Veröffentlichungsverzeichnis/ Volltextarchiv: Ein digitales Archiv für Veröffentlichungen der Universität Karlsruhe

von Günter Radestock

Das Veröffentlichungsverzeichnis/Volltextarchiv (VVV) ist der Volltextserver der Universitätsbibliothek (UB) Karlsruhe. *Das VVV enthält Dissertationen und andere wissenschaftliche Veröffentlichungen der Universität und dem Forschungszentrum Karlsruhe. Grundlage des VVV ist die Software Pscript, die Postscriptdokumente aufbereitet und dem Benutzer am Bildschirm anzeigt.*

The digital archive of publications of the Karlsruhe University Library. *This article describes the archiv of digital University publications (VVV) i.e. the document server of the Karlsruhe University Library. VVV is based on the software Pscript, which preprocesses postscript documents and displays them to the user through the web.*

Un archive digital des publications de l'Université de Karlsruhe *Le registre des publications/archive des textes complèts (VVV) est le serveur des textes complèts des la BU Karlsruhe. Le VVV contient des dissertations et des autres publications scientifiques provenant de l'université et du centre de recherche à Karlsruhe. Le VVV est basé sur la software Pscript, qui traite des documents postscript et les montre auc client sur l'écran.*

Seit 1972 wird von der Universitätsbibliothek (UB) Karlsruhe in Zusammenarbeit mit dem Forschungszentrum Karlsruhe (FZK, früher Kernforschungszentrum) eine Bibliographie wissenschaftlicher Veröffentlichungen von Universität und FZK herausgegeben: das Veröffentlichungsverzeichnis. Seit 1989 ist das Veröffentlichungsverzeichnis auch als Online-Datenbank verfügbar (http://www. ubka.uni karlsruhe.de/ hylib/ vv_ suchmaske.html). Im Januar 1997 haben wir damit begonnen, einen Teil der im Veröffentlichungsverzeichnis verzeichneten Dokumente als Volltexte auf unserem WWW-Server zur Verfügung zu stellen. Inzwischen enthält das Volltextarchiv (http://www.ubka.uni-karlsruhe.de/vvv) über 700 Dokumente, davon über 20 Dissertationen. Die Software für das System wurde zum großen Teil an der UB entwickelt.

Das Pscript genannte System übernimmt die Aufbereitung und Darstellung der Dokumente, die meist als Postscriptdateien angeliefert werden. Pscript erlaubt auch die Indexierung der Volltexte mit einer beliebigen Suchmaschine, die Einbindung anderer Dateiformate wie PDF und Latex, sowie die Zusammenarbeit mit der Software Dienst, die vom verteilten Suchdienst NCSTRL verwendet wird. Der Zugang zu den Dokumenten erfolgt entweder über die Olix-Datenbank „Veröffentlichungsverzeichnis" (die Datenbank enthält Dokumentnummern, die bei der Präsentation in Links auf den Dokumentenserver umgewandelt werden) oder über eine Volltextsuchmaschine (momentan setzen wir die frei verfügbare Suchmaschine HTDig (http://htdig.sdsu.edu) ein).

1 Voraussetzungen

Fast alle Publikationen von Mitarbeitern der Universität Karlsruhe werden mit Textverarbeitungssystemen erstellt und liegen primär in digitaler Form vor. Damit sind die Voraussetzungen für das elektronische Publizieren und die Verbreitung dieser Texte über Datennetze – insbesondere über das Internet – gegeben. In der Tat bieten bereits zahlreiche Mitarbeiter der Universität über die WWW-Server von Instituten, Fakultäten oder privat ihre Publikationen im Internet an.

Bei näherem Betrachten werden dabei einige Probleme deutlich, welche die Benutzbarkeit der Publikationen zum Teil einschränken:

- Die Texte liegen in unterschiedlichen Datenformaten vor
- die Dokumente sind meist uneinheitlich oder gar nicht erschlossen
- die Dokumente sind meist nicht in Katalogen verzeichnet, sondern nur über Suchmaschinen u.ä. unzuverlässige Hilfsmittel auffindbar
- Die Server, auf denen die Dokumente aufliegen, werden z.T. unzureichend gewartet, mit der Folge, daß hohe Ausfallzeiten und Datenverluste vorkommen
- Die Adressen können sich ändern
- Die Langzeitsicherung der Dokumente ist ungeklärt. Die Flüchtigkeit der elektronischen Dokumente birgt die Gefahr in sich, daß wichtige wissenschaftliche Erkenntnisse nach einigen Jahren nicht verfügbar sind, wenn nicht rechtzeitig Maßnahmen zur Archivierung getroffen werden.

