

Verwaltung kostenpflichtiger digitaler Dokumente im eVerlage-System

Dietrich Boles, Cornelia Haber, Frank Oldenettel*
Informatik-Institut OFFIS
Escherweg 2
26121 Oldenburg

Zusammenfassung

eVerlage ist eine Sonderfördermaßnahme im Rahmen des BMBF-Förderkonzeptes Global Info. Zehn deutsche Verlage aus Wissenschaft und Technik haben sich zu einer Arbeitsgemeinschaft zusammengeschlossen mit dem Ziel, einen relevanten Teil ihrer Bücher Wissenschaftlern und Studierenden als hypermediale Dokumente über das Internet anzubieten. Partner dieser Arbeitsgemeinschaft sind sieben Forschungseinrichtungen, Hochschulen und Bibliotheken, die die Modellierung, die technische Realisierung und den Betrieb des Online-Angebots übernehmen.

Das eVerlage-System dient den Verlagen als Testumgebung für die Erprobung verschiedener elektronischer Angebotsformen, Abrechnungsmodelle und Zahlungsverfahren für den Zugriff auf die gespeicherten Dokumente. Es ist als verteiltes System realisiert und über Standard-WWW-Browser nutzbar. Nach einem kurzen Überblick über die Architektur und Funktionalität des eVerlage-Systems werden in diesem Artikel schwerpunktmäßig Aspekte der Dokumentenverwaltung beschrieben.

1 Einleitung

Das eVerlage-Projekt ist eine Sonderfördermaßnahme (SFM9) im Rahmen des *Global-Info* Programms, welches vom BMBF gefördert wird. Das Projekt läuft seit 01. August 1999 und hat eine Laufzeit von 3 Jahren. Aufgabe des Projektes ist die Entwicklung und Erprobung eines Systems für die kommerzielle Bereitstellung und Nutzung von multimedialen Volltextangeboten im Internet.

Diese kurze Aufgabenbeschreibung gibt bereits einige Rahmenbedingungen für das Projekt fest vor: Um ein Volltext-System zu erproben und nach Abschluss des Projektes wirtschaftlich tragfähig zu betreiben, muss eine kritische Masse an hochwertiger Fachliteratur im Volltext zur Verfügung stellen. Dies wird über die Einbindung namhafter Verlage in das Projekt erreicht. Ein wei-

terer wichtiger Faktor für den Erfolg des eVerlage-Systems ist die Anzahl der Nutzer. Um möglichst viele Nutzer zu erreichen, ist das System ohne vorherige Installationen über einen Standard-WWW-Browser zugreifbar. Ein zusätzlicher Nutzerkreis wird durch die Anbindung an existierende Bibliothekssysteme angesprochen. Damit das System sowohl für einzelne Nutzer als auch für Nutzergruppen attraktiv ist, müssen die Abrechnungs- und Zahlungsverfahren flexibel sein.

Neben diesen eher allgemeinen Rahmenbedingungen gibt es noch eine Reihe konkreter Anforderungen an das System, die im folgenden Abschnitt diskutiert werden.

2 Anforderungen

Die Anforderungen an das eVerlage-System lassen sich nach den verschiedenen Nutzergruppen gliedern, die mit dem System zu tun haben. In erster Linie sind hier die klassischen Nutzer zu nennen, die das System zur Literaturrecherche und zum Zugriff auf die Volltexte der Dokumente nutzen. Neben diesen Nutzern gibt es die Gruppe der Administratoren, die für den technischen Betrieb des Systems verantwortlich sind. Die dritte Nutzergruppe bilden die Informationsanbieter, die ihre Dokumente über das eVerlage-System anbieten.

2.1 Anforderungen aus Sicht der Nutzer

Die klassischen Nutzer (Wissenschaftler, Studierende, ...) wollen das System zur Literaturrecherche und zum Zugriff auf die Dokumente nutzen. Das System muss also eine intuitive und dennoch mächtige Suche sowohl auf den bibliographischen Daten als auch auf den Volltexten ermöglichen. Nachdem der Nutzer ein Dokument gefunden hat, muss es möglich sein, direkt auf das Dokument zuzugreifen bzw. online eine Lizenz für den Zugriff auf das Dokument zu erwerben. Zu diesem Zweck muss das System elektronische Zahlungsverfahren unterstützen.

* e-mail: {boles, haber, oldenettel}@offis.de

Um den Nutzern einen Überblick über den vorhandenen Dokumentenbestand zu ermöglichen, ist eine Navigationsfunktion nötig, die den Dokumentenbestand anhand vorgegebener Kriterien wie Autorennamen, Klassifikation o.ä. auflistet.

