Konzepte und Systemarchitektur des Dokumentenservers KLUEDO

Wolfgang Eiden

4. MathBibNet-Treffen in Göttingen
14. bis 15. Dezember 1999

Zusammenfassung

Auf KLUEDO, dem Kaiserslauterer uniweiten elektronischen Dokumentenserver, können Angehörige der Universität Kaiserslautern wissenschaftliche Dokumente in elektronischer Form im Internet und damit weltweit veröffentlichen. Der Dokumentenserver entstand auf Initiative der Fachbereiche und der Universitätsbibliothek Kaiserslautern im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten *GLOBAL INFO* - Vorprojektes. Hierbei basiert der Dokumentenserver stellenweise auf dem (während des *MathNet*-Projektes des deutschen Forschungsnetzes (DFN) und des *MathBibNet*-Projektes der deutschen Forschungsgemeinschaft (DFG) aufgebauten) Mathematik-Preprintserver der Universität Kaiserslautern. Dieser Artikel umreißt die Aspekte, die hinter der Entwicklung des Dokumentenservers in der vorliegenden Form gestanden haben und gibt einen groben Überblick über die Systemarchitektur. Detailinformationen und eine Beschreibung des Systems findet man in meiner Arbeit *Propädeutik metadatenbasierter Publikationsserver: Ontologie und konkordante Implementierung* [2].

1991 Mathematics Subject Classification: 68M05, 68P20, 68T30, 68T35

1 Einleitung

Gerade in einer von Globalisierungstendenzen geprägten Zeit wie heute - in der sowohl das Angebot als auch die Nachfrage an Informationen exponentiell wächst - gewinnen Dokumentenserver an zunehmender Bedeutung: Die weltweite Verfügbarkeit und Recherchierbarkeit von elektronisch gespeicherten Ressourcen (bei gleichzeitig minimalen Kosten) sind zentrale Forderungen der Informationsgesellschaft und Antriebsmotor für internationale Informationsstrukturen. Daher hat bei der Konzeption und der Realisierung des Dokumentenservers KLUEDO a priori globales Denken im Vordergrund gestanden. Hieraus ergaben sich eine Vielzahl von Forderungen aus unterschiedlichsten Bereichen (wie beispielsweise Qualitätsmanagment, Informationsrepräsentation, Informationsretrieval oder Verfügbarkeit) deren Umsetzung international genormte Konzepte voraussetzten.

2 Konzepte

Das System sollte auf internationalen Standards und offenen Schnittstellen basieren. Daher schien die Nutzung des Internets mit seinen weltweit anerkannten Informationsdiensten (wie beispielsweise *World Wide Web* oder *electronic mail*) mit implizierter Plattformunabhängigkeit die logische Konsequenz. Ein weiterer Vorteil hierbei ist die Möglichkeit der Nutzung des Angebotes ohne die Installation spezieller Software. Aufgrund der jedoch rasanten Entwicklung des Internets und der sich damit ändernden Standards, sollte der Publikationsserver auf einem zentralen System basieren, da es nur dann mit einem vertretbarem Aufwand möglich ist, mit den aktuellen Entwicklungen Schritt zu halten und den sich ändernden Bedürfnissen zu genügen. So erfordert beispielsweise auch der Datenaustausch mit Bibliothekskatalogen und fachbezogenen Servern zentrale Anknüpfungspunkte zur Realisierung eines Systems mit offenen Schnittstellen. Diese sind für die Akzeptanz eines Publikationsservers von zentraler Bedeutung, da nur durch Einbettung in bestehende Infrastrukturen die Informationen einem Spektrum von möglichst vielen Benutzern zur Verfügung gestellt werden. Entscheidend ist desweiteren die Art und Weise des Angebots: Das Web-User-Interface (als Verbindung zwischen Internet und dem verwendeten Datenbankmanagmentsystem) sollte daher eine benutzerfreundliche Nutzung erlauben, wobei eine multilinguale Ausrichtung, eine klare Struktur, die Konzentration auf das Wesentliche und die Beschränkung auf ein qualitativ hochwertiges Informationsangebot von oberster Bedeutung waren und sind. Die fachspezifische inhaltliche Erschließung von Ressourcen (auf verbale und klassifikatorische Art), themenspezifische Veröffentlichungslisten und übersichtlicher Suchergebnisse leisten hierbei einen nicht unwesentlichen Beitrag. Um eine effizientes Retrieval zu gewährleisten, verwendet KLUEDO zur Beschreibung der Ressourcen das mittlerweile international verbreitete *Dublin Core Metadata Set* und setzt als Retrievalsystem *HARVEST* ein.

3 Systemarchitektur

KLUEDO setzt einen Webserver mit PHP-Unterstützung und die Installation der Programmiersprache Perl voraus. Als HTTP-Server wird der *Apache* Webserver verwendet. Es handelt sich hierbei um einen freien, open-source Webserver mit Distributionen für diverse Server-Betriebssysteme. Bei PHP handelt es sich um einen Hypertext-Preprocessor, also um eine serverseitig interpretierte, in HTML eingebettete Skriptsprache. KLUEDO verwendet PHP um ein einheitliches Aussehen aller Seiten des Dokumentenservers sicherzustellen und ein Session-Management zu ermöglichen. Dazu wurden Routinen entwickelt, die das eigentliche HTML-Dokument in einen Rahmen mit integriertem Navigationssystem einbettet.

