

Verlagsbibliothek ProPrint

1993 meldeten die RZ-Mitteilungen: „Im Zuge der Realisierung des SERVUZ-Projektes in der Humboldt-Universität wurde im Rechenzentrum jetzt die erste Stufe eines Peripherie-Service in Betrieb genommen.“ Hinter dieser Meldung verbarg sich, dass es unter anderem nun im Rechenzentrum möglich ist, einen Print-Service auf Basis eines Tintenstrahl Druckers (Phaserdruckers) und eines Farbsublimationsdruckers mit einer maximalen Auflösung von 300 dpi zu nutzen [1].

Dieser Blick zurück macht es augenscheinlich, in welcher Geschwindigkeit sich die Welt im Bereich des Druck-Services verändert hat. Lagen die Probleme in der Vergangenheit in der Beschaffung und Installation von Druckern, sind es heute darüber hinausgehende Aufgaben, die sich aus der Nutzung des Internets mit seinen Sammlungen von elektronischen Publikationen auf Dokumentenservern ergeben. Um sich diesen Aufgaben zu stellen, wird seit dem November 2000 im Rahmen des vom DFN Verein geförderten Programms „Einsatz von Netzdiensten im Wissenschaftlichen Informationswesen“ an einer Beispiellösung am Rechenzentrum gearbeitet, die im Folgenden dargestellt wird.

Das Projekt „Verlagsbibliothek ProPrint: Benutzerorientierte dynamische Bereitstellung von digitalen und digitalisierten Dokumenten in elektronischer und gedruckter Form (Printing on Demand)“ so der ausführliche Titel, kurz „ProPrint“, stellt sich folgende Ziele:

- Zusammenschluss der Dokumentenserver der Universitätsbibliothek der Humboldt-Universität (HU) und der Staats- und Universitätsbibliothek (SUB) Göttingen
- Installation eines Workflows, der ein Serviceangebot von der Autorenbetreuung bis zur Aufbereitung, Archivierung und Distribution von digitalen Dokumenten und Image Digitalisaten innerhalb der Digitalen Bibliothek an beiden Standorten (Berlin/Göttingen) umfasst
- Bereitstellung eines Dienstes innerhalb der Bibliotheken Berlin – Göttingen zur Erstellung von Publikationen unter Nutzung von Publishing on Demand

Mit dieser technisch-organisatorischen Beispiellösung soll ein neuer Dienst für das wissenschaftliche Informationswesen, genannt *Verlagsbibliothek*, geschaffen werden. Der Dienst Verlagsbibliothek erfüllt eine Doppelfunktion. Zum einen ist er Verlag, indem er seinen Autoren im elektronischen Publizieren und bei der Distribution ihrer Dokumente Hilfestellung bietet. Zum anderen ist er Bibliothek, indem er die elektronischen Dokumente aufbereitet, erschließt und archiviert.

An dem Projekt sind unter der Leitung des Rechenzentrums der HU die Universitätsbibliothek der HU und die Staats- und Universitätsbibliothek Göttingen beteiligt.

Was ist eine Verlagsbibliothek

Die *Verlagsbibliothek* ist ein Teil der Digitalen Bibliothek, deren Dokumentenbestand aus einem Zusammenschluss von einzelnen Dokumentenservern besteht. Die Verbindung der Dokumentenserver erfolgt nicht durch die Spiegelung oder Übernahme von elektronischen Dokumenten, sondern über eine definierte Schnittstelle beider Server. Die Kommunikation beider Server wird dabei über einen RDF-Metadatenatz (RDF: Resource Description Framework), der die bibliographischen Daten, Strukturdaten, Daten der Distribution und der Prüfung nach Authentizität beinhaltet, realisiert. Nur in dieser virtuellen Zusammenfassung aller digitalen Dokumente beider Dokumentenserver bietet die Verlagsbibliothek dem Wissenschaftler einen großen Bestand von Volltexten.