Der konkrete Zugriff auf die elektronischen Dokumente gestaltet sich deshalb zur Zeit noch recht mühsam.

2 Ziel des Projektes VVV

Traditionell versorgt die Universitätsbibliothek die Universität mit wissenschaftlicher Literatur aller Art. Sie sieht es als eine wichtige Aufgabe der Zukunft an, den Universitätsangehörigen auch Zugriff auf elektronische Dokumente zu geben. Zudem hat sie als zentrale Archivbibliothek der Universität die Verpflichtung, die langfristige Archivierung dieser Dokument ebenso zu gewährleisten, wie es bei Printmedien üblich ist.

Da bisher noch wenig Erfahrungen im routinemässigen Umgang mit elektronischen Dokumenten vorliegen, wurde damit begonnen, ein Volltextarchiv aufzubauen, das alle elektronischen Dokumente enthält, die in der Universität erzeugt werden und von der Bibliothek in dieser Form angeboten werden dürfen. Ausgangsbasis hierfür ist das konventionelle Veröffentlichungsverzeichnis, das seit 1972 von der UB in Printform und seit 1989 als Datenbank angeboten wird und alle Publikationen nachweist.

Neben anderen wissenschaftlichen Veröffentlichungen enthält das VVV auch die technischen Berichte der Fakultät für Informatik und Dissertationen.

NCSTRL (National Computer Science Technical Reports Library) ist eine verteilte Datenbank von technischen Berichten aus dem Gebiet Informatik. Das Rückgrat von NCSTRL ist ein auf dem Internet aufgesetztes Netzwerk von Servern, auf denen Dokumente gespeichert sind und die bibliographische Angaben mit Hilfe des Protokolls „Dienst" untereinander austauschen. Die Dokumente werden von Dienst als Postscript und (optional) als GIF-Dateien in Bildschirmauflösung gespeichert. An der UB-Karlsruhe wird ein solcher Server für die Fakultät für Informatik betrieben. Die Konventionen zur Speicherung von Dokumenten und Metadaten wurden von NCSTRL/Dienst übernommen, so daß die technischen Berichte Informatik nur einmal gespeichert sind und sowohl über das VVV als auch über NCSTRL zugänglich sind.

Die Universitätsbibliothek Karlsruhe ist auch am Projekt „Dissertationen Online" (http://www.educat.hu-berlin.de/diss-online/) beteiligt – wir möchten dadurch u.a. erreichen, daß an der Universität Karlsruhe erstellte und im VVV gespeicherte Dissertationen auch über überregionale Verzeichnisse zugänglich werden.

3 Konzeption

Das VVV stellt ein Konzept dar, wie elektronische Dokumente einheitlich präsentiert, umfassend recherchiert und langfristig archiviert werden können.
Im einzelnen bietet das VVV:

- Zugriff auf die Texte direkt nach der Katalogrecherche
- Einfacher und komfortabler Zugriff auf die Dokumente (Bildschirmlesen und Ausdruck)
- Gute Recherchemöglichkeiten im Text der Dokumente
- Rund um die Uhr-Verfügbarkeit
- Sicheres Backup
- Sicherung der langfristigen Verfügbarkeit, Archivierung (gegebenenfalls mit Überführung in neues Datenformat)
- Sicherstellung der Authentizität der Dokumente

3.1 Verfügbarkeit und Archivierung

Alle Dokumente werden auf dem Volltext-Server der UB aufgelegt – wenn ein Dokument bereits auf einem anderen Server aufliegt, wird es kopiert. Dadurch können Ausfallzeiten oder Reorganisationen der Ursprungsserver die Verfügbarkeit über das VVV nicht beeinträchtigen. Bei von anderen Servern kopierten Dokumenten wird die Quell-URL als Metainformation mitgespeichert. Die Quell-URL wird dem Benutzer mitgeteilt und dazu verwendet, eine Änderung des Ursprungsdokumentes automatisch festzustellen (noch nicht realisiert).

Die Archivierung erfolgt mit dem Backupsystem ADSM. Bei ständig sinkenden Preisen für Plattenkapazität rechnen wir nicht damit, daß unsere Dokumente in Zukunft auf andere Medien ausgelagert werden müssen.