3.1 Struktur der einzelnen Fachbereiche

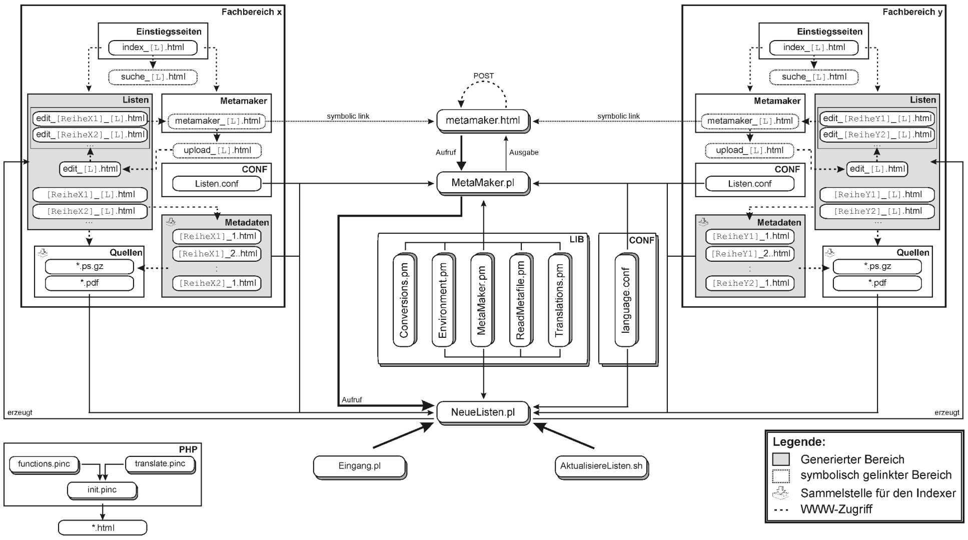
Alle Fachbereiche sind bezüglich der Struktur identisch. Die HTML-Seiten für den Index, den Metamaker, die Suche und den Dateiupload sind symbolische Links auf zentrale PHP-Seiten. Im Wesentlichen unterscheiden sich die Fachbereiche nur in der Konfigurationsdatei welche die Veröffentlichungsreihen des entsprechenden Fachbereiches definiert. Die Indexseiten verweisen (in ihrer Funktion als Einstiegsseiten) via URL auf die Suche und die einzelnen Reihen des Fachbereiches. Desweiteren wird dem Benutzer die Möglichkeit der Neuerzeugung und Bearbeitung von Metadaten gegeben.

3.2 Konfigurationsdateien und Bibliotheken

Als zentrale Konfigurationsdatei ist zur Zeit nur eine Datei vorhanden, die zur Sprachübersetzung einer (mit einem Keyword assozierten) Phrase dient. Die Übersetzungsroutine und der Dateiaufbau sind so konzipiert, dass eine Sprachunterstützung für weitere Sprachen problemlos möglich ist. Um die, in Perl implementierten, Verwaltungsskripte möglichst einfach zu halten, wurden zentrale Bibliotheken geschaffen, die sowohl bei der Metadaten- als auch bei der Listenverwaltung verwendet werden: Die Bibliothek *Environment.pm* stellt Routinen zur Verfügung mit dessen Hilfe die Umgebungsvariablen (wie z.B. Basisverzeichnisse und interne Sprachcodes) ermittelt werden können. In *Conversions.pm* werden diverse Konvertierungs- und Verifikationsoperationen für Datentypen, CGI, HTML, Normen (beispielsweise ISO8601 normierte Datumsangaben) und Klassifikationen (beispielsweise MSC1991-Klassifikation) definiert. Zum Erzeugen von Metatags gemäß Dublin Core dient die Bibliothek *Metamaker.pm*. Hierbei werden die zu erzeugenden Informationen vorher (gemäß Normierung und verwendeten Klassifikationen) geprüft und gegebenenfalls korrigiert. *ReadMetafile.pm* stellt eine effiziente Routine zum Einlesen der Metadaten aus einer Datei zur Verfügung. Die Übersetzung von Begriffen in verschiedene Sprachen wird durch die die Bibliothek *Translations.pm* realisiert.

3.3 Metadatenmanagement

Zur Erreichung eines minimalen Aufwandes zur Administration des Dokumentenservers war die Entwicklung zentraler Module zur Erzeugung und Verarbeitung von Metadaten notwendig. Neben den Bibliotheken, die unter anderem das Grundgerüst zum Metadatenmanagement zur Verfügung stellen, bedeutete dies sowohl die Bereitstellung eines zentralen Web-Interfaces (zur Neuaufnahme und Bearbeitung von Metadaten), als auch die eines zentralen Skriptes zur Erzeugung der verschiedenen Veröffentlichungsübersichten. Die Metamaker der einzelnen Fachbereiche sind als symbolische Links auf einen zentralen Metamaker realisiert. Das zugrundeliegenden Skript kann aufgrund der Struktur des Dokumentenservers durch URL-Splitting zwischen den verschiedenen Fachbereichen und Sprachen differenzieren.



Literatur

[1] Dalitz W., Grötschel M., Heyer G., Lügger J., Sperber W.: *Informationsdienste für die Mathematik im Internet (Math-Net)*. DFN-Vorhaben DT11. Technical Report TR 96-13. Dezember 1996. Konrad-Zuse-Zentrum für Informationstechnik Berlin

[2] Eiden W.: *Propädeutik metadatenbasierter Publikationsserver: Ontologie und konkordante Implementierung*. Oktober 1999. Universität Kaiserslautern.

[3] Grötschel M., Dalitz W., Winczewski V.: *Math-Net: Ein Informationssystem der Mathematik.*. Preprint SC 98-25. Oktober 1998. Konrad-Zuse-Zentrum für Informationstechnik Berlin

[4] Lehmann A.: *Empfehlungen zum Aufbau eines Servernetzwerkes für elektronische Hochschulpublikationen*. September 1998. Arbeitsgruppe Volltexte und Hochschulpublikationen des Ministeriums für Wissenschaft, Forschung und Kunst Baden-Württemberg.