitenabfragen in Abhängigkeit von den Filtern

Wird das Programm *analog* gänzlich ohne Filter benutzt, so würden sich die **täglichen** Zugriffsraten um 600 % erhöhen. Insgesamt 2538 Anfragen pro Tag gehen durch die Filter-Einstellungen in der Konfigurationsdatei nicht in die statistischen Analysen ein. Abb. 7.2 stellt die prozentuale Verteilung dar, auf welche Filterprozesse wieviel Prozent der herausgefilterten Anfragen entfallen. Danach entfallen auf den Grafik-Filter Prozeß 82 % der herausgefilterten Dateien.

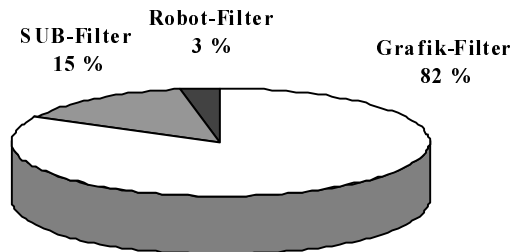


Abb. 7.2: Prozentuale Verteilung der herausgefilterten Anfragen auf die Filter

Da bei den meisten Servern, die Angaben über die Zugriffsstatistik auf ihrer Seite geben, nicht nachzuvollziehen ist, ob und welche Anfragen herausgefiltert wurden, wird im SSG-FI Projekt die Kategorie *Server Statistics* nicht mehr benutzt. Es lassen sich nur dann mehrere statistische Angaben miteinander vergleichen, wenn sie derselben Auswertungs-Methode unterliegen. Allerdings kann im Internet in der Regel davon ausgegangen werden, daß hohe Zugriffsraten erwünscht sind und deshalb nicht jeder Server-Betreiber seine Analyse einem Filterprozeß unterzieht.

Die folgende Auswertung wurde einem rigorosen Filter-Prozeß unterzogen, so daß Anfragen an Grafiken, eigene Anfragen und die der Robots nicht mit in die Statistik eingehen. Aus praktischen Gründen wurden Seiten mit weniger als 20 Anfragen nicht berücksichtigt. Dies sind vor allem einzelne Metadaten-Templates, die wahrscheinlich noch nicht lange in dem jeweiligen Guide enthalten sind.

7.3.2 Statistische Auswertung und Nutzerprofile am Beispiel der SSG-FI-Nutzung 1998

Die SSG-FI Server Statistik liegt für den Zeitraum vom 16.12.1997 bis zum 13.11.1998 (332 Tage). Insgesamt erfolgten 170.476 Anfragen an die gesamten SSG-FI Seiten bei durchschnittlich **512 Anfragen pro Tag** (vgl. Tab. 7.3).

Durchschnittlich bearbeitete Seitenanfragen pro Tag	512
Erfolgreich bearbeitete Seitenanfragen insgesamt	170.476
Anzahl unterschiedlich verlangter Dateien	4.241
Anzahl unterschiedlich anfragender Hosts	42.678
Durchschnittliche Menge verschickter Daten pro Tag [kByte]	4.577

Tab. 7.3: Server-Statistik SSG-FI, Übersicht

Die Zahl der Hosts liegt mit ca. 42.000 sehr hoch. Dies deutet darauf hin, daß die SSG-FI Seiten von einem breiten Publikum besucht werden. Da die Auswertung für den oben genannten Zeitraum keine Aussage über die aktuellen Zugriffsraten erlaubt, wurde die Zeit vom 30.09.1998 bis 13.11.1998 (44 Tage) separat analysiert. Danach gab es **durchschnittlich 1062 Anfragen pro Tag** an die SSG-FI Seiten.

Da in den Anfangsmonaten der Bekanntheitsgrad der Guides naturgemäß nicht sehr hoch war, reduziert sich bei einer 11-monatigen Berechnungsgrundlage der tägliche Durchschnittswert der Seitenanfragen beträchtlich. Für die weitere statistische Analyse wird allerdings der lange Berechnungszeitraum zu Grunde gelegt, um eine möglichst repräsentative Auswertung zu erhalten.

Tab. 7.4 gibt die Anfrage-Statistik für die Monate Januar bis Oktober wieder. Der Dezember 1997 und der November 1998 wurden nicht mit berücksichtigt, da sie nur zur Hälfte ausgewertet wurden. Die Monate Juli und September zeigen mit über **35.000 Anfragen** ein deutliches Anwachsen gegenüber den Monaten Anfang des Jahres.

Januar	1894 Anfragen
Februar	4373 Anfragen
März	5289 Anfragen
April	10.207 Anfragen
Mai	8913 Anfragen
Juni	7669 Anfragen
Juli	32.630 Anfragen
August	17.004 Anfragen
September	36.275 Anfragen
Oktober	28.852 Anfragen

Tab. 7.4: Server-Statistik SSG-FI, Monatsberichte 1998

Tab. 7.5 faßt die Ergebnisse für die einzelnen Monate zusammen. Ein erstes deutliches und beständiges Anwachsen der Anfragen ist im April mit ca. 3000 zu beobachten. Ein weiterer deutlicher Schub erfolgte im Juli mit Wochen, in

denen die Anfragen deutlich zwischen 7.000 und 10.000 liegen. Im August flaut es dann wieder etwas ab (urlaubsbedingt?), um im September sich in einigen Wochenabschnitten der 10.000 Marke zu nähern. Während sich in der 1. Hälfte des Jahres die Zugriffe erst allmählich steigern, so ist ca. ab der 2. Jahreshälfte (Anfang Juli) ein deutlicher Anstieg zu beobachten.

Aus dem in Tab. 7.6 vorgestelltem Wochentagsbericht geht, wie nicht anders zu erwarten war, hervor, daß sich die Anfragen auf die SSG-FI Seiten auf die Wochentage Montag bis Donnerstag konzentrieren. In der Summe ergeben die Anfragen von Samstag und Sonntag aber eine genau so hohe Zahl wie die einzelnen Wochentage Montag bis Donnerstag. Vermutlich ist dies darauf zurückzuführen, daß doch eine erhebliche Anzahl Nutzer aus dem privaten Bereich auf die Guides zugreift.

Sonntag	14.193 Anfragen
Montag	28.462 Anfragen
Dienstag	30.127 Anfragen
Mittwoch	29.347 Anfragen
Donnerstag	28.825 Anfragen
Freitag	24.603 Anfragen
Samstag	14.778 Anfragen

Tab. 7.6: Server-Statistik SSG-FI, Wochentagsbericht 1998

Der Bereichsbericht („domains“) ist in der Abb. 7.3 zusammengefaßt und gibt die Zugriffszahlen aufgeschlüsselt nach Ländern bzw. Bereiche (.com, .edu, .gov etc.) wieder.

Danach erfolgen die meisten Anfragen auf die SSG-FI Seiten von kommerziellen Servern aus (46.762), wobei über 17.700 Anfragen Zugriffe von Privatrechnern über AOL darstellen, die sich länderspezifisch nicht weiter unterscheiden lassen. Darauf folgen Anfragen von Servern aus Deutschland (28.638, z.B. Uni Hannover, Uni Würzburg, Uni Duisburg, Uni Bonn, Uni Stuttgart, Uni Hamburg, DBI Berlin etc.). Das SSG-FI Projekt findet allmählich auch internationale Akzeptanz, worauf Anfragen aus den USA von Educational Servern mit 9.995 Anfragen hinweisen. Erfreulich sind auch die hohen Zugriffsraten von Rechnern aus Großbritannien (5066), Kanada (3.059) und Australien (3.477). Durchschnittlich sind dies immerhin 10 - 30 Anfragen pro Tag an die SSG-FI-Seiten.