3.2 Authentizität

Die Authentizität von Dokumenten, die auf dem VVV-Server aufliegen, gewährleistet die Universitätsbibliothek. Das Einbringen von neuen oder geänderten Dokumenten geschieht nur nach Absprache mit dem Autor.

3.3 Urheberrecht

Urheberrechtliche Probleme treten vor allem bei Texten auf, die in Zeitschriften, Kongressbänden und Büchern erscheinen. Es gibt Verlage, die den Autoren die elektronische Parallelveröffentlichung von Aufsätzen, die in Printform erscheinen, verbieten. Eventuelle urheberrechtliche Fragen im Zusammenhang mit der elektronischen Verbreitung seiner Texte zu regeln, obliegt dem Autor. Jeden Einzelfall mit dem Verlag abzuklären, kann die UB nicht leisten. Sollten nach der Veröffentlichung auf dem Server der Bibliothek rechtliche Probleme auftauchen, so kann der Zugang zu den entsprechenden Dokumenten sehr schnell gesperrt werden.

Eine Abhandlung über Autorenrechte und andere urheberrechtliche Fragen findet sich in ①.

3.4 Organisation

Neue Dokumente für das VVV erreichen uns auf zwei Wegen. Entweder der Autor des Dokumentes überläßt uns eine (elektronische) Kopie seiner Arbeit oder wir „finden" seine Arbeit auf einem seinem Institut zugeordneten Webserver.

Die aktive Anmeldung einer Veröffentlichung bei der Bibliothek ist das von uns präferierte Verfahren, es erfordert jedoch eine gewisse Motivation der Autoren zu diesem bürokratischen Schritt. Wir versuchen durch den Einsatz eines WWW-Formulars die Anmeldung möglichst einfach zu machen und bieten durch die Aufbereitung und Archivierung der Dokumente einen gewissen Mehrwert gegenüber der Veröffentlichung auf institutseigenen Webservern. Außerdem haben wir durch Absprache mit einigen Fakultäten erreicht, daß die Veröffentlichung im VVV einer Veröffentlichung im Verlag gleichgestellt wird und der Doktorand dadurch die Kosten für eine Verlagsveröffentlichung sparen kann, wenn er sich für die Veröffentlichung seiner Arbeit im VVV entscheidet.

Das Absuchen von institutseigenen WWW-Servern führen wir durch, um die für eine sinnvolle Recherche im Volltextarchiv notwendige kritische Masse an Dokumenten zu bekommen. Langfristig möchten wir das Verfahren ablösen oder durch das Anmelden von Publikationslisten durch die Autoren ersetzen, da momentan viel Aufwand durch unterschiedliche Versionen desselben Dokumentes, unterschiedliche Rechtsauffassungen bei Dokumenten von mehreren Autoren usw. entsteht.

Die neuen Dokumente werden in der Regel als Postscriptdateien beschafft oder angeliefert und dann für das VVV aufbereitet. Die Aufbereitung ist weitgehend automatisch und umfaßt die Konvertierung in Textformat, die Extraktion von Seitennummern, Markierung des Inhaltsverzeichnisses (falls vorhanden), die Erzeugung einer Bitmapversion, einer PDF-Version, sowie die Eingabe von Metadaten zum Dokument. Ebenfalls möglich ist die Speicherung von anderen Dateiformaten, entweder optional oder ausschließlich. Neben der Dokumentaufbereitung wird jedes Dokument wie bisher für das Veröffentlichungsverzeichnis katalogisiert. Von Zeit zu Zeit muss der Volltextindex aller Dokumente neu erstellt werden. Der Betrieb des Servers erfordert ansonsten nur den für den Betrieb eines Rechners mit Webserver nötigen Aufwand.