2.2 Anforderungen aus Sicht der Anbieter

Die Anbieter möchten ihre Dokumente im eVerlage-System möglichst einfach verwalten können. Dazu zählt sowohl das Einspielen neuer Dokumente als auch das Ändern vorhandener Dokumente oder das Löschen alter Dokumente.

Ein anderer wichtiger Aspekt aus Sicht der Anbieter ist der Schutz ihrer Dokumente vor Veränderung, Weitergabe oder unrechtmäßiger Nutzung.

2.3 Anforderung aus Sicht der Administratoren

Das Hauptanliegen der Administratoren ist es, den Administrationsaufwand möglichst gering zu halten. Das System muss also möglichst stabil laufen, keine Handarbeit erfordern und muss im laufenden Betrieb wartbar sein.

3 Architektur

Unter Berücksichtigung der Anforderungen an das System aus der Sicht der verschiedenen Nutzergruppen wurde die Systemarchitektur in Abbildung 1 entwickelt.

Das eVerlage-System besteht damit im wesentlichen aus drei Komponenten: Nutzeragenten, Zentralagent und Anbieteragenten. Nutzeragenten bilden die Zugriffsschnittstelle der Nutzer auf das System. Anbieteragenten verwalten die Dokumente und der Zentralagent vermittelt zwischen Nutzeragenten und Anbieteragenten. Die einzelnen Agenten sind als unabhängige Java-Komponenten realisiert, die über RMI kommunizieren.

3.1 Zentralagent

Der Zentralagent besteht aus sieben Managern: Der Nutzermanager verwaltet sämtliche Nutzerdaten (Namen, Logins, ...). Der Gruppenmanager kapselt alle Gruppen Daten wie bspw. die Nutzer, die zu einer Gruppe gehören u.ä. Der Agentmanager gibt Auskunft über alle angeschlossenen Anbieter- und Nutzeragenten. Der Angebotsmanager verwaltet die Angebote zu den einzelnen Dokumenten (Kosten, Lizenzmodelle, usw.). Der Bestellmanager ist verantwortlich für den gesamten Bestellvorgang (in Zusammenarbeit mit den anderen Managern). Um die Abwicklung der Zahlungsformalitäten kümmert sich der Zahlungsmanager und für die Verwaltung der Lizenzdaten zeichnet sich der Lizenzmanager verantwortlich.

3.2 Nutzeragent

Nutzeragenten bilden die Zugriffsschnittstelle der Nutzer zum eVerlage-System. Sie ermöglichen die Suche auf dem Dokumentenbestand, die Navigation durch den Dokumentenbestand, die Bestellung eines oder mehrerer Dokumente und den Zugriff auf die Dokumente. Darüber hinaus bieten die Nutzeragenten Hilfe zum System an.

Wie die Abbildung 1 zeigt, lassen sich auch spezielle Nutzeragenten integrieren, um z. B. das eVerlage-System an existierende Bibliothekssysteme anzubinden.

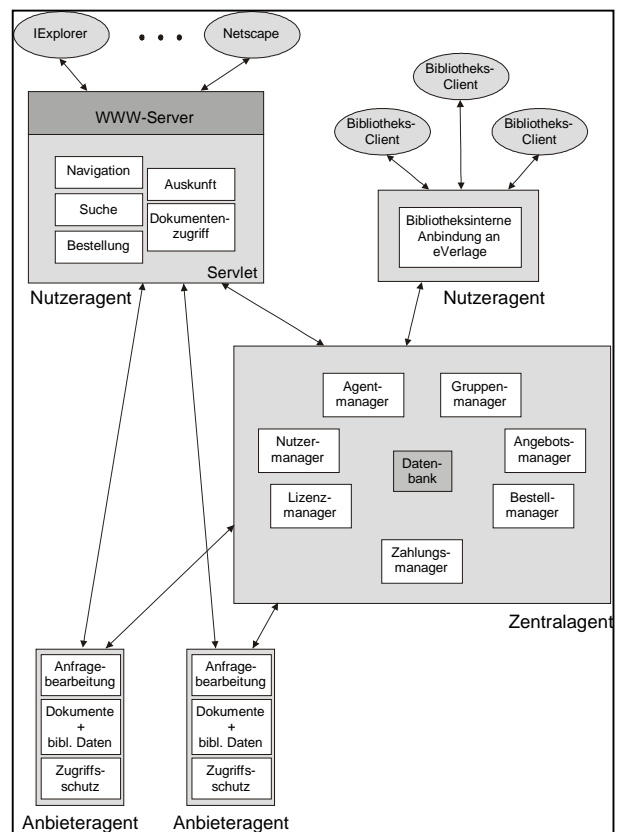


Abbildung 1: Architektur des eVerlage-Systems

3.3 Anbieteragent

Anbieteragenten kapseln die Spezifika der darunterliegenden Datenbanken. Sie verwalten sowohl die bibliographischen Daten zu den Dokumenten als auch die Dokumente selbst. Anbieteragenten bearbeiten Rechercheanfragen, die sie von den Nutzeragenten erhalten und liefern die Anfrageergebnisse zurück. Eine weitere Aufgabe, die den Anbieteragenten zufällt, ist der Schutz der Dokumente vor unberechtigten Zugriffen und vor Missbrauch.

