

Anhang 3

Veröffentlichungen

Inhaltsverzeichnis

EIN INTERNET-BASIERTES UMWELTDOKUMENTENSYSTEM FÜR KLEINE NUTZERGRUPPEN.....	179
Veröffentlichung auf dem Workshop des Arbeitskreises Umweltdaten- banken, 10.06.-11.06.1999, Karlsruhe	
INTERNETBASIERTE DOKUMENTENVERWALTUNG HETEROGENER UMWELTDOKUMENTENBESTÄNDE	197
Veröffentlichung auf der Umweltinformatiktagung, 30.08.-01.09.1999, Magdeburg	
INTRANET-BASIERTE VERWALTUNG UND VERFÜGBARMACHUNG MULTIMEDIALER UMWELTDOKUMENTENBESTÄNDE	215
Einreichung für den 3. Workshop des Arbeitskreises Hypermedia im Umweltschutz, 23.03.-24.03.2000, Ulm	

Ein Internet-basiertes Umweltdokumentsystem für kleine Nutzergruppen

Klaus Tochtermann, Andreas Kussmaul und Katrin Pursche¹

Zusammenfassung

Die Entwicklungen der vergangenen Jahre erlauben in Zukunft auch Nicht-Experten kleinere Umweltdokumentbestände aufzubauen, zu warten und inhaltlich zu erschließen. Vor diesem Hintergrund wird sich die Nutzung von Internettechnologien in Zukunft weg von großen, nur von Experten zu beherrschbaren Dokumentbeständen und zugehörigen Systemen, hin zu kleineren und auf spezielle Bedürfnisse der Nutzer zugeschnittene Dokumentbeständen und Systemen bewegen. Dieser Beitrag stellt ein solches System und mögliche technische Realisierungen zur Anbindung von (Umwelt-)Datenbanken über das Internet vor. Die verschiedenen Alternativen zur technischen Realisierung werden in dem Beitrag diskutiert und bewertet.

1. Einleitung

Im vergangenen Jahr endete die erste Phase der amerikanischen Initiative „Digital Libraries“. Während in dieser Phase sechs Universitäten mit Arbeiten zu Grundlagenproblemen, wie Verfügbarmachung riesiger Dokumentbestände, gefördert wurden, werden in der derzeit vorbereiteten Phase 2 mehr Schwerpunkte auf Bereiche wie Erzeugung von Inhalten, Zugriff auf Inhalte sowie Nutzen und Nutzbarkeit von Inhalten gelegt.

¹ K. Tochtermann, A. Kussmaul, K. Pursche

Forschungsinstitut für anwendungsorientierte Wissensverarbeitung, Postfach 2060,
D-89081 Ulm; E-Mail: {tochterm | kussmaul | pursche}@faw.uni-ulm.de.

Die im Special Issue „Digital Libraries“ der Zeitschrift IEEE Computer (Computer 1999) dargelegten Erfahrungen aus der Phase 1 zeigen, dass für das 21. Jahrhundert erwartet wird, dass sich das Arbeiten mit Wissen und Informationen dramatisch ändern wird. In den vergangenen Jahren wurden riesige Datenbestände von professionellen Fachkräften katalogisiert, automatisch indiziert und anschließend der Öffentlichkeit über das Internet zugänglich gemacht (Schatz 1999). Aufgrund dieser Vorgehensweise konnte sich die Öffentlichkeit in den letzten 10 Jahren mit Internettechnologien vertraut machen. Inzwischen entsteht mehr und mehr das Bedürfnis, nicht nur recherchierend mit Dokumentbeständen zu arbeiten, sondern diese auch verändern zu dürfen. Die Möglichkeit der Veränderung spielt sich auf mindestens zwei Stufen ab. In der ersten Stufe werden von Systemen Möglichkeiten angeboten, die es erlauben, die Systeme an persönliche Anforderungen anzupassen. Ein solches System ist etwa PADDLE (Personal Adaptable Digital Library Environment) (Hicks et al. 1999). Damit nehmen Nutzer eine aktive Rolle in Form von Autoren ein. Um sie in ihrer Funktion als Autor mit der damit verbundenen Komplexität nicht zu überfordern, ist es sinnvoll, dass zunächst Systeme entwickelt werden, mit deren Hilfe von Nicht-Experten kleinere organisationsinterne Datenbestände aufgebaut, erschlossen und nutzbar gemacht werden können. Mit PADDLE ist es bspw. möglich, erhobene Metainformation zu verändern und zu personalisieren, ohne dass der Originaldatenbestand davon betroffen ist. In der zweiten Stufe werden Systeme ihren Nutzern erlauben, auch die Dokumentbestände an sich zu verändern, und zwar unabhängig davon, wer der Eigentümer dieser Dokumentbestände ist. Diskussionen über diese zweite Stufe werden derzeit unter dem Schlagwort „cross-publishing“ geführt. Zukünftige Entwicklungen im Bereich Bereitstellung von Online-Informationen müssen daher diesem Trend Rechnung tragen.