4 Pscript

Zur Aufbereitung und Präsentation der Dokumente haben wir die Software Pscript entwickelt. Pscript wurde ursprünglich als Werkzeug zum Konvertieren von Postscriptdokumenten in ASCII-Text entwickelt, dann als System zur Präsentation von Postscriptdokumenten im WWW erweitert. Pscript besteht aus Komponenten zur Aufbereitung und zur Darstellung der Dokumente:

- Zur Aufbereitung eines Dokumentes wird zunächst aus der Postscriptdatei der Text des Dokumentes sowie die Dokumentstruktur (Inhaltsverzeichnis, Seitennummern, Absätze) extrahiert.
- Bei Dokumenten, die außer Text auch Formeln, Bilder und Grafiken enthalten, sollte zum Lesen am Bildschirm eine Grafikversion erzeugt werden. Diese besteht aus GIF-Dateien in Bildschirmauflösung
- Zu dem konvertierten Dokument müssen Metadaten eingegeben werden, die zur Darstellung benötigt werden. Die benötigten Metadaten sind der Dokumenttitel, eine Liste der Autoren sowie Informationen über die Formatierung: sollen Links auf deutsch oder in einer anderen Sprache angezeigt werden, liegt das Dokument in anderen Dateiformaten vor usw.
- Der Benutzer greift auf die in den vorangegangenen Schritten erzeugten Daten über ein CGI-Programm zu, das jeweils einen Ausschnitt des Dokumentes anzeigt. Dieses Programm erzeugt zu jedem Dokument eine Übersichtsseite mit bibliographischen Angaben, Links auf die Seiten des Volltextes, einer Kurzfassung oder dem Inhaltsverzeichnis, sowie einem Formular zur Stichwortsuche innerhalb des Dokumentes und einer Liste aller Formate, in denen das Dokument als ganzes heruntergeladen werden kann.

Für das VVV erzeugen wir mit Adobe Acrobat (http://www.adobe.com) eine zusätzliche PDF-Version der Dokumente. Pscript erkennt PDF automatisch und führt es in der Liste der Dokumentformate auf der Übersichtsseite mit auf. Andere Dateiformate können ebenfalls eingebunden werden, müssen aber in den Metadaten des Dokumentes spezifiziert werden.

Falls ein Dokument nicht als Postscriptdatei vorliegt und keine brauchbare Postscriptversion erzeugt werden kann, besteht auch die Möglichkeit, auf die aus Postscript generierten Text- und Grafikversionen zum Lesen am Bildschirm zu verzichten und in der Übersichtsseite nur die Liste der Formate zum Herunterladen anzuzeigen.

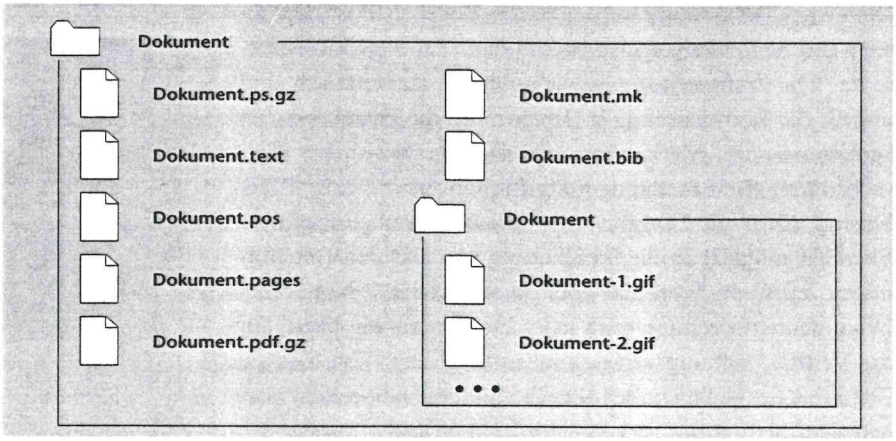
5.1 Aufbereitung von Postscriptdateien

Das Dateiformat Postscript kann von jedem gebräuchlichen Textsystem ohne Mühe erzeugt werden und ist in der Lage, beliebige Dokumente exakt in der vom Autor gewünschten Form (auf Papier) zu reproduzieren. Leider hat Postscript auch Eigenschaften, durch die es schwierig wird, Postscript in andere Formate wie Text zur Indexierung oder HTML zur Ansicht am Bildschirm zu übersetzen.

Postscriptdateien sind Programme, geschrieben in der Programmiersprache Postscript, die beim Ausdrucken im Drucker ausgeführt werden. Postscriptdateien enthalten Anweisungen, u.a. zur Ausgabe von Zeichen an bestimmten Positionen auf einem Blatt Papier. Informationen über die Dokumentstruktur: Wortgrenzen, Absätze, Markierungen von Überschriften, Indexeinträgen oder Inhaltsverzeichniseinträgen sind in der Postscriptdatei nicht mehr vorhanden.

