

Multimedia Volltextserver mit benutzerprofilgestütztem Push-Mechanismus

Projektstand Januar 1999

Vorbemerkung zur Zielsetzung

Der schnelle Zugang zu aktuellen wissenschaftlichen Informationen wird für die effiziente Arbeit an den Hochschulen immer wichtiger. Um dies zu realisieren, kommt nur eine elektronische Form der Informationsspeicherung und -distribution in Frage.

Das Land Baden-Württemberg stellt über die AGs "Volltexte und Hochschulpublikationen" sowie „Multimedia“ der "Zukunftsoffensive Junge Generation" Mittel zum Aufbau einer virtuellen Forschungsbibliothek an den Universitäts- und Landesbibliotheken zur Verfügung. Damit soll ganz allgemein ein leichter und besserer Zugriff auf die Forschungserträge im Land erreicht werden.

Das hier beantragte und geförderte Projekt, das an der Universität Ulm gemeinsam von der Abteilung LOMI und der Universitätsbibliothek durchgeführt wird, sieht den Aufbau eines Online-Bestandes multimedialer Dokumente (z.B. Dissertationen, allgemeine Hochschulpublikationen, „graue Literatur“) vor, die auf der Basis eines vom Autor selbst erstellten Präkatalogisats von der Universitätsbibliothek in den normalen Katalogbestand aufgenommen und über den Horizon WebPAC recherchierbar gemacht werden. Darüber hinaus wird ein Mechanismus entwickelt, der es den Nutzern des Systems ermöglicht, ein individuelles Profil für ihre Informationspräferenzen zu definieren. Der Mechanismus soll dann aktiv die Nutzer gemäß den hinterlegten Profilen mit Informationen versorgen.

Stand des Projekts im Januar 1999

Obwohl die Sach- und Personalmittel für dieses Projekt erst am Jahresende 1998 zur Verfügung standen, wurde bereits im zweiten Halbjahr 1998 mit den Arbeiten am Projekt begonnen:

- Die konzeptionellen Vorarbeiten des Projekts sind so weit fortgeschritten, daß innerhalb der nächsten Wochen mit dem Beginn der Implementierung auf einem Testsystem zu rechnen ist. Die Dokumente werden in einer ORACLE Datenbank gespeichert, der Upload und die Erfassung des Präkatalogisats erfolgen über ein Web-Interface.
- Der Workflow für die Einspielung des vom Autor erfaßten Präkatalogisats in die Verbunddatenbank des SWB wurde definiert und greift auf Verfahren zurück, die bereits erprobt sind. Es erfolgt dann ein Rückfluß als Titelaufnahme im Rahmen der normalen Katalogisierungsarbeiten der UB Ulm.

- Sehr viel Aufwand wurde in die Entwicklung des Datenmodells investiert, das einerseits nicht zu komplex sein darf und andererseits leistungsfähig genug sein muß, um zukünftigen Anforderungen oder Änderungen gewachsen zu sein. Das Datenmodell wurde darüber hinaus so konzipiert, daß es zu anderen Systemen (z.B. Landesprojekt IBM Digital Library) kompatibel ist. Eine Umsetzung dieses Datenmodells auf eine reale ORACLE Testinstallation ist zur Zeit in Arbeit.
- Eine grobe Designstudie für die Gestaltung des webbasierten Userinterface wurde entwickelt und wird in Kürze in überarbeiteter Form umgesetzt.
- Bei der für das Projekt und den späteren Produktionsbetrieb erforderlichen Hardware handelt es sich um ein zertifiziertes Cluster-System der Fa. Dell, das unter *NT Server 4.0 Enterprise Edition* betrieben wird. Die Serverhardware setzt sich zusammen aus 2 Dell PowerEdge 2300 mit je einer PII-400 MHz CPU, 128 MB ECC SDRAM, 2 * 4.5 GB Systemplatten in RAID Level 1, einem internen und einem Cluster RAID-Controller, die auf ein gemeinsames externes Stagesystem in RAID Level 5 mit einer Bruttospeicherkapazität von 54 GB zugreifen. Die Anbindung an das LAN der Universität wird über je ein Fast-Ethernet Interface erfolgen. Mit dem an der Universität Ulm erstmaligen Einstieg in die Clustertechnik unter NT bei einem Produktionssystem, soll zunächst Know-how in diesem innovativen Bereich aufgebaut und letzten Endes die Verfügbarkeit des Volltextservers selbst hohen Ansprüchen genügen können. Zudem kann bei entsprechendem Loadbalancing der Cluster auch weitere Dienste erbringen.

Der systemtechnische Aufbau des Clusters ist zur Zeit gerade in Arbeit; erste Evaluierungen der Clusterfunktionalität konnten bereits durchgeführt werden. Daran wird sich ein Test der beiden wichtigsten Anwendungsprogramme IIS4.0 und ORACLE in dieser Umgebung anschließen. Mit der Fertigstellung der Systemtechnik ist erst nach Abschluß der zwingend erforderlichen Testläufe zu rechnen; das wird voraussichtlich Mitte bis Ende Februar der Fall sein.