Vor diesem Hintergrund wurde am FAW im Rahmen einer Diplomarbeit ein Internet-basiertes Dokumentenarchiv erstellt, das kleinere Nutzergruppen (ca. 10

Personen) in Aufbau, der Erschließung und der Nutzung von Umweltdokumenten unterstützt. Es wurde bewußt Wert darauf gelegt, zunächst nur einer überschaubaren Zahl von Nutzern Autorenrechte einzuräumen. Damit ist es möglich, die entstehenden Phänomene, aber auch Schwierigkeiten besser zu erfassen und zu kontrollieren. Auf der Basis der gesammelten Erfahrungen können zu einem späteren Zeitpunkt Anforderungen abgeleitet werden, die zu berücksichtigen sind, um einer beliebig großen Zahl von Nutzern Autorenrechte für ein Dokumentenarchiv einzuräumen. Die Bedeutung dieser Arbeit wird dadurch unterstrichen, dass das Ministerium für Umwelt und Verkehr Baden-Württemberg das FAW mit der Erstellung eines solchen Systems für die Verwaltung von Umweltdokumenten in Behörden beauftragt hat (Henning et al. 1999).

Der Beitrag stellt nun die technischen Aspekte der Realisierungsmöglichkeiten in den Vordergrund und ist wie folgt aufgebaut: Abschnitt 2 stellt das System an sich mit prinzipiellen Arbeitsabläufen dar. Abschnitt 3 diskutiert die technische Realisierung, bevor in Abschnitt 4 der Beitrag mit einer Bewertung und einem Ausblick beendet wird.

2. Anforderungen an das System und pinzipieller Arbeitsablauf

Mit dem System sollen Nutzer bei dem Aufbau, der Erschließung und der Nutzung von ihren täglichen Arbeitsdokumenten unterstützt werden. Darunter sind Dokumente zu verstehen, die schnell zugreifbar und recherchierbar sein müssen, ohne dass zuvor umfangreiche Metadaten (Daten über Daten) erhoben werden müssen. Versionenverwaltung sowie andere typische Eigenschaften von (Internet-basierten) Dokumentverwaltungssystemen, wie BSCW (BSCW 1999) oder AltaVista-Forum (AltaVista 1999), wurden in dieser Phase nicht berücksichtigt. Bei der Zielgruppe der Nutzer wurde ferner davon ausgegangen,

dass kein umfangreiches technisches Know-how vorliegt. Vor diesem Hintergrund ist es wichtig, die Komplexität des Systems hinter einfach zu bedienenden Benutzungsoberflächen zu verbergen.

Für das Dokumentenarchiv wurden drei Komponenten vorgesehen (vgl. Abb. 3), welche jeweils auf einem separaten Rechner in einem Intranet oder im Internet platziert werden können.