Die restlichen Anfragen verteilen sich mehr oder weniger gleichmäßig auf die restlichen europäischen Länder (Österreich, Schweden, Schweiz, Niederlande etc.) und auf Länder außerhalb Europas wie Neu-Seeland, Japan, Israel, Brasilien, Süd Afrika etc. Verblüffend sind die Zugriffsraten von USA-Regierungs-Servern mit insgesamt 789 Anfragen und USA-Militär mit insgesamt 562 Anfragen.

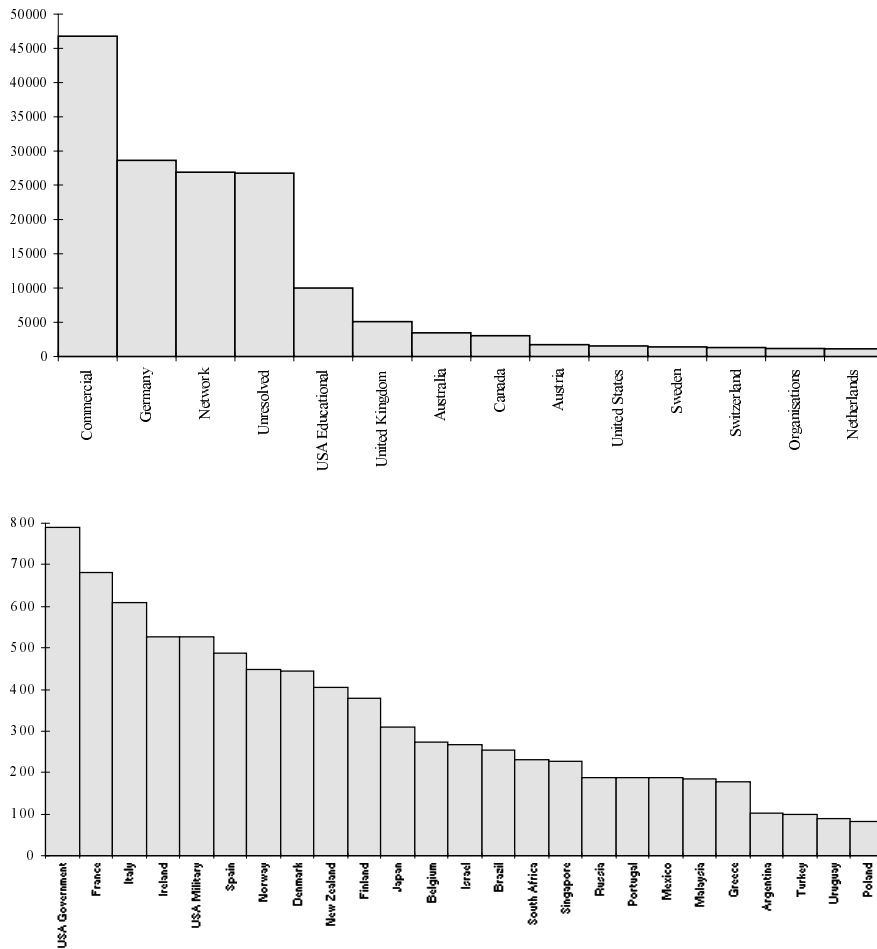


Abb. 7.3: Bereichsbericht (unterschiedliche Skalierung!)

Die Statistik über die Ausgabe aller verlangten Dateien (vgl. Abb. 7.5) besagt, daß die GeoGuide Einstiegsseite die höchsten Zugriffsraten zu verzeichnen hat, gefolgt von den Einstiegsseiten von SSG-FI und des MathGuide. Danach sind der Subject Catalog und die einfache Suche des GeoGuide die Seiten, die am meisten besucht werden. Dies bedeutet, daß der GeoGuide bisher der am stärksten frequentierte Guide ist (vgl. Abb. 7.5). Bei allen Guides wird der Source Type Catalog (Formal) erst nach anderen relevanten Seiten (Subject Catalog, Simple Search) besucht.

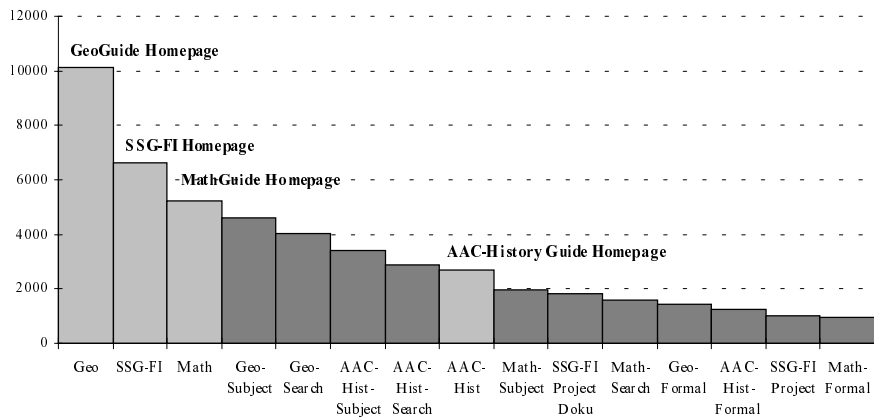


Abb. 7.4: Reihenfolge der 15 gefragtesten SSG-FI Seiten

Bei der Betrachtung der prozentualen Verteilung der Anfragen auf die Guides (Abb. 7.5) muß berücksichtigt werden, daß der AngloAmericanCulture - Literature Guide (AAC-LitGuide) erst im September 1998 der Öffentlichkeit zugänglich gemacht wurde und so noch keine große Verbreitung gefunden haben kann. Zudem geht er bei dem Auswertungs-Zeitraum ab Dezember 1997 in der Statistik unter.

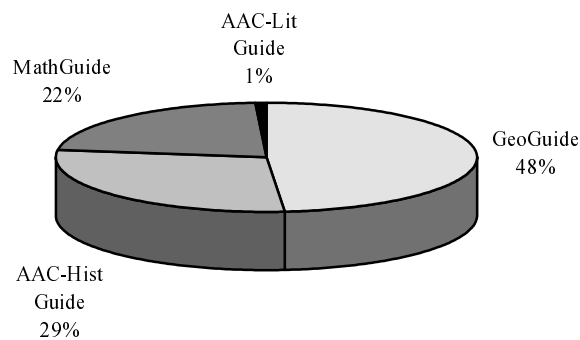


Abb. 7.5: Prozentuale Verteilung aller Anfragen auf die jeweiligen Guides

Abb. 7.6 verdeutlicht, daß von den allgemeinen SSG-FI Seiten die Einstiegseite am häufigsten verlangt wird. Aber auch die allgemeine Projektseite (Project) mit der Projekt-Dokumentation (Doku), der Kontaktseite (Contact), der Vortragsseite (Present), der Beitragsseite (Form) und dem Nachnutzungsvermerk werden aufgerufen. Die doch relativ hohen Zugriffsraten auf die Projekt-

Dokumentationsseite verdeutlichen, wie wichtig sie den Nutzern ist und daß sie regelmäßig gepflegt und ergänzt werden sollte.