Die Konvertierung von Postscript zu Text basiert auf dem Konverter Prescript (▶ ②), der für die New-Zealand Digital Library eingesetzt wird. Um Text und Struktur wiederzugewinnen wird das Postscriptprogramm zunächst von Ghostscript ausgeführt und dabei die Ausgabe von Text mitprotokolliert. Dieses Protokoll wird ausgewertet – zuerst um durchschnittlichen Zeilenabstand, Zeilenhöhe und Buchstabenbreite zu errechnen, dann um die Zeichen in Worte und Absätze aufzuteilen.

Da die Kodierung der Zeichensätze in Postscript nicht vorgegeben ist, müssen unterschiedlich kodierte Zeichensätze bei der Protokollierung und bei der Auswertung des Protokolls berücksichtigt werden. Einen Sonderfall bieten



▲ *Abbildung 1: Ein Dokument im Dateisystem des Webservers*

Postscriptdateien, die vom Textsystem LaTex erzeugt wurden – hier werden Umlaute aus mehreren Zeichen (eines für die Punkte, eines für den Vokal) erzeugt, die von Pscript wieder zu einem Umlaut zusammengefügt werden müssen.

Um in der Übersichtsseite eines langen Dokumentes ein Inhaltsverzeichnis anzeigen zu können, muß zum einen das Inhaltsverzeichnis im Dokument gefunden werden, zum anderen müssen die im Inhaltsverzeichnis angegebenen Seitennummern den Seiten des Dokumentes zugeordnet werden. Dazu werden die im folgenden kurz angerissenen heuristischen Methoden verwendet. Eine ausführlichere Beschreibung dieser Methoden finden Sie in ③.

Pscript markiert die Zeilen des Inhaltsverzeichnisses, die mit einer Seitennummer enden. Bei der Anzeige werden die Seitennummern im Text als Links formatiert, alle markierte Zeilen werden auf der Titelseite des Dokumentes ausgegeben. Folgende Merkmale werden außerdem beim Erkennen von Inhaltsverzeichniszeilen ausgewertet:

- die Zeile besteht aus Text, einer Reihe von Punkten und endet mit der Seitennummer
- aufeinanderfolgende Zeilen enden rechtsbündig mit einer Seitennummer, die Seitennummern sind dabei aufsteigend sortiert
- mehrzeilige Inhaltsverzeichniseinträge werden zu einer Zeile zusammengefaßt, wenn alle bis auf die erste Zeile nach rechts eingerückt sind

In vielen Fällen werden auf diese Weise auch Zeilen erkannt, die nicht zum Inhaltsverzeichnis gehören, z.B. Tabellen die zufällig mit rechtsbündigen, aufsteigenden Zahlen enden oder ein Teil des Indexes. In diesem Fall löschen wir die falsch erkannten Inhaltsverzeichnis-Markierungen manuell.

Die Seitennummern versucht Pscript zuerst aus dem Text der Seite zu gewinnen. Gelingt das nicht, so wird in der Postscriptdatei nach in Kommentaren eingebetteten Seitennummern gesucht (die leider in vielen Fällen nicht korrekt sind), oder die Seiten werden aufsteigend durchnumeriert.

Das Ablesen der Seitennummern erfolgt in zwei Durchläufen. Im ersten Schritt wird die Anzahl aufeinanderfolgender Nummern auf aufeinanderfolgenden Seiten an verschiedenen Positionen auf der Seite gezählt, z.B. die Anzahl aufeinanderfolgender Nummern im letzten Wort der vorletzten Zeile jeder Seite. Von der so ermittelten Position werden die Seitennummern im zweiten Schritt abgelesen. Beide Schritte werden für gerade und ungerade Seiten getrennt ausgeführt. Schließlich werden fehlende Seitennummern ergänzt oder falsche korrigiert (falsche Seitennummern entstehen z.B. durch Kapitelanfangsseiten, bei denen die Kapitelnummer an der Stelle steht, wo das Programm die Seitennummer vermutet). Die Heuristik zum Ablesen der Seitennummern arbeitet sehr zuverlässig, mir ist kein Dokument mit gültigen Seitennummern bekannt, bei dem sie versagt.

```
BIB-VERSION:: CS-TR-v2.1
ID:: ubka-1998-2
ENTRY:: January 20, 1998
AUTHOR:: Sivawan Phoolphundh
TITLE:: The Degradation of 2-Chlorophenol in an Upflow
Anaerobic Sludge Blanket (USAB) Reactor
DATE:: 1997
FormatType:: originale
FormatName:: Einzeldateien, Microsoft Word Format
FormatUrl:: parts
END:: ubka-1998-2
```

▲ *Abbildung 2: Metadaten zu einem Dokument*

5.2 Speicherung von Dokument und Metadaten

Die Dokumente des VVV werden im Dateisystem des Webservers gespeichert, jedes Dokument in einem eigenen Verzeichnis (▶ *Abb.1*).