Eine Komponente bearbeitet Anfragen an eine „Accounts“-Datenbank. In dieser Datenbank werden Informationen über die Benutzer festgehalten, wie zum Beispiel der Status des Benutzers (Gast, Autor oder Administrator) oder seine Emailadresse. Die Komponente ist erforderlich, um prinzipiell nachhalten zu können, welcher Nutzer welche Änderungen am Dokumentbestand vorgenommen hat. Da nur ein kleiner Nutzerkreis angesprochen werden soll, wurde kein besonderer Wert auf Sicherheitsaspekte, etwa bei der Vergabe und Verwaltung von Paßwörtern, gelegt.

Eine weitere, sehr mächtige Komponente beinhaltet einen WWW-Server mit einem Index Server und einem FTP-Dienst sowie alle Dokumente, welche über das Dokumentenarchiv verwaltet werden und vom Index Server indiziert werden. Neben den rechenintensiven Suchanfragen an den Index Server über Netzwerkverbindungen mit dem HTTP-Protokoll werden auch FTP-Sitzungen zur Dateiübertragung abgehalten. Hierüber wird den Nutzern die Möglichkeit eingeräumt, in der Rolle eines Autors Dokumente zu verändern. Über die Dateitransfers mit FTP hinaus werden weitere Dateioperationen auf dem Dateisystem ständig zur Aktualisierung der Indizes des Index Servers ausgeführt.

Die dritte Komponente verwaltet eine Datenbank. Diese Datenbank ist der Umweltthesaurus GEMET (General European Multilingual Environmental Thesaurus) des Umweltbundesamtes und findet bei der Suche nach synonymen Begriffen, Unter- und Oberbegriffen sowie Schwesterbegriffen Anwendung. Die Komponente ist im Vergleich zu den anderen nicht erforderlich, um Nutzer in

ihrer Rolle als Autor zu unterstützen. Allerdings erleichtert sie die inhaltliche Erschließung von Dokumenten, da anhand des kontrollierten Vokabulars eine systematische Verschlagwortung möglich ist. Der prinzipielle Arbeitsablauf mit dem System wird in den folgenden Abschnitten beschrieben (mehr Informationen hierzu in (Henning et al. 1999)).

2.1 Autorenkomponente

Die Oberfläche der Autorenkomponente wird in Abbildung 1 dargestellt. Die Autorenkomponente wird im Prototypen aus der Rechercheoberfläche heraus gestartet. Im linken Teil des Fensters wählt der Autor das Verzeichnis aus, in welches er das von ihm bearbeitete Dokument einstellen möchte. Es handelt sich hier um die physischen Verzeichnisse der Server, die die Dokumente des Dokumentbestandes enthalten. Im rechten Teil des Fensters werden die lokalen Verzeichnisse des Autorenrechners angezeigt. Der Autor sucht hier nach dem Dokument, welches in den Dokumentbestand eingestellt werden soll. Dieses Dokument wird dann auf das ausgewählte Verzeichnis eines Servers kopiert. Weiterhin kann der Autor sowohl auf dem Server als auch auf seinem lokalen Rechner neue Verzeichnisse erstellen oder Dateien löschen, wenn er die notwendigen Zugriffsrechte besitzt.