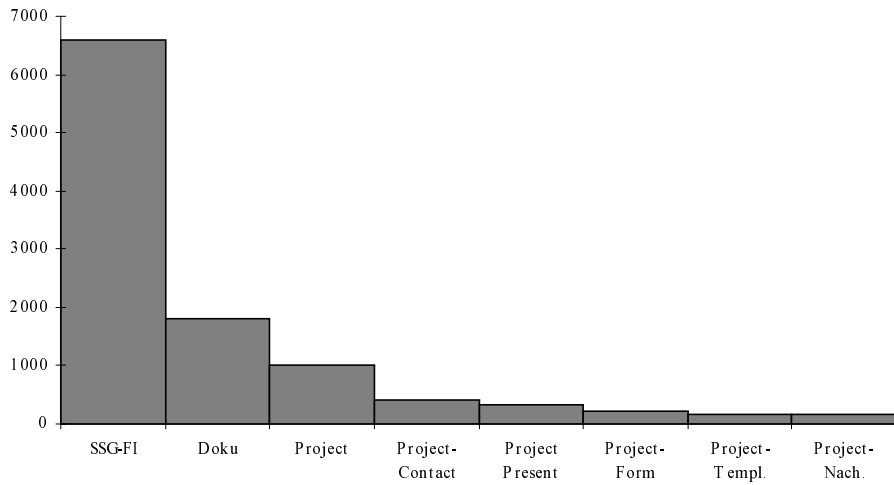


Abb. 7.6: Reihenfolge der allgemeinen SSG-FI Seiten

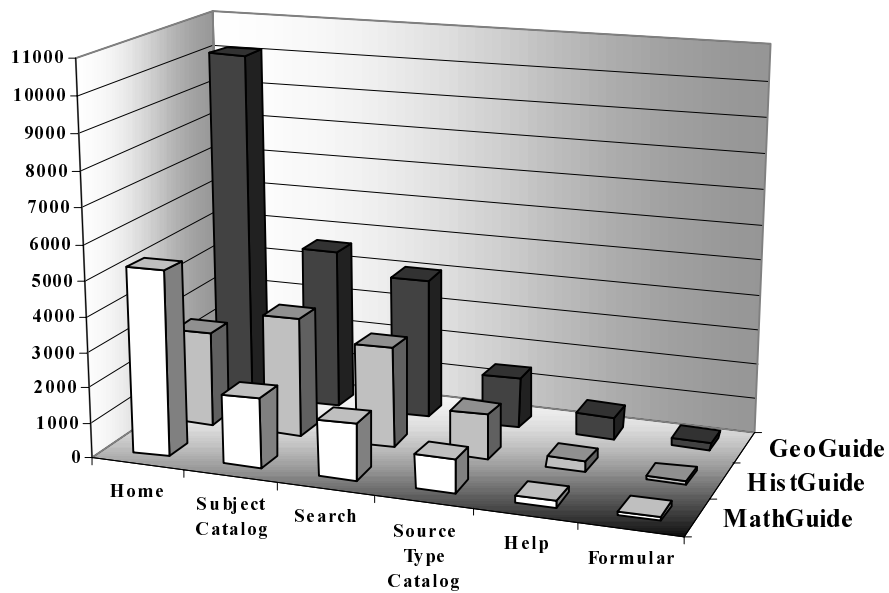


Abb. 7.7: Reihenfolge der Guide-Seiten

Bei den einzelnen Guides zeigt Abb. 7.7 den schon oben genannten Trend, daß der Source-Type Catalog (Formal) bei den Nutzern erst nach dem Subject Catalog und der Suche eine Rolle spielt. Besonders deutlich wird dies beim GeoGuide, der beim Source Type Catalog ähnlich niedrige Zugriffsraten verzeichnet wie der MathGuide und AAC-Hist Guide, obwohl er bei den anderen GeoGuide Einstiegsseiten wie Search- und Subject Catalog Seite wesentlich höhere Anfragen hat.

Die Hilfedatei wird bei allen Guides relativ oft benutzt, obwohl sie z.T. nur in deutscher Fassung vorliegt. Zudem ist sie auch erst seit wenigen Monaten veröffentlicht. Auffällig beim AAC-History Guide ist, daß die eigentliche Einstiegsseite in den Zugriffszahlen erst nach den Einstiegsseiten des Subject Catalog und der Suche auf Platz 3 steht. Dieses Phänomen könnte damit zusammenhängen, daß mehr Nutzer über kommerzielle Suchmaschinen zu dem Guide finden und direkt auf die Metadaten-Templates verwiesen werden (s. u.).

Einen Überblick über die genutzte Bandbreite der einzelnen Kategorien des Subject Catalog und Source Type Catalog inklusive der möglichen ausführlichen Infodateien (Metadaten-Templates) gibt Tab. 7.7. Darin wird die Zahl der Kategorien, die über 20 mal nachgefragt worden sind, mit der Zahl der vorhandenen Kategorien verglichen.

	Abfrage ins- gesamt	Aufgerufene Kategorien	Vorhandene Kategorien	Prozentuale Nutzung
AAC-History Guide				
Source-Type Catalog	829	18	45	40 %
Subject Catalog	1.514	40	186	22 %
Infodata (= Template)	22.061	275	747	37 %
GeoGuide				
Source-Type Catalog	2.859	49	86	57 %
Subject Catalog	8.327	140	174	80 %
Infodata (= Template)	26.217	408	1.000	41 %
MathGuide				
Source-Type Catalog	1.110	27	50	54 %
Subject Catalog	2.881	65	86	76 %
Infodata (= Template)	12.725	326	804	41 %

Tab. 7.7: Abfragedichte der Verteilerdokumente und der Infodateien

Tab. 7.8 bietet einen Überblick über die drei gefragtesten Kategorien des Source-Type und Subject Catalogs und der Info-Datensätze (Template) der jeweili-

gen Guides. Die erste Spalte gibt die Gesamt-Anfrage für die jeweilige Kategorie bzw. Info-Datensatz wieder, die 2. Spalte die abgeschnittene URL und die letzte Spalte schlüsselt die Kategorie des Source-Type und Subject Catalogs auf bzw. gibt den Titel der URL des Info-Datensatzes wieder.

Beim AAC-History Guide fällt auf, daß die Summe der abgefragten Seiten des Source Type und Subject Catalogs (Tab. 7.7, Spalte 1) im Vergleich zu den anderen Guides relativ niedrig ist, hingegen die Anzahl der aufgerufenen Templates mit 22.061 relativ hoch ist. Obwohl die Seitenabfrage bei den Templates hoch ist, so sind es wenige verschiedene Templates mit sehr hohen Zugriffsraten (vgl. Tab. 7.8). So sind von 747 möglichen Templates (Anzahl der Datensätze im Guide, Stand Anfang November) nur 275 aufgerufen worden, diese dann allerdings z.T. sehr häufig. Auch das Angebot des Browsens im Source Type oder Subject Catalog wird durch den Nutzer nicht ausgeschöpft. Nur 40 % der im Source Type zur Verfügung stehenden Kategorien wurden aufgerufen, im Subject Catalog sind es sogar nur 22 %.