Der Dokumentenserver der SUB Göttingen hält vor allem historisch wertvolle Publikationen in Form von Bilddaten vor. Diese retrodigitalisierten Dokumente sind durch bibliographische Daten und durch Metadaten, die die Struktur des Dokumentes beschreiben, erfasst.

Der Dokumentenserver der Humboldt-Universität hält vor allem aktuelle Publikationen aus Forschung und Wissenschaft vor. Darunter befinden sich Dissertationen aus allen Fakultäten der Universität, die „Öffentlichen Vorlesungen“ – dies sind Antritts- und Festvorträge von Professoren der Universität – und wissenschaftliche Zeitschriften. Die elektronischen Dokumente sind dabei als SGML/XML Instanzen, die mit der DiML-Dokumenttypdefinition (DiML: Dissertation Mark Language) strukturiert sind [2], abgespeichert.

Wie werden die digitalen Dokumente der Verlagsbibliothek vorgehalten?

Die Verlagsbibliothek versteht sich nicht nur als Literaturvermittlungsstelle, die ihren Lesern Informationen vorhält, sondern als aktiver Partner des wissenschaftlichen Autors. Die Verlagsbibliothek steht dabei im Mittelpunkt eines Informations- und Publikationskreislaufes, der das Schreiben des zu publizierenden Textes, das Verlegen des Textes in einer Publikation sowie das Erfassen, Archivieren und Bereitstellen beinhaltet.

Dem Problem des Schreibens, das an erster Stelle des Publikationskreislaufs steht, widmet sich das Projekt ebenfalls. Hier kann es auf bewährte Arbeitsabläufe und Methoden aufbauen, die Ergebnisse des HSP III-Projektes „Digitale Dissertationen“ [3, 4] waren. Zu diesen gehören Formatvorlagen für die Erstellung einer Publikation im Textverarbeitungssystem Microsoft Word und angepasste Konvertierungswerkzeuge.

Nur durch ein gutes Serviceangebot kann den Autoren geholfen werden, die richtigen Werkzeuge für ihre Publikationen auszuwählen und zu bedienen. Der Erfolg dieser Leistungen schlägt sich in der Senkung des Aufwandes bei der Aufbereitung der Dokumente für die Veröffentlichung und bei der späteren Archivierung nieder. Der Autor erschließt nahezu automatisch seinen Text durch die konsequente Handhabung von Formatvorlagen mit Metadaten, die den Text strukturell und inhaltlich beschreiben.

Die spätere Konvertierung nach SGML/XML kann so reibungs- und verlustfrei durchgeführt werden.

Aufbereitung, Archivierung und Distribution

Die Aufbereitung, Archivierung und Distribution der elektronischen Texte in der Verlagsbibliothek soll dem Wissenschaftler aus dem Dokumentenbestand zielgenau die gewünschten Inhalte erschließen und übermitteln. Er kann vollständige Quellen oder einzelne Teile abrufen, individuell zusammenstellen und sich auf Wunsch ausdrucken lassen bzw. an seinem Arbeitsplatz selbst drucken. Bei der Belieferung der Leser werden Techniken verwendet, die unter dem Begriff Publishing on Demand zusammenzufassen sind. Dies bedeutet, dass die Verlagsbibliothek Kleinst- und Einzelausgaben auf verschiedenen Medien druckt oder über das Internet verbreitet.