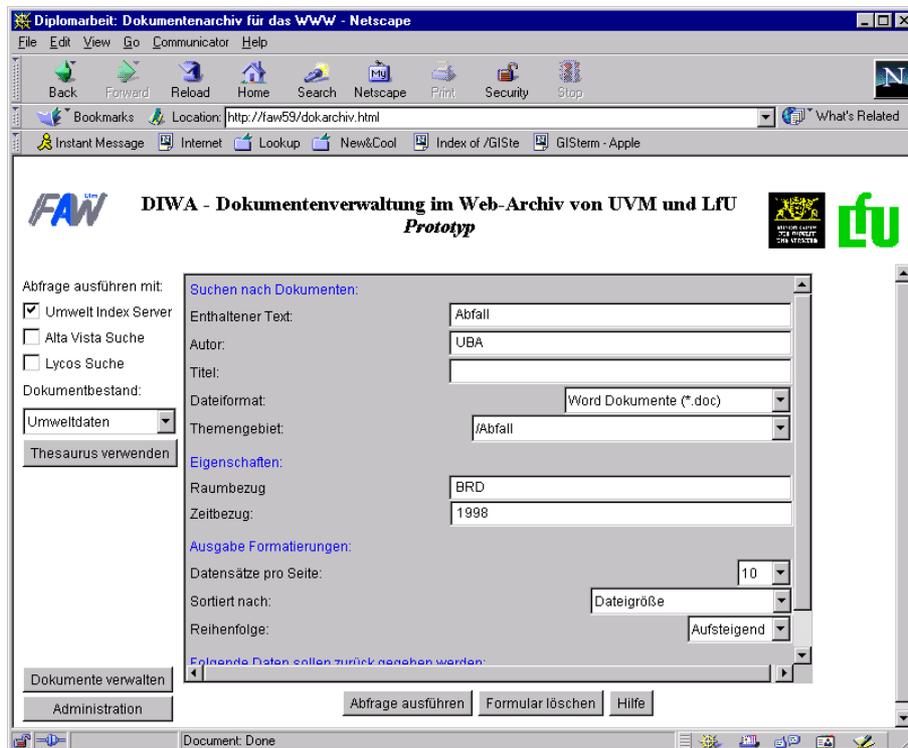


Abbildung 2: Rechercheoberfläche des Prototypen

Dabei kann der Nutzer neben einer Volltextsuche mit logischen Verknüpfungen auch nach Dateieigenschaften (Metainformationen) suchen.

Zur Begriffsfindung für die Volltextsuche kann der Thesaurus verwendet werden. Die in ihm selektierten Begriffe werden direkt in die Volltextsuche übernommen. Es kann weiterhin das Dateiformat der zu suchenden Dokumente sowie ein Themenbereich festgelegt werden. Ebenfalls im Suchformular wird das Format der Ausgabe der Ergebnisseite eingestellt. Nach dem Start der Anfrage wird ein neues Fenster für die Ergebnisanzeige geöffnet. Über die angezeigte URL können nun die Dokumente zur Ansicht oder weiteren Verarbeitung auf den lokalen Rechner heruntergeladen werden.

Der zugrunde liegende Datenbestand, welcher von diesem System derzeit verwaltet wird und recherchiert werden kann, besteht aus über 1000 Berichtsdocumenten, die in dem vom Umweltbundesamt erstellten Umweltbericht „Daten zur Umwelt 1997“ Verwendung finden. Die Datenbank wird bis zum Jahr 2000

ein einzigartiges Archiv an Informationen über den Zustand der Umwelt in Deutschland enthalten, welche zur Umweltberichterstattung herangezogen werden.

Um das System für Nicht-Experten einfach nutzbar zu machen, wurde eine überschaubare Anzahl an Feldern für die inhaltliche Erschließung des Datenbestandes über Metadaten ausgewählt. Dieser Sachverhalt ermöglicht es, Benutzern nur die für ihre Aufgabe erforderlichen Metadaten zu erheben, so dass der gesamte Dokumentenbestand auf ihre Bedürfnisse zugeschnitten ist.

Dies ist ein wesentlicher Unterschied zu Systemen wie der Datenbank Grunddaten (DBG 1999) oder dem Umweltdatenkatalog (UDK 1999). Bei der Datenbank Grunddaten stehen bspw. etwa ca. 40 Felder für die inhaltliche Erschließung der eingestellten Umweltdokumente zur Verfügung. Viele davon sind jedoch nur für Experten in bestimmten Fachgebieten erforderlich. Erfahrungen zur Befüllung der Datenbank Grunddaten werden in diesem Tagungsband von (Dombeck et al. 1999) vorgestellt.