4 Dokumentenverwaltung

Die Verwaltung der Dokumente ist eine der Kernaufgaben eines Anbieteragenten. Ziel der Dokumentenverwaltung muss es sein, möglichst gute Rechercheergebnisse zu liefern, wobei die Verwaltung der Dokumente möglichst einfach und ohne Aufwand möglich sein soll. Eine wichtige Grundlage für die Recherche stellen die bibliographischen Daten dar.

4.1 Bibliographische Daten

Die Forderungen nach guten Rechercheergebnissen und einer einfachen Dokumentenverwaltung stellen konkurrierende Ziele dar. In der Praxis finden sich zwei grundsätzlich verschiedene Ansätze zur Dokumentenverwaltung in digitalen Bibliotheken und Suchmaschinen, je nachdem, welches Ziel höher bewertet wird. Die erste Möglichkeit, die von klassischen Literaturdatenbanken und einigen Technical Report Systemen wie z. B. NCSTRL [LSDK95] unterstützt wird, ist es, Metadaten zu den einzelnen Dokumenten wie bspw. die bibliographischen Daten, als strukturierte Daten in einer Datenbank vorzuhalten und diese Datensätze per Hand zu erstellen. Die andere Alternative ist es, diese Daten nicht explizit zu speichern, sondern die Suche in den bibliographischen Daten durch eine eingeschränkte Volltextsuche zu approximieren. Dieser Ansatz wird von z. B. von NZDL [WNC96] verfolgt. NZDL realisiert beispielsweise die Suche nach Autoren durch eine Volltextsuche nach den Autorennamen innerhalb der ersten Seite eines Dokumentes. Für das eVerlage-System haben wir uns bewusst für die manuelle Erstellung der Metadaten entschieden, da diese Variante bessere Rechercheergebnisse liefert und damit in unseren Augen den Mehraufwand rechtfertigt.

4.2 Dokumentenhierarchie

Zusätzlich zu den rein bibliographischen Angaben werden im eVerlage-System Metadaten zur Dokumentenhierarchie und damit zur logischen Verknüpfung von Dokumenten vorgehalten. Oberste Elemente der Dokumentenhierarchie sind Zeitschriften und Serien. Eine Zeitschrift besteht aus einzelnen Bänden, die wiederum aus Ausgaben bestehen. In einer Ausgabe befinden sich verschiedene Artikel, die weiter in Abschnitte untergliedert werden können. Abschnitte lassen sich rekursiv in (Unter-)Abschnitte unterteilen. Abbildung 2 zeigt, welche Beziehungen zwischen Dokumenten das eVerlage-System unterstützt.

Auf der Basis der Dokumentenhierarchie kann der Nutzer dann ausgehend von einzelnen Dokumenten in der Dokumentenhierarchie navigieren und beispielsweise

von einem Artikel einer Zeitschrift über den Zeitschriftenband zu den Artikeln der nächsten Ausgabe gelangen.

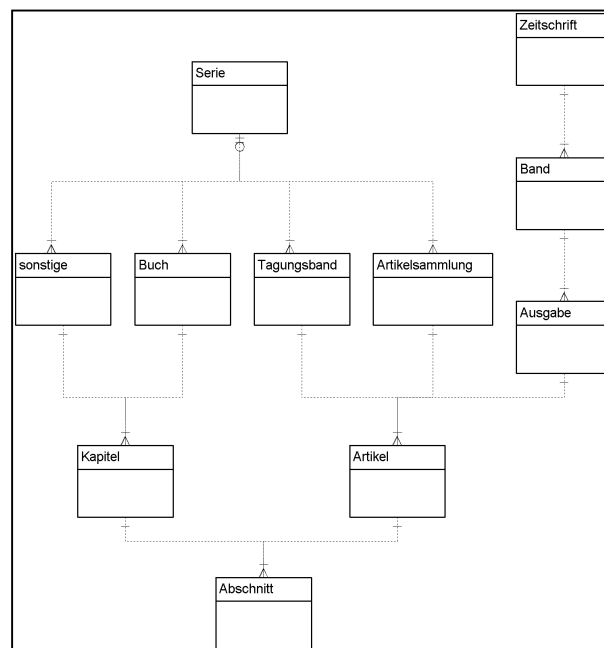


Abbildung 2: Dokumentenhierarchie

4.3 Dokumentenstruktur

Die Dokumente im eVerlage-System sind speziell für eine Bereitstellung über das Internet aufbereitet. Sie liegen in den Formaten pdf oder html vor. Um lange Ladezeiten zu vermeiden und eine übersichtliche Darstellung am Bildschirm zu ermöglichen, können die einzelnen Dokumente in mehrere Dateien untergliedert sein. Bei HTML-Dokumenten, wo die Bilder erst vom Browser in den Text eingebunden werden, liegen auch die Graphiken in separaten Dateien vor.