Im ersten Jahr der Projektarbeit sind vor allem Arbeiten an der Infrastruktur erfolgt. Diese beinhalteten die Erarbeitung eines Schemas zur Arbeitsweise der Verlagsbibliothek ProPrint und Planung, Erstellung und Umsetzung von weiteren Publikationstypen. Die Abbildung 1 soll die Funktionsweise des ProPrint-Servers verdeutlichen. Zuerst werden die Metadaten, die im ProPrint-Server abgespeichert und dem Nutzer über eine Suchmaschine zur Recherche angeboten werden, von den einzelnen Dokumentenservern automatisch abgefragt. Die abgespeicherten Metadaten werden zum einen für die Erzeugung dynamischer Listen, die einen systematischen Zugang damit schaffen, verwendet und zum anderen für die

gezielte Suche über eine Suchmaske. Das Resultat der Suche werden Dokumente oder Teile von Dokumenten sein, die von den Servern abgerufen werden. Der ProPrint-Server setzt in einem Transformationsvorgang, gesteuert durch XSL Stylesheets, die Dokumente zusammen und präsentiert dem Benutzer das Ergebnis dann als ein Dokument am Bildschirm.

Die oben beschriebene Technologie hat die Projektgruppe erstmalig bei der Produktion des Konferenzbandes für die Tagung EUNIS 2001 (siehe <http://www.hu-berlin.de/EUNIS2001>) angewandt. Hierbei wurde den Autoren aus über 27 Ländern der Welt eine auf Konferenzbeiträgen abgestimmte Formatvorlage zur Verfügung gestellt. Ohne zusätzliche Ausbildung waren die Autoren in der Mehrzahl der Fälle in der Lage, die Formatvorlage korrekt anzuwenden, so dass die Beiträge mühelos in SGML/XML konvertiert werden konnten. Für den Druck des Konferenzbandes wurde das Druckhaus Köthen ausgewählt, da dieser Betrieb in der Lage ist, SGML-Daten für den Offsetdruck aufzubereiten.

Parallel zum Druck wurden die Beiträge auf den WWW-Seiten der EUNIS-Konferenz als HTML-Dateien veröffentlicht. Dieser Vorgang des Umsetzens von SGML/XML nach HTML wurde vollautomatisch vorgenommen. Somit konnte zur Konferenz den Teilnehmern der Tagungsband und die WEB-Publikation angeboten werden (<http://dochostrz.hu-berlin.de/overview.php3?doctype=Text.Conferences&lang=ger>).

Der Konferenzband der EUNIS 2001-Tagung umfasst 120 Beiträge. Bei dieser Fülle ist es oft der Wunsch der Leser, die für sie wichtigen Beiträge auszuwählen und als eigenen individuellen Konferenzband zusammenzustellen. Da die Daten in SGML/XML vorliegen, ist dies kein Problem. Durch Nutzung des XML-Publishing-Framework Cocoon wurde exemplarisch