3. Technische Realisierung

Abbildung 3: stellt die Systemarchitektur sowie die Kommunikationswege zwischen den in Abschnitt 2 beschriebenen Systemenkomponenten dar. Um möglichst viel Komplexität nutzerfreundlich darzustellen, wurde für die Benutzungsschnittstelle ein Java Applet entwickelt. Dieses Java Applet verbirgt z.B. die Nutzung von FTP hinter intuitiv bedienbaren Funktionskomponenten. Die Kommunikation zwischen dem Java Applet und den beiden Datenbankservern erfolgt über Java Database Connectivity (JDBC) und Remote Method Invocation (RMI). RMI wird verwendet, da momentan für Microsoft Access Datenbanken noch keine JDBC-Treiber existieren, welche den Zugriff auf die Access-Datenbanken mit JDBC direkt über das Netz ermöglichen. Um mit JDBC auf Access Datenbanken zuzugreifen, benötigt man einen speziellen JDBC-Treiber

(JDBC-ODBC-Bridge) von SUN (JDBC 1999), welcher JDBC-Befehle in ODBC-Befehle umwandelt und dann an den lokal installierten ODBC-Treiber für Access-Datenbanken weiterleitet. Ein RMI Server nimmt aus diesem Grund die Datenbankabfragen der Clients über das Netz entgegen, wandelt sie in JDBC-Befehle um und leitet diese über die JDBC-ODBC-Bridge an die lokale Datenbank weiter. Die Ergebnisse werden dann über RMI an die Clients zurück geschickt.

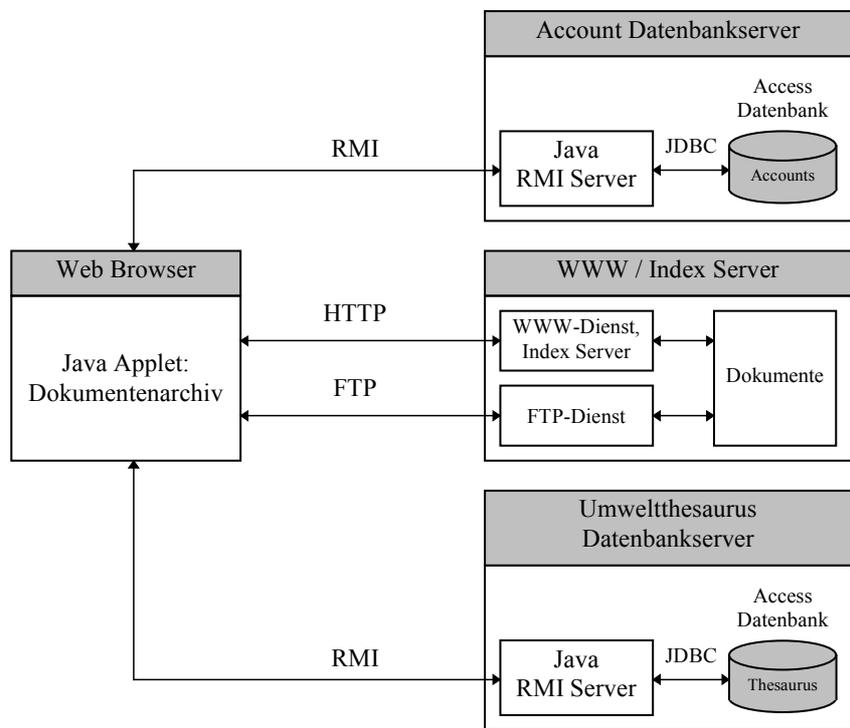


Abbildung 3: Systemarchitektur

3.1 Möglichkeiten zur technische Anbindung der Datenbanken

3.1.1 Java Database Connectivity (JDBC)

Der Begriff JDBC (Java Database Connectivity) bezeichnet ein von Sun entwickeltes Paket, das auf den Java Basisklassen aufsetzt und über die Bereitstellung von relationalen Datenbankobjekten sowie der entsprechenden Methoden den Zugriff aus Java Applets und Anwendungen auf beliebige Datenbanken ermöglicht. JDBC setzt auf dem X/OPEN SQL-Call-Level-Interface (CLI) auf und besitzt damit die gleiche Basis wie die ODBC-Schnittstelle.