AAC-History Guide:

Source Type Kategorien/Titel

201 Anfragen	Maps
79 Anfragen	Libraries
63 Anfragen	Archives

Subject

94 Anfragen	Colonialism and Imperialism
92 Anfragen	History of the United States
87 Anfragen	USA 1776-1918: Civil War

Info-Datensätze

2241 Anfragen	U.S. Immigration and Naturalization Service (INS)
1550 Anfragen	Texas Maps
989 Anfragen	Australian Bureau of Statistics

GeoGuide:

Source-Type Kategorien/Titel

261 Anfragen	Maps
193 Anfragen	(Collections of) Pictures, Photos
151 Anfragen	Lists of Periodicals/Journals

Subject

650 Anfragen	Maps: North America
205 Anfragen	Maps: Africa
182 Anfragen	History of Carthography

Info-Datensätze

1759 Anfragen	World Road Atlas (Mapquest)
702 Anfragen	Africa: Geographic Maps
653 Anfragen	Field Museum of Natural History, Chicago

MathGuide:

Source-Type

Kategorien/Titel

102 Anfragen	Series of Preprints, Reports, and Working Papers
90 Anfragen	Journals
76 Anfragen	Resaearch Institutes

Subject

147 Anfragen	History and biography
127 Anfragen	Mathematical logic and foundations
115 Anfragen	Statistics

Info-Datensätze

232 Anfragen	The National Council of Teachers of Mathematics
202 Anfragen	Princeton Institute for Advanced Study, School of Mathematics
189 Anfragen	London School of Economics and Political Science; Department of Statistics

Tab. 7.8: Die drei meistgefragten Kategorien und Infodatensätze der einzelnen Guides

Diejenigen Kategorien, die bevorzugt beim Browsen im AAC-History Guide gewählt wurden, sind in Tab. 7.8 aufgelistet. Danach werden am häufigsten die Kategorien *Maps*, *Libraries* und *Archive* des Source Type Catalogs besucht, im Subject Catalog die Kategorien *Colonialism and Imperialism* (GOK PFM 000), *History of the United States* (GOK PVK 000) bzw. *USA 1776-1918: Civil War* (GOK PVM 245), also eine Konzentration auf die Geschichte der USA. Sehr auffällig ist die hohe Zugriffsrate auf den Infodatensatz *U.S. Immigration and Naturalization Service (INS)* mit über 2.200 Treffern. Auch die beiden nachfolgenden Infodatensätze liegen mit insgesamt 1550 bzw. 989 Zugriffen im Vergleich zu anderen Infodatensätze des MathGuide oder GeoGuide sehr hoch.

Eine mögliche Erklärung könnte darin liegen, daß diese Sites sehr populär sind und von den Nutzern über Suchmaschinen wie AltaVista gefunden werden. Dabei könnten sie an das SSG-FI System verwiesen werden. Das steht im Einklang mit der Beobachtung, daß diese Datensätze wesentlich häufiger als die zugehörigen Kategorien des Source Type bzw. Subject Catalog nachgefragt werden. Die Anzahl der Abfragen zeigt, daß der Aufruf auch nicht über die Einstiegseite oder die Suchfunktionen erfolgt sein kann.

Insgesamt kann aus der Analyse des Nutzerverhaltens für den AAC-History Guide geschlossen werden, daß viele Nutzer indirekt über Suchmaschinen den Weg zum AAC-History Guide finden und dort auf die Beschreibung des von ihnen gesuchten Site stoßen. Die drei meist frequentierten Metadaten-Templates gehören zu sehr populären Sites, wie die Backlink-Statistik zeigt. Bei dem *U.S. Immigration and Naturalization Service (INS)* liegt sie bei über 1.600, bei *Texas Maps* sogar über 7.000.

Bei dem GeoGuide und dem MathGuide werden die Kategorien des Source Type und des Subject Catalogs ähnlich stark genutzt: ca. 55 % aller vorhandenen Kategorien des Source Type Catalog sind im Laufe der letzten 11 Monate mehr als 20 mal aufgerufen worden, beim Subject Catalog liegt die Zahl sogar bei 80 %. Dies bestätigt auch die schon oben getroffene Aussage, daß der Subject Catalog öfter aufgerufen wird als der Source Type Catalog (vgl. Abb. 7.7). Dies mag vielleicht damit zusammenhängen, daß im Internet bei Subject Gateways die Möglichkeit des zwei-dimensionalen Browsens nicht so stark verbreitet ist. Die meisten Fachinformationsführer bieten dem Nutzer in erster Linie eine fachliche Gliederung an und nur, wenn überhaupt, eine stark vereinfachte formale Navigationshilfe. In den meisten Fällen beschränkt sich diese auf Journals, Organizations, Universities und vielleicht noch Software oder Linklisten. Die Zahlen zeigen aber auch, daß der Source-Type Catalog als zusätzliche Browse-Funktion akzeptiert wird.

Bei beiden Guides werden ca. 40 % aller vorhandenen Templates aufgerufen, bei 1000 Datensätzen im GeoGuide und 804 Datensätzen im MathGuide (jeweils Stand Anfang November). Während beim MathGuide sich die Anfragen mehr oder weniger gleichmäßig zu verteilen scheinen, fällt beim GeoGuide die hohe Zugriffsrate auf den Infodaten-Satz 557 mit über 1700 auf. Auch hier liegt, wie beim AAC-History Guide, der Verdacht nahe, daß die Nutzer weniger durch das Browsen im GeoGuide als vielmehr durch Benutzung der Suchmaschinen auf den Datensatz stoßen. Immerhin handelt es sich bei dem *World Road Atlas* von *Mapquest* ebenfalls um einen äußerst populären Site mit über 40.000 Backlinks. Leider können zur Zeit noch keine Aussagen darüber getroffen werden, ob die Nutzer, die indirekt über Suchmaschinen zu den SSG-FI Guides finden, sich diese näher anschauen und schätzen lernen oder ob sie nach erfolgreicher Suche nach der URL, die sie in den Templates finden, die Guides

verlassen, sich diese aber für das nächste Mal als Anlaufstelle vormerken oder nicht.

Beim GeoGuide sind die Seiten sehr gefragt, die sich mit thematischen Karten beschäftigen (vgl. Tab. 7.8). Sowohl im Source-Type Catalog als auch im Subject Catalog und bei den Templates spielen sie eine sehr große Rolle. Es wäre für den GeoGuide daher zu überlegen, inwieweit er als Einstiegspunkt diesen Bereich weiter ausbauen sollte bzw. weiter differenzieren könnte. Als Konsequenz kann weiterhin gelten, daß bei der Recherche und Aufnahme geowissenschaftlich relevanter Seiten das Hauptaugenmerk bei den Karten liegen könnte. Anscheinend können Nutzer über diesen Themenbereich am ehesten an den GeoGuide herangeführt werden. Allerdings wirft dies auch Fragen nach der Zielgruppe des SSG-FI Systems auf. Eine der wichtigsten Prämissen des SSG-FI Projektes ist, daß nur wissenschaftlich relevante Ressourcen, die in erster Linie für die Forschung wichtig sind, erfaßt werden sollen. Deshalb gilt es für die Zukunft zu klären, inwieweit auch eher populärwissenschaftliche Internet-Ressourcen die Akzeptanz der Guides erhöhen würde und ob dies nur zu einer unnötigen Belastung des SSG-FI Servers führen würde. Auch wäre zu überlegen, inwieweit die wissenschaftliche Fachwelt auch an populärwissenschaftlichen Seiten interessiert ist.