Die Dateien Dokument.text, Dokument.pos und Dokument.pages enthalten die aus Postscript konvertierte Textversion mit Seitennummern (Dokument.pages) und Inhaltsverzeichnismarkierungen (Dokument.pos). Die Datei Dokument.mk enthält Parameter zur Darstellung des Dokumentes, die Datei Dokument.bib Metadaten im Format RFC-1807. Die Eintragung im OPAC wird separat von den hier gespeicherten Metadaten vorgenommen, es ist aber prinzipiell möglich, die Katalogisierung anhand der Metadaten beim Dokument zu automatisieren. Zur Speicherung der Dokumente wird keine Datenbank benötigt – alle Informationen liegen im Dateisystem, was die Datensicherung stark vereinfacht. Die Metadaten zu den Dokumenten (▶ *Abb. 2*) werden in Textdateien nach RFC-1807 gespeichert. Dieses Format kann einfach von Hand editiert oder von einem zu Pscript gehörenden Hilfsprogramm erzeugt/bearbeitet werden.

Neben den bibliographischen Angaben (wir verwenden momentan Autor, Titel und Erscheinungsjahr) können hier Angaben zu den Formaten gemacht werden, in denen das Dokument vorliegt. Das Dokument in Abbildung 3 besitzt neben den Postscript und PDF-Versionen, die automatisch gefunden und nicht definiert werden müssen, ein Unterverzeichnis mit dem Namen „originale", das Worddateien der Kapitel der Arbeit enthält. In der Übersichtsseite erscheint ein Link auf dieses Verzeichnis mit dem Namen „Einzeldateien, Microsoft Word Format".

5.3 Volltextsuche

Im VVV wird Stichwortsuche im Volltext der Dokumente unterstützt. Der Benutzer kann in einem allgemeinen Formular nach Dokumenten suchen, die angegebene Stichworte enthalten. Hat er ein Dokument gefunden, so kann er die Suche von der Titelseite des Dokumentes aus innerhalb dieses Dokumentes fortsetzten. Er erhält eine Liste aller Treffer innerhalb des Dokumentes, kann diese direkt anspringen und von jedem Treffer zum vorherigen oder nächsten navigieren.

Die Suche innerhalb eines Dokumentes findet (ohne Index) direkt in der zugehörigen Textdatei statt. Das Anzeigen und Hervorheben von Trefferstellen erledigt das CGI-Programm Makehtml (oben als Teil von Pscript beschrieben). Um die Volltextsuche in allen Dokumenten anbieten zu können, haben wir das VVV mit der frei verfügbaren Suchmaschine HT Dig indexiert. Bei der Indexierung unserer Dokumente stellten sich folgende Probleme:

- die Dokumente sind zu groß, um auf einer Webseite dargestellt werden. Die Indexierung von mehreren Webseiten führt aber zu unschönen Trefferlisten, in denen meistens unterschiedliche Teile desselben Dokuments, nicht unbedingt direkt hintereinander, aufgeführt sind
- Suchmaschinen indexieren normalerweise nicht die Ergebnisse von CGI-Programmen, unsere Dokumente werden von CGI-Porgammen angezeigt. Zwingt man die Suchmaschine zur Indexierung von CGI-Ausgaben, so kann das zu Endlosschleifen führen

Wir haben diese Probleme mit einem Trick gelöst: die Suchmaschine indexiert nicht die URLs der Dokumente, sondern einen separaten Indexierungszugang. Dieser Indexierungszugang ist ein (getarntes) CGI-Programm mit eigener URL, das bei jedem Aufruf überprüft, ob der Aufruf von einer Suchmaschine oder vom Browser eines Benutzers stammt.