Die API-Spezifikation für JDBC wurde von JavaSoft im März 1996 als vorläufige Version 0.50 vorgestellt. Sie liegt nunmehr in der Version 2.0 vor und wird im Rahmen des JDK 1.2 (Sun's Java Development Kit) ausgeliefert.

3.1.1.1 Einsatz von JDBC

Bei clientseitigem Zugriff auf eine Datenbank wird mit Java Applets unter Verwendung von JDBC eine direkte Verbindung zwischen dem Klienten und dem Datenbankmanagementsystem hergestellt. Die Verbindung erfolgt auf der Grundlage der von Sunsoft entwickelten JDBC-API (Java-Database-Connectivity), welche Java-Anwendungen und Java Applets den direkten Zugriff auf Datenbanksysteme ermöglicht.

Die Vorteile dieser Lösung liegen auf der Hand: Der Web-Server wird von intensiven Datenbankabfragen entlastet, die Programmierung über CGI oder eine spezifische API entfällt, und die Portabilität ist durch die Verwendung der Java-Schnittstelle JDBC gewährleistet.

3.1.1.2 Architektur von JDBC-Anwendungen

Ein Java Applet, das mit einem Browser vom Server auf den Klienten geladen wird, kann aufgrund der strengen Sicherheitsvorschriften für Applets nur mit

dem Rechner in Verbindung treten, von dem es geladen wurde. Wenn das Applet direkt mit dem Datenbankserver kommunizieren soll, ist es daher notwendig, dass Datenbank und Web-Server auf der gleichen Maschine installiert werden. Man spricht in diesem Fall von einer "Two-Tier-Architektur" (vgl. Abbildung 4), die im wesentlichen der klassischen Client-Server-Architektur entspricht.

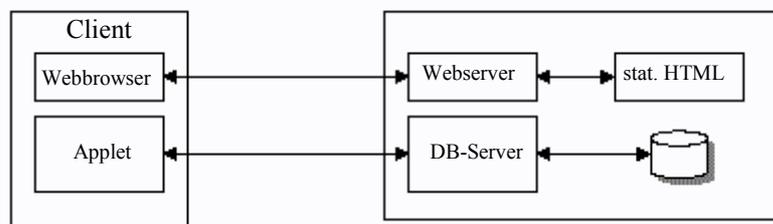


Abbildung 4: Two-Tier-Architektur

Der Klient lädt ein Java Applet vom Server, welches sich über ein Datenbankmanagementsystem (DBMS) spezifisches Protokoll (hier JDBC) mit der Datenbank verbindet, die auf der gleichen Maschine liegt wie der Web-Server.

Eine andere Möglichkeit, bei welcher der Web-Server und der Datenbankserver auf getrennten Maschinen installiert werden können, ist in Abbildung 5 dargestellt. Bei dieser sogenannten "Three-Tier-Architektur" wird ein Applet vom Web-Server geladen, das sich über ein DBMS-unabhängiges Protokoll mit einem Gateway (auf dem Web-Server) verbindet, welches dann als Datenbank-Klient die Abfrage für das Applet übernimmt und die Ergebnisse anschließend an das Applet zurücksendet (ein Beispiel für eine solche Three-Tier-Architektur ist die Verbindung eines Java Applets mit einer Java Anwendung über die Remote Method Invocation (RMI)).