Die unterschiedlichen Nutzergruppen mit unterschiedlichen Interessenschwerpunkten zeigen sich besonders deutlich im MathGuide bei den am häufigsten aufgesuchten Kategorien und Info-Datensätzen (Tab. 7.8). Die drei meist frequentierten Kategorien des Source-Type Catalogs (*Series of Preprints...*, *Journals* und *Research Institutes*) deuten auf eine Nutzergruppe hin, die primär an Forschung im eigentlichen Sinne interessiert ist. Im Gegensatz dazu zeigen die aufgerufenen Kategorien des Subject Guide (*History and biography*, *Mathematical logic and foundations*, *Statistics*), daß die Nutzer dieser Kategorien eher an allgemeinen Grundlagen und Lehrmaterial interessiert sind. Somit zeichnen sich zwei Nutzerprofile ab, die auf unterschiedlichen Bedürfnissen beruhen.

Die drei meist gefragten Infodaten-Sätze weisen eventuell noch auf eine dritte Nutzergruppe hin (vgl. Tab. 7.8): Der Infodaten zu *The National Council of Teachers of Mathematics* verweist auf einen eher populärwissenschaftlichen Site, der mit über 2.000 Backlinks die Schulmathematik und nicht forschungsrelevante Themen behandelt, und auch die beiden anderen Infodaten-Sätze (*Princeton Institute for Advanced Study*, *School of Mathematics*, *London School of Economics and Political Science*; *Department of Statistics*) lassen eher auf ein allgemeines als auf ein Forschungsinteresse schließen. Ob der Aufruf der Templates primär von Quereinsteigern über Suchmaschinen oder durch gezieltes Browsen bzw. Suchen im MathGuide erfolgt, kann nicht eindeutig geklärt werden. Da allerdings die Anzahl der Anfragen an die Templates gleichmäßiger

verteilt ist als beim AAC-History Guide und GeoGuide, liegt es nahe, daß sie durch den MathGuide selbst erfolgen.

Es ist derzeit nicht zu klären, ob die vorangeschaltete Liste der Mini-Templates ausreichend Informationen bietet, um sich einen Überblick über die Ressourcen zu verschaffen oder ob sie vorwiegend dem Verweis auf die vollen Datensätze dienen. Die ausführlichen Metadaten-Templates ermöglichen auf jeden Fall, in der lokalen Suche vielfältige Optionen anzubieten, d.h. gezielte Suche nach Titel, Autor, Publisher, Keywords etc. Sie ermöglichen es zudem dem Quereinsteiger, der über kommerzielle Suchmaschinen nach Informationen sucht, über die ausführlichen Templates indirekt auf die Guides zu führen, wie dies z.B. bei dem AAC-History Guide der Fall zu sein scheint.

Sollten in Zukunft Suchmaschinen angeboten werden, die auch Dublin Core Metadaten indexieren und demzufolge suchbar machen, so dürfte das SSG-FI System als einer der wenigen Metadaten-Fachinformationssysteme an welcher Akzeptanz gewinnen.

8 Modelle kooperativer Pflege der SSG-FI-Guides

Von Wilfried Enderle

Die laufende Pflege eines Fachinformationssystems allein durch eine SSG-Bibliothek wird in den meisten Fällen sicherlich nicht vollständig und mit den regulären Personalressourcen einer SSG-Bibliothek möglich sein, da es sich, zumindest bei der derzeitig immer noch hohen Bedeutung der gedruckten Publikationen, eindeutig um eine neue, zusätzliche Aufgabe und Dienstleistung handelt. Die alleinige Pflege sollte aber auch nicht als Ziel angestrebt werden, da gerade der laufende Betrieb eines Internet-Fachinformationssystems am besten in Kooperation mit Nutzern, Autoren sowie anderen Bibliotheken und Fachinformationseinrichtungen stattfinden sollte. Gerade das Internet bietet die Möglichkeit, neue Formen und Modelle wechselseitiger Kooperation zu entwickeln, wie sie bislang nicht möglich waren. Gerade in diesen neuen organisatorischen Möglichkeiten liegt eines der innovativen Momente solcher Subject Gateways, die in ihrer Bedeutung denen der technischen Innovationen nicht nachstehen dürften. Daher sollen kurz die möglichen und konkret auch vorgesehenen Kooperationsmodelle umrissen werden, mit deren Hilfe der weitere Ausbau und laufende Betrieb der SSG-FI-Guides gewährleistet werden kann. Dies sind zusammenfassend:

- Kooperation mit den Nutzern.
- Kooperation mit den Autoren.
- Kooperation mit Bibliothekaren und Fachdokumentaren.
- Kooperation mit anderen Subject Gateways.

Die einfachste Form der Kooperation stellt diejenige mit den *Nutzern* des SSG-Fachinformationssystems dar, die selbst aufgrund ihrer Fachkenntnis Ressourcen kennen, die noch nicht im SSG-FI-Guide, den sie nutzen, verzeichnet sind. Diese Nutzer können über das vorhandene Webformular die URL dieser Ressource als Vorschlag zur Aufnahme in den Guide melden (vgl. Abb. 8.1). Dieser Dialog mit dem Nutzer ist wichtig und derzeit auch bereits für alle Guides implementiert. Er bedeutet freilich auch, daß die Erfassung eines Datensatzes weiterhin Aufgabe des SSG-Referenten bleibt.

Diese Aufgabe bleibt auch bei einer weiteren, anzustrebenden Kooperationsform: der Meldung neuer thematischer Sites durch die *Autoren* selbst. Auf mittlere Sicht sollte angestrebt werden, daß ein SSG-FI-Guide innerhalb seines Faches sich so etabliert, daß die Autoren neuer Sites selbst daran interessiert sind, ihre Sites in den Guide aufgenommen zu sehen und aus eigener Initiative die entsprechenden Formulare ausfüllen, um so die SSG-Bibliothek auf ihre Ressource aufmerksam zu machen.



The image shows a screenshot of a Netscape browser window. The title bar reads "GeoGuide - Subject Gateway in the Earth Sciences [SUB Göttingen, Germany] - Netscape". The address bar shows "http://www.SUB.uni-goettingen.de/ssgi/geo/". The main content area displays the following text:

SSG-FI Notification: Anglo-American History Guide

Please enter your information in the following fields and send it to SSG-FI.
Your contribution is highly appreciated.

Title:

Description:

URL:

Please include your address:

Name: E-Mail:

At the bottom, there are two buttons: "Send the Form to SSG-FI" and "Discard all entries". The status bar at the bottom of the browser window shows "Document Done".

Abb. 8.1: Webformular zur Meldung neuer Internet-Sites

Kooperationsmodelle im engeren Sinne umfassen die konkrete Mitarbeit, also das Beschreiben thematischer Internet-Ressourcen durch fachwissenschaftliche *Bibliothekare und Dokumentare* anderer Einrichtungen. Das ideale, wenn auch in der konkreten bibliothekarischen Praxis der Bundesrepublik sicherlich etwas utopische Modell, bestünde in einer nationalen Fachreferentenkooperation. Immerhin wäre ein nationaler Subject Gateway ein Dienst, der im Grunde allen Bibliotheken zugute käme, da sie sich auf diese Weise den Aufbau und Betrieb eigener fachlicher Linklisten oder Fachinformationssysteme sparen können. Um den jeweiligen Anteil einer anderen Bibliothek an einem Guide zu dokumentieren, sollte jeder Datensatz, der von einem anderen Referenten erstellt würde, ein Emblem der zugehörigen Einrichtung enthalten. Der SSG-

Referent würde sich dann schwerpunktmäßig auf die organisatorischen und redaktionellen Aufgaben einer derartigen Kooperation konzentrieren können.