Zusätzlich zur Dokumentenhierarchie, die die Verknüpfung der Dokumente untereinander festlegt, wird daher in der Dokumentenverwaltung die Struktur eines Dokumentes abgespeichert. Strukturinformationen geben in erster Linie Auskunft darüber, welche Dateien zu einem Dokument gehören. Sie können jedoch auch Angaben zu einzelnen ausgezeichneten Dokumentbestandteilen enthalten, wie etwa eine Festlegung, welche Datei das Inhaltsverzeichnis enthält, welche Dateien als Leseprobe für jedermann kostenlos zugänglich sein sollen oder welche Dateien das Literaturverzeichnis bilden. Die Speicherung der Dokumentenstruktur ermöglicht es Nutzern, gezielt auf einzelne Dokumentbestandteile zuzugreifen.

4.4 Verwaltung der Volltexte

Neben der Verwaltung der bibliographischen Daten und der Einordnung innerhalb der Dokumentenhierarchie müssen die Volltexte verwaltet werden. Dazu wird der Volltext der Dokumente in das (volltextfähige) Datenbanksystem eingespielt und über eine Relation mit den bibliographischen Daten verknüpft.

4.5 Zugriff auf Dokumente

Dokumente können aus einzelnen Dateien bestehen. Bei der Dokumentenrecherche wird daher zusammen mit der Dokumentenbeschreibung die Startdatei eines Dokumentes an den Nutzer geliefert. Greift der Nutzer nun über den WWW-Browser auf die Startseite oder eine andere Seite des Dokumentes zu, so ist dies ein Zugriff auf eine einzelne Datei und nicht auf das Dokument als Ganzes. In der Datenbank wird dann überprüft, zu welchem Dokument die Datei gehört und ob der Nutzer eine Lizenz für das entsprechende Dokument besitzt. Besitzt der Nutzer eine Lizenz, so kann er auf das Dokument zugreifen. Ist keine Lizenz vorhanden, so kann der Nutzer online eine erwerben.

4.6 Wartung des Dokumentenbestandes

Wie in Abschnitt 4 bereits ausgeführt, haben wir uns für das eVerlage-System bewusst für eine manuelle Erstellung und Wartung der Metadaten entschieden. Da mit dem Einspielen, Ändern und Löschen eines Dokumentes jedoch ein gewisser Aufwand verbunden ist, werden Tools benötigt, um den Aufwand möglichst gering zu halten. Im Rahmen des eVerlage-Projektes wird daher ein spezielles Tool entwickelt, das Unterstützung für das Einspielen und Löschen von Dokumenten bietet. Über dieses Tool werden die bibliographischen Angaben eingegeben, wobei auf bereits existierende Autoren, Editoren, Verlage, Serien, usw. zurückgegriffen wird. Im nächsten Arbeitsschritt wird das Dokument mit Hilfe des Dokumentenverwaltungstools entweder in die Dokumentenhierarchie eingeordnet oder als unabhängiges neues Dokument deklariert. In einem letzten Arbeitsschritt wird dann die Struktur des Dokumentes definiert.

5 Zusammenfassung

Mit dem eVerlage-System wurde ein erster Prototyp einer kostenpflichtigen digitalen Bibliothek erstellt. Die leistungsfähige Recherchefunktion kombiniert mit einer Navigation innerhalb des Dokumentenbestandes erfordert hierbei eine komplexe Dokumentenverwaltung sowie Administrationstools, die die Verwaltung des Dokumentenbestandes unterstützen. Die Akzeptanz des Systems bei den Nutzern wird stark von den Inhalten des

Systems und den angebotenen Lizenzmodellen und Zahlungsverfahren abhängen.

Literatur

- [LSDK95] Carl Lagoze, Erin Shaw, James R. Davis and Dean B. Krafft. Dienst: Implementation Reference Manual. Technical Report TR95-1514 Computer Science Department, Cornell University, May 1995.
- [WNC96] Ian H. Witten, Craig, G. Nevill-Manning and Sally Jo Cunningham. Building a Digital Library for Computer Science Research: Technical Issues. In *Proceedings of the 19th Australasian Computer Science Conference, Melbourne, Australia, January 31-February 2, 1996*.