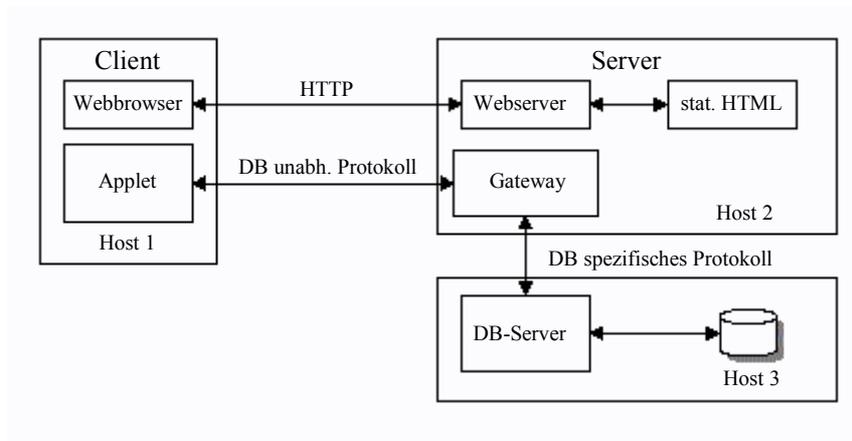


Abbildung 5: Three-Tier-Architektur

3.1.1.3 JDBC-Zugriffe über RMI

Eine Möglichkeit, die Three-Tier Architektur umzusetzen, besteht in der Verwendung der Remote Method Invocation (RMI). RMI ermöglicht beliebigen Java Applets und Anwendungen untereinander zu kommunizieren, um gegenseitig Funktionalitäten, Objekte und Methoden der einzelnen Java Anwendungen zu nutzen. Mit RMI ist es also auch möglich, ein komplexes System auf mehrere kleinere Systeme zu verteilen. Diese Möglichkeit existiert ab der Version 1.1 des JDK in Form des java.rmi-Packages.

Dieses Package stellt Java-Klassen mit den entsprechenden Methoden bereit, um vom Client-Applet aus mit Serverobjekten zu kommunizieren und Daten auszutauschen. In einer solchen Form kann die Three-Tier-Architektur folgendermaßen implementiert werden:

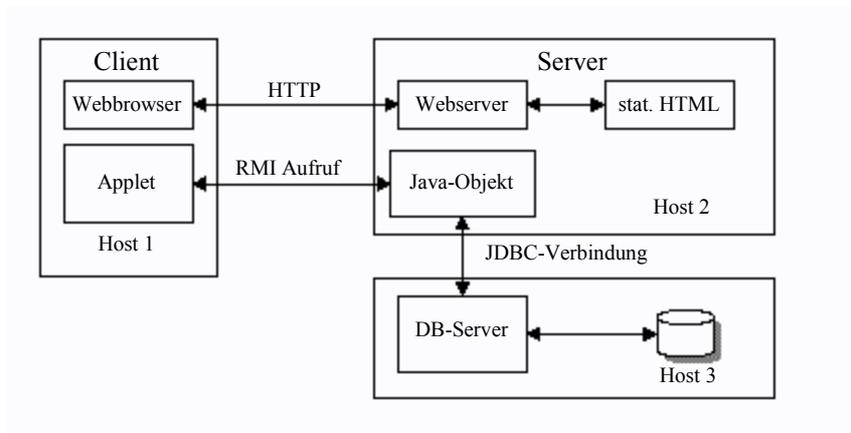


Abbildung 6: Three-Tier-Architektur mit RMI und JDBC

Der Klient lädt ein Java Applet vom Server, welches sich indirekt über RMI mit einem Serverobjekt verbindet, das seinerseits über JDBC die Verbindung mit einem ausgelagerten Datenbankserver herstellt.

3.1.2 Remote Method Invocation (RMI)

Das Package `java.rmi` stellt die Klassen und Methoden zur Realisierung einer verteilten Systemarchitektur mit Java bereit.