Eine konkrete Kooperation wird zur Zeit für den GeoGuide zwischen der SUB Göttingen und der Universitätsbibliothek „Georgius Agricola“ in Freiberg vorbereitet. Beide Bibliotheken wollen in Zukunft den GeoGuide gemeinsam pflegen.



Abb. 8.2: Demo-Version einer Isaac Distributed Search Engine

Ein viertes, vielversprechendes Kooperationsmodell sieht die Zusammenarbeit verschiedener selbständiger *Subject Gateways* vor. Dies dürfte mittelfristig der interessanteste Ansatz sein, da zum einen die technischen Voraussetzungen dafür vorhanden sind und dieses Modell zugleich auf real existierenden Gateways, die bereits Internet-Ressourcen in gewissem Umfang nachweisen, aufsetzen kann. Theoretisch sieht dieses Modell vor, daß in mehreren, verteilt existierenden Subject Gateways parallel recherchiert werden kann. Ein Nutzer könnte über die Oberfläche des SSG-FI-Systems – zumindest mit eingeschränkter Suchfunktionalität (also rein verbale Recherche über Stich- oder

Schlagwörter) – gleichzeitig in mehreren thematisch zusammengehörenden Subject Gateways recherchieren. Die Ergebnisliste zeigt ihm über entsprechende Logos an, von welchem der kooperierenden Gateways die jeweilige Ressource erschlossen wurde. Eine praktische technische Lösung hierfür wird derzeit im Rahmen des internationalen Projektes Isaac⁶⁷ erprobt, an dem auch das SSG-FI-Projekt als Partner teilnimmt (vgl. Abb. 8.2).



Abb. 8.3: Demo-Version der Isaac Distributed Search Results unter Wahrung der eigenen Identitäten der beteiligten Einrichtungen

Der interessante und zukunftsweisende Aspekt bei diesem Modell besteht darin, daß technisch wie organisatorisch neue, bis dahin nicht vorhandene Kooperationsformen realisiert werden können. Die im Bibliotheksbereich bekannten

67 Siehe Michael Roszkowski/Christopher Lukas: A Distributed Architecture for Resource Discovery Using Metadata. In: D-Lib Magazine, June 1998 (URL <http://www.dlib.org/dlib/june98/scout/06roszkowski.html>) Sowie: Project Isaac Architecture Overview for Collaborators (URL <http://scout.cs.wisc.edu/scout/research/arch/index.html>)

klassischen Modelle der Zusammenarbeit basieren bis dato entweder auf dem Verbundkonzept, einer gemeinsamen Datenbank also, die alle Partner nutzen und ausbauen, oder dem Nachnutzungskonzept; in diesem Fall wurde eine von einer Bibliothek erarbeitete Lösung unmittelbar übernommen und an einer anderen Bibliothek implementiert. Beide Lösungen, die in konkreten Anwendungsfällen natürlich immer noch ihren Sinn haben, implizieren indes, daß eine Bibliothek lokalspezifische Interessen und Identitäten teilweise aufgeben muß. Genau dieser Nachteil existiert bei einer gleichberechtigten Kooperation mehrerer WWW-basierter Subject Gateways nicht mehr, da jeder Gateway seine eigene lokale und fachliche Identität über seine Einstiegsseiten wahren kann, zugleich aber einen größeren Raum von Daten als den eigenen erschließt. Dem Nutzer wird stets deutlich, welches sein Einstiegssystem ist, zugleich aber auch, daß die Daten auch von anderen Institutionen kommen können. Eine Kooperation in dieser Form erlaubt also eine vollständige Wahrung der eigenen corporate identity bei gleichzeitig voller Nutzung der Daten anderer Partner (vgl. Abb. 8.3).

Voraussetzung für eine solche Kooperation ist zunächst einmal, daß die beteiligten Gateways sich auf einen gemeinsamen Minimalset von Metadatenkategorien verständigen und im besten Fall auch noch mindestens eine gemeinsame Sacherschließungskomponente pflegen. Ansonsten dürfte es im wechselseitigen Interesse sinnvoll sein, daß die kooperierenden Gateways, abgesehen von einer gemeinsamen fachlichen Ausrichtung, eine jeweils vergleichbare Größenordnung haben, damit die Nutzung der Daten anderer auch durch die eigenen Beiträge angemessen ausgeglichen werden kann.

Das SSG-FI-Projekt wird versuchen, alle hier skizzierten Kooperationsformen und -modelle zu realisieren. Zum einen soll – auch über entsprechende Informationsaktivitäten in fachlichen Diskussionslisten und Zeitschriften – dafür geworben werden, daß individuelle Nutzer und Autoren ihre Sites melden, so daß sich ein regelmäßiger Austausch mit den wissenschaftlichen Nutzern im Laufe der Zeit entwickeln kann. Zum zweiten wird auch versucht werden, andere wissenschaftliche Bibliothekare zur Mitarbeit zu gewinnen. Dafür können zunächst vorhandene Strukturen innerhalb des regionalen Bibliotheksverbundes, wie existierende Facharbeitsgruppen, genutzt werden, aber auch die Einrichtung einer elektronischen Diskussionsliste zum Austausch von Fachreferenten und Fachdokumentaren. Zum dritten werden durch die Teilnahme am Project Isaac bereits modellhaft internationale Kooperationsbeziehungen aufgebaut.

9 Ausblick und weitere Entwicklungsmöglichkeiten

Von Wilfried Enderle

Die Erschließung wissenschaftlicher Internetressourcen durch die fachlichen Guides des SSG-FI-Systems stellt ein konsistentes Konzept dar, um zum einen thematische Sites des Internet in die Erschließungsarbeit einer wissenschaftlichen Bibliothek mit Sammelschwerpunktgebieten zu integrieren und um zum anderen den wissenschaftlichen Nutzern mit dieser Erschließung eine neue Dienstleistung als SSG-Bibliothek anbieten zu können: gleichsam in klassischer Terminologie eine „Bibliographie der Bibliographien“ zu elektronischen Ressourcen. Die Konsistenz des im Rahmen des Projektes erarbeiteten Konzeptes bedeutet freilich nicht, daß es nicht noch Ausbau- und Entwicklungsmöglichkeiten gäbe. Diese liegen zum einen in alternativen technischen Lösungen (Kap. 9.1), vor allem aber auch in der Integration der Guides in die bestehenden klassischen Katalogstrukturen der Bibliothek (Kap. 9.2 und 9.3). Dazu kommt, daß die Datenbasis der SSG-FI-Guides durchaus auch als Grundlage für die Entwicklung hybrider fachlicher Suchmaschinen dienen kann, welche automatisiert Kataloge einzelner Internetdokumente erstellen (Kap. 9.4).