Browseraufrufe werden an den „normalen" Zugang weitergeleitet (▶*Abb.3b*). Aufrufe durch eine Suchmaschine (▶*Abb.3a*) erzeugen eine spezielle Version des Dokumentes, bestehend aus Titel, Metaangaben im Dublin-Core-Format, sowie dem Volltext des Dokumentes an einem Stück. Durch diesen Trick kann die Suchmaschine leicht gegen ein leistungsfähigeres Produkt ausgetauscht werden. Der Aufwand zur Realisierung war außerdem geringer als bei der Verwendung von Programmierschnittstellen.

6 Zusammenfassung und Ausblick

Mit dem VVV bietet die UB-Karlsruhe Wissenschaftlern der Universität die Möglichkeit, ihre Publikationen im Internet zu veröffentlichen. Ein Teil der im Veröffentlichungsverzeichnis verzeichnis nachgewiesenen Arbeiten ist inzwischen online verfügbar.

Das VVV basiert auf der an der UB-Karlsruhe entwickelten Software Pscript, die Postscriptdokumente automatisch für das WWW aufbereitet. Die mit Pscript aufbereiteten Dokumente können ohne Zusatzprogramme am Bildschirm betrachtet werden, Pscript ermöglicht die Suche im Volltext der aufbereiteten Dokumente. Der NCSTRL-Server mit internen Berichten der Fakultät für Informatik ist in das System integriert.

In Zukunft möchten wir versuchen, die Volltextsuche mit der Suche im Katalog zu integrieren (ein Formular, ein Suchergebnis). Die Anmeldung neuer Dokumente, die momentan über FTP-Lieferung und Email-Formular erfolgt, kann durch ein verbessertes Formular und durch Angeben von URLs weiter verbessert werden. In diesem Zusammenhang fehlt auch noch eine automatische Überprüfung möglicher Änderungen bei kopierten Dokumenten.

Falls Sie den Aufbau eines ähnlichen Servers vorhaben, stellen wir die für das VVV entwickelte Software auf Anfrage gerne zur Verfügung.

Veröffentlichungen

① Levinski, S. von: Paten der Autorenrechte im Cyberspace; In: Wissenschaftsmanagement, Heft 2, 1998

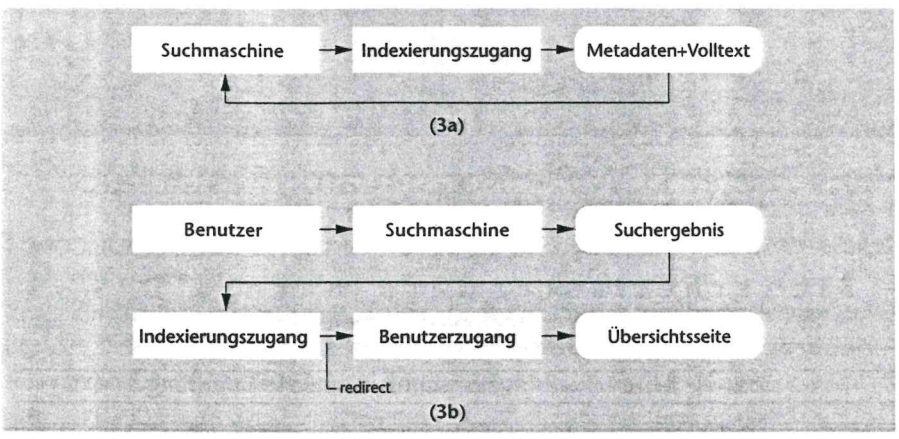
② Todd Reed, Craig Neville-Manning: A Postscript to Plain Text Converter. URL:http://www.cs.waikato.nz/~nzdl/technology/prescript.ps.gz

③ Radestock, Günter: Pscript: Aufbereitung von Postscriptdokumenten für das World-Wide-Web. URL: http://www.ubka.uni-karlsruhe.de/~guenter/pscript/pscript.html

Zum Autor

Dipl. Informatiker **Günter Radestock** ist Wiss. Mitarbeiter in der EDV-Abteilung der Universitätsbibliothek Karlsruhe und bearbeitet im Rahmen des Projektes „Wissensbank Informatik" versch. elektr. Volltextprojekte

▶ Universitätsbibliothek Karlsruhe
Postfach 6920
D-76049 Karlsruhe
E-Mail:
radestock@ubka.uni-karlsruhe.de



▲ *Abbildung 3a und 3b: Indexierungsschnittstelle für Suchmaschinen*