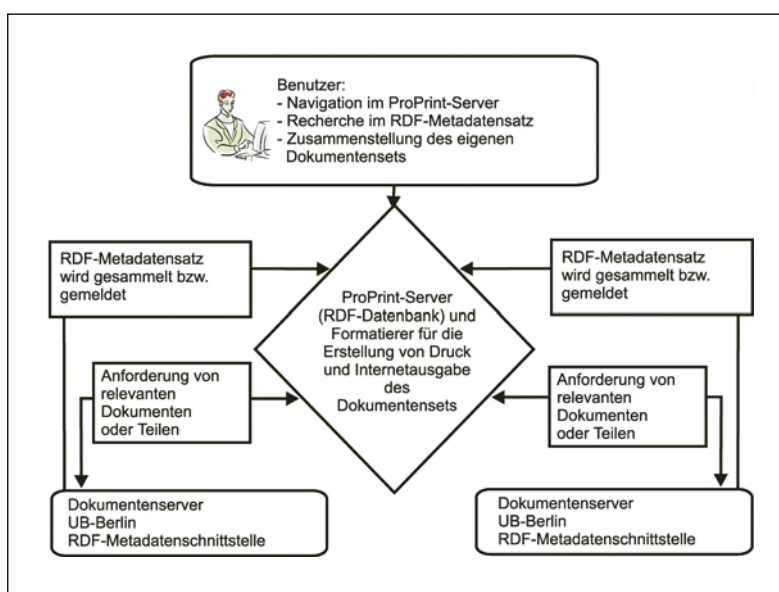


Abb. 1: Schematische Darstellung der Arbeitsweise des ProPrint-Servers

ein Publishing on Demand Dienst eingerichtet. Cocoon ist ein serverbasiertes Publishingsystem zur dynamischen Umsetzung von XML-kodierten Daten in verschiedene Formate. Es arbeitet dabei mit einem Webserver zusammen. Das Framework wird innerhalb des Apache Projektes entwickelt und steht unter einer Open-Source-Lizenz der Apache-Software-Lizenz. Mit Cocoon können XML-Dokumente medienneutral in Formate für die digitale Druckvorstufe (PDF) und für die WWW-Präsentation (HTML) transformiert werden. Neben diesen Formatierungsmöglichkeiten bietet der Konferenzband On-Demand eine Auswahlfunktion der einzelnen Beiträge und der anschließenden Lieferung als PDF-Datei auf den Bildschirm des Benutzers, oder er kann in der automatisch erstellten HTML-Version am Bildschirm im Inhaltsverzeichnis navigieren und sich dadurch einen Überblick verschaffen.

Ausblick

Im kommenden Jahr werden die erarbeiteten Konzepte durch den Aufbau entsprechender Datenbanken und Anschaffung von Publikationswerkzeugen in breitem Stil in die Praxis umgesetzt. Die Betreuung der verschiedenen Publikationsreihen und Typen wird weiter

ausgebaut, so dass am Ende des Jahres die Verlagsbibliothek jedem Lesenden, Suchenden und Schreibenden im Internet zur Verfügung steht. Eventuell werden im Jahr 2002 dann die RZ-Mitteilungen melden:

„Im Zuge der Realisierung des ProPrint-Projektes an der Humboldt-Universität wurde im Rechenzentrum jetzt die zweite Stufe eines Peripherie-Services in Betrieb genommen.“

Literatur

- [1] ÖLZOG, F.: Printservice des Rechenzentrums. *RZ-Mitteilungen* Nr. 6/1993, S. 13.
- [2] DOBRATZ, S.: XML – eXtensible Markup Language. *RZ-Mitteilungen* Nr. 18/1999, S. 24–29.
- [3] MARTIN, N., SCHIRMBACHER, P.: Die elektronische Publikation von Dissertationen an der Humboldt-Universität zu Berlin. *RZ-Mitteilungen* Nr. 14/1997 S. 41–49.
- [4] DOBRATZ, S.: Die „Humboldt-Dissertationen“ im Internet. *RZ-Mitteilungen* Nr. 16/1998, S. 13–16.

Matthias Schulz
matthias.schulz.1@rz.hu-berlin.de

Notebook-Programm an der Humboldt-Universität

Das Rechenzentrum der Humboldt-Universität hat im Zusammenwirken mit Computerfirmen ein Notebook-Programm aufgelegt, mit dem insbesondere Studierenden Sonderkonditionen zum Erwerb und zur Nutzung angeboten werden. Im Zusammenhang mit dem Ausbau des Wireless-LAN in den unterschiedlichen Gebäuden sowie auch auf Freiflächen der Universität werden den Teilnehmern damit besondere Bedingungen zur Nutzung von Netzdiensten auch außerhalb stationärer Computerbereiche geboten.

In diesem Programm wird unterschiedlichen Anbietern die Möglichkeit eingeräumt, auf Web-Seiten der Humboldt-Universität auf ihre speziellen Angebote und Konditionen zum Erwerb von Notebooks hinzuweisen. Parallel dazu werden vom Rechenzentrum PCMCIA-WLAN-Karten gegen eine Schutzgebühr von 150 € ausgeben (bei Rückgabe natürlich Erstattung der Schutzgebühr), über die der Netzzugang ermöglicht wird. Genauere Informationen sind unter <http://www.hu-berlin.de/rz/wlan/> nachlesbar.