9.1 Dynamische http-Schnittstelle

Die realisierte technische Lösung der SSG-FI-Guides basiert auf einer Datenbank, aus der über Exportskripte die Datensätze in feste HTML-Dateien exportiert werden, die in einer definierten Verzeichnisstruktur auf einem Web-Server abgelegt werden und so den Benutzern über die vorgegebenen Suchmöglichkeiten zur Verfügung stehen. Die Vorteile dieser technischen Lösung liegen darin, daß sie einfach und stabil ist. Es gibt zudem keine Zugriffs- und Antwortzeitprobleme, solange der Web-Server über eine ausreichende Kapazität verfügt. Außerdem sind diese Daten für andere fachliche Suchmaschinen ohne Probleme indizierbar. Der einzige Nachteil dieser Lösung liegt darin, daß Aktualisierungen der Datenbank für den Benutzer erst nach dem Export der Datensätze in das WWW sichtbar sind. Diese regelmäßige Aktualisierung erfordert zudem ein gewisses technisches Know-How und zeitlichen Aufwand.

Das Konzept der SSG-FI-Guides kann natürlich auch jederzeit mit einer Datenbank realisiert werden, welche über eine dynamische http-Schnittstelle verfügt. Diese alternative technische Konzeption zur existierenden Lösung kann mit jedem Datenbanksystem implementiert werden, das über eine derartige Schnittstelle verfügt. Da das im SSG-FI-Projekt eingesetzte Allegro mit dem Avanti-Server über eine solche Schnittstelle verfügt, könnten auch die derzeit eingesetzten Datenbanken relativ einfach auf dieses Modell umgestellt werden. In diesem Fall würde der Zugriff über den Subject Catalog oder den Source Type

Catalog über fest eingestellte Suchkombinationen von Fachnotationen oder Formal Keys erfolgen, welche über ein CGI-Script eine entsprechende Suchanfrage in der Datenbank auslösen würde. Die Suchergebnisse, Kurztitel und Einzeltitelanzeige, würden dann dynamisch als HTML-Datei aufbereitet.

Welche technische Lösung für die Realisierung eines SSG-FI-Guides nun gewählt wird, hängt von den lokalen Voraussetzungen ab. Entscheidend ist, daß das fachliche und bibliothekarische Konzept der SSG-FI-Guides weitgehend unabhängig von konkreten technischen Lösungen mit verschiedenen Datenbanksystemen realisiert werden kann. Die einzigen Voraussetzungen sind die Übernahme des Metadatenformats und die Vergabe fachlicher Notationen für die einzelnen Datensätze, um ein fachsystematisches Browsen zu ermöglichen und so inhaltliche Konsistenz zu gewährleisten. Wer die im Rahmen des Göttinger SSG-FI-Projektes erarbeiteten Programme und Skripte nachnutzen will, kann die auf Allegro basierende Lösung direkt übernehmen.

9.2 Ausbau um die Daten gedruckter Informationsquellen

Zur Zeit werden über die SSG-FI-Guides thematische Internetsites erschlossen, also elektronische Ressourcen, die weiterführende Informationen und Materialien enthalten. Die Funktion dieser Guides besteht mithin darin, daß Wissenschaftler für ihre fachspezifischen Fragen auf Internet-Ressourcen geführt werden, die dazu weitere Informationen enthalten, seien es weiterführende Informationen in Form fachlicher Bibliographien oder Linklisten, oder primäre Informationen in Form von Working Papers, elektronischen Zeitschriften oder Sammlungen digitalisierter Quellentexte. Es steht außer Frage, daß, insbesondere in den Kulturwissenschaften, diese elektronischen Informationsmöglichkeiten für den Wissenschaftler nur einen Teil, oft sogar einen eher marginalen Teil derjenigen Hilfsmittel darstellen, die er für seine fachwissenschaftlichen Recherchen nutzt. Gedruckte Fachwörterbücher und Bibliographien stellen hier immer noch ein wichtiges Recherche-Instrument dar. Daher liegt es eigentlich nahe, die SSG-FI-Guides um eine weitere, optional zu wählende Funktion zu erweitern: die Suche nach gedruckten Informationsquellen zu fachspezifischen Themen. Damit würde eine SSG-Bibliothek einen einheitlichen und aus der Sicht des Nutzers integrierten Zugriff auf sekundäre Informationsmittel geben.

Der konzeptionelle Ausbau der SSG-FI-Guides um gedruckte Informationsmittel liegt auch deshalb nahe, weil er mit einem relativ geringen zusätzlichen Aufwand unter Nutzung vorhandener Daten einen im Vergleich dazu beachtlichen Zugewinn an Informationsmöglichkeiten bietet. Anders formuliert: Die gedruckten fachlichen Informationsmittel einer Bibliothek sind in der Regel im lokalen Online-Katalog erschlossen. Im Falle der SUB Göttingen sind sie zum Beispiel ebenfalls mit fachlichen Notationen der Göttinger Online-Klassifikation und Formalschlüsseln versehen. Damit sind sie im Grunde in derselben Form e r-

geschlossen wie die elektronischen Internet-Ressourcen in den SSG-FI-Datensätzen. Da der Online-Katalog der SUB Göttingen über eine http-Schnittstelle verfügt, ist es somit prinzipiell kein Problem, aus den Datensätzen der SSG-FI-Guides, also dem Subject Catalog und dem Source Type Catalog, mit einem CGI-Script eine Suchanfrage an den Katalog zu starten. Diese würde aus einer Kombination von GOK-Notation und einem Formalschlüssel bestehen, der das fachliche Nachschlagewerk kennzeichnet. Voraussetzung ist nur, daß die entsprechende Kategorie indiziert ist. Das SSG-FI-System würde damit gleichsam als Metakatalog für den lokalen Online-Katalog fungieren, aus der Sicht des Benutzers hingegen wie ein konsistentes System aussehen, das ihm gleichermaßen elektronische wie gedruckte Fachinformationsmittel erschließt. Für ihn ist es letztlich unwichtig, daß die Daten, die er über eine Oberfläche recherchiert in technischer Hinsicht aus zwei unterschiedlichen Systemen kommen.

Ein derartiger Ausbau der SSG-FI-Guides würde damit eine partielle Mehrfachnutzung vorhandener Daten darstellen. Es wäre ein Beispiel dafür, wie vorhandene Daten durch spezifische Suchoberflächen für die individuellen Bedürfnisse der Nutzer gleichsam maßgeschneidert in unterschiedlicher Form angeboten werden könnten. Aus dem umfassenden Datenmaterial, das eine SSG-Bibliothek in ihren Katalogen vorhält, würden unterschiedliche, spezialisierte Dienste für den Nutzer zur Verfügung gestellt.

9.3 Einspielen der SSG-FI-Datensätze in den lokalen Online-Katalog

Ebenso kann und sollte man, um das Potential der vorhandenen Daten so weit wie möglich auszunutzen, auch den Weg in die andere Richtung gehen: Die Integration aller Daten in eine Datenbank mit der Möglichkeit einer einheitlichen Recherche in allen vorhandenen Daten der Bibliothek. Derzeit stellen die SSG-FI-Guides ein spezielles Recherche-Instrument für spezifische wissenschaftliche Recherchebedürfnisse dar. Sie sind damit ein Instrument für den qualifizierten Nutzer, das in dieser Form auch als Teil eines umfassenden, differenzierten Informationsangebotes einer SSG-Bibliothek notwendig ist. Zugleich liegt es aber natürlich nahe, diese Daten im Online-Katalog der SSG-Bibliothek zusammenzuführen, da bei sachlichen Recherchen dort, seien sie über simple Stichworte oder über systematische Notationen durchgeführt, dann auch die elektronischen Ressourcen, welche die Bibliothek als Teil ihrer Erschließungsarbeit erfaßt hat, mit angeboten werden. Die Datensätze der SSG-FI-Guides sollten in Zukunft also auch in den Online-Katalog der SUB Göttingen eingespielt werden. Dies ist möglich, da es sich bei den SSG-FI-Daten um strukturierte Datensätze handelt und das Sacherschließungskonzept mit dem für die gedruckte Literatur konform ist. Die Daten fügen sich somit, ohne daß eine Überarbeitung notwendig wäre, konsistent in den Online-Katalog der Bibliothek ein. Der einzige wesentliche Unterschied zu den anderen Daten besteht darin,

daß sie nicht auf ein konkretes, physisches Objekt vor Ort, also ein Buch oder eine Zeitschrift, verweisen, sondern auf einen thematischen Site im Internet.

Dieser Unterschied sollte freilich nicht gänzlich außer acht gelassen werden. Die Daten des SSG-FI-Systems verlangen eine andere Form der Pflege als die Katalogdaten für gedruckte Bücher. Die Gültigkeit von URLs muß regelmäßig geprüft werden, mögliche inhaltliche Veränderungen thematischer Sites in den Datensätzen nachgeführt werden usw. Insofern stellen diese Daten für den Bibliothekar durchaus eine spezifische Kategorie innerhalb seiner Datenbank dar. Diese Besonderheit freilich muß sich nicht in der Form widerspiegeln, wie diese Daten dem Nutzer zur Verfügung gestellt werden, da dieser sich nun einmal in der Regel nur für den wissenschaftlichen Informationsgehalt der nachgewiesenen Ressourcen interessiert, seien sie elektronisch oder gedruckt.

9.4 Das SSG-FI-System als Basis für die automatisierte Katalogisierung einzelner Internetdokumente

Der bibliothekarische Kern des SSG-FI-Konzeptes entspricht durchaus bisherigen Aufgaben und Konzepten einer wissenschaftlichen Bibliothek: Wissenschaftliche Informationsressourcen werden mit Hilfe einer Datenbank erfaßt. Der Vorteil eines solchen Ansatzes besteht auch darin, wie gerade die Entwicklungsperspektiven (vgl. Kap. 9.2 und 9.3) belegen, daß das SSG-FI-Konzept direkt in vorhandene bibliothekarische Katalog- und Erschließungsstrukturen integriert werden kann. Zugleich bietet dieses Konzept aber auch eine Basis für neue, informationstechnisch innovative Ansätze, die z.B. die Erschließung der thematischen Internet-Sites im SSG-FI-System als Ausgangspunkt für den Einsatz von Suchmaschinen und Verfahren automatisierter Katalogisierung nutzen.

Ähnlich wie bei gedruckten Monographien oder Periodika, bei denen mittlerweile Bibliotheken mit Hilfe moderner Scan- und OCR-Technologien versuchen, eine weitergehende Erschließung als bislang durchzuführen, indem sie Inhaltsverzeichnisse scannen und in ihre Online-Kataloge integrieren, ist es auch bei Web-Sites möglich, bis auf Dokumentenebene zu erschließen. Im Unterschied zum gedruckten Medium bietet es sich hier sogar im weitaus stärkeren Maße an, dies zu tun, da die technischen Voraussetzungen dafür ungleich günstiger sind. Da die Dokumente der Web-Sites ohnehin in digitaler Form vorliegen, können sie auch automatisiert indiziert und erschlossen werden. Mit anderen Worten: Auf der Basis der intellektuellen Auswahl der SSG-FI-Guides können fachliche Suchmaschinen aufgesetzt werden, welche über die Ebene des Web-Site hinaus, ggf. bis auf die Ebene der einzelnen Dokumente dieser Web-Sites, eine Indizierung durchführen.

Ein konkretes System, mit dem eine derartige fachliche Suchmaschine aufgebaut werden kann, wäre zum Beispiel die Harvest-Software. Im Elib-Projekt der UB Osnabrück wird sie genutzt, um fachliche elektronische Zeitschriftenbiblio-

theken aufzubauen. Dieses Grundkonzept kann auch auf andere Dokumentformen übertragen werden.

Technisch denkbar wäre es, alle über die SSG-FI-Guides erschlossenen Web-Sites bis zu einer definierten Ebene zu indizieren, die ggf. auch von Site zu Site verschieden sein kann. Damit hätte man einen fachlichen „Basic Index“ aufgebaut, der zu einzelnen spezifischen Fragen, die sich durch spezielle Suchbegriffe formulieren lassen, sinnvolle Resultate erbringen kann. Insgesamt freilich dürfte ein solcher „Basic Index“ für eine qualitative Fachrecherche noch keine ausreichend präzisen Ergebnisse erzielen. Sinnvoller dürfte es sein, gerade solche Web-Sites für eine automatisierte Erschließung auszuwählen, die vorwiegend einzelne digitale Dokumente enthalten und so zum Beispiel elektronische Zeitschriften- und Textbibliotheken zu erzeugen. Im ersten Fall würden alle elektronischen Zeitschriften bis zur Ebene der Inhaltsverzeichnisse und, soweit vorhanden, der Abstracts erschlossen. Eine Suche über die Suchmaschine einer solchen Zeitschriftenbibliothek würde damit über Stichworte der Aufsatztitel einschlägige Texte finden können. Eine Textbibliothek würde diejenigen Sites erschließen, die Texte, seien es von einzelnen Autoren im Bereich der Philologie oder allgemeine Quellentexte, aufliegen haben. Diese Texte hätten im Prinzip analogen Charakter wie Monographien, so daß eine entsprechende Suchmaschine das Pendant zum Online-Katalog gedruckter Monographien darstellen würde. Die Suchmaschine müßte also für die Web-Sites bis auf die Ebene der jeweiligen Inhaltsverzeichnisse, welche die Autoren und Titel enthalten, eine Indizierung vornehmen.

Möglicherweise könnten die so gewonnenen Indices, welche auf einzelne Dokumente im Internet verweisen, sogar, *analog* zu dem Modell des Einspielens der SSG-FI-Datensätze in den lokalen Online-Katalog (siehe Kap. 9.3), ebenfalls in den lokalen Katalog integriert werden, so daß im Falle einer einfachen Stichwortrecherche ein Nutzer als Suchergebnis auch auf diese elektronischen Dokumente im Internet geführt wird. Da es sich auf alle Fälle um einen fachlich klar eingegrenzten Bereich elektronischer Dokumente handelt, könnten hier auch Verfahren automatisierter Sacherschließung sinnvoll eingesetzt werden, sei es durch Rückführung von Stichworten auf ihre Grundformen, wie es in den Projekten MILOS⁶⁸ und KASCADE⁶⁹ entwickelt wurde, sei es durch ein Expertensystem, welches Stichworten Notationen zuteilt, wie im Falle von OSIRIS⁷⁰.

68 MILOS: Maschinelle Indexierung zur erweiterten Literaturschließung in Online-Systemen (URL: http://www.uni-duesseldorf.de/WWW/ulb/mil_home.htm)

69 KASCADE: KAtalogerweiterung durch SCanning und Automatische DokumentErschließung (URL: http://www.uni-duesseldorf.de/WWW/ulb/kas_home.htm)

70 OSIRIS: OSnabrück Intelligent Research Information System (URL: <http://osiris1.ub.uni-osnabrueck.de/isis/demo/start.htm>)