Allgemein geschieht das in der Form, dass ein Serverobjekt definiert wird, das die Methoden bereithält, die von den Klienten aufgerufen werden können. Zu diesem Serverobjekt muß ein Interface bereitgestellt werden, das der Klient vom Server lädt und das ihm die Schnittstelle zu den Serverkomponenten bereitstellt. Der Server selbst wird einmalig gestartet und wartet dann auf Anfragen der Klienten nach Ausführung seiner Methoden. Dies geschieht in der Form, dass der Klient das Serverobjekt ermittelt, und die Methode genauso aufruft, als wäre es ein lokales Objekt.

Die Weitergabe der Aufrufe zwischen dem Server und dem Klient geschieht über sogenannte Stubs und Skeletons, die stellvertretend für die jeweilige Gegenseite mit der aufrufenden Komponente kommunizieren.

RMI stellt somit eine geeignete Methode zur Realisierung verteilter Systemarchitekturen in einem Netzwerk zur Verfügung, sei es, um einem Applet den Zugriff auf externe Server (hier auf einen Datenbankserver) zu ermöglichen, oder aber spezielle Anwendungen (zum Beispiel umfangreiche Berechnungen) von einem eigens dafür eingerichteten Rechner durchführen zu lassen.

3.1.3 Bewertung der Architekturen

Aufwand

Die Two-Tier-Architektur verlangt weniger Aufwand bei der Implementierung und Wartung. Der gesamte Code wird auf dem Klient ausgeführt, es werden keine zusätzlichen Programme benötigt.

Performance

Ein Two-Tier-Treiber kann häufig nicht die Möglichkeiten des Multithreading unter Java und dessen Vorteile ausnutzen, wenn der Klient nicht selbst multithreadfähig implementiert wurde. Threads sind für eine Steigerung der Performance sehr wichtig, da dadurch mehrere Datenbankzugriffe annähernd gleichzeitig und mit einer sicher eingehaltenen Reihenfolge ausgeführt werden können.

Die Two-Tier-Architektur erfordert eine Realisierung der JDBC Implementation explizit für jedes Applet. Dieser zusätzliche Code muß immer an den Klienten übermittelt werden und verursacht zusätzlichen Verbrauch von Systemressourcen und Overhead zu Lasten des Klienten.

Durch die Verwendung der Three-Tier-Architektur verringert sich die Komplexität des Client-Applets. Das Applet empfängt lediglich Daten und stellt diese auf der Oberfläche dar.

Realisierung

Die enge Bindung der JDBC Implementierung an die Benutzungsoberfläche bei der Two-Tier-Architektur birgt die Gefahr, dass diese beiden Komponenten oft nicht mehr sauber voneinander getrennt implementiert werden; die Wiederverwendbarkeit von einzelnen Modulen und Komponenten wird eingeschränkt.

Einen weiteren Nachteil der Two-Tier-Architektur stellen einige JDBC-Treiber dar, die nativen, plattformabhängigen Code beinhalten, welcher von den Web Browsern aus Sicherheitseinschränkungen nicht ausgeführt werden kann.

Sicherheit

Bei der Two-Tier-Architektur müssen das Applet und die Datenbank auf demselben Rechner liegen, da das Applet aus Sicherheitsgründen nur auf die Maschine zugreifen darf, von der es geladen wurde. Alle Klienten, welche dieses Applet laden und ausführen, greifen direkt auf den Datenbankserver zu, wozu jeder Client ein verschlüsseltes Paßwort und einen verschlüsselten Benutzernamen benötigt. Dies erschwert die Konzeption eines Sicherheitskonzepts.

In einer Three-Tier-Architektur werden alle Informationen des Datenbankservers innerhalb der Mittelschicht (Gateway) von dem Applet versteckt gehalten. Dieses Gateway bildet eine Art Schutzschild für den direkten Zugriff von Klienten auf das Datenbankmanagementsystem.

4. Bewertung und Ausblick

Für kleine Nutzergruppen, die in einem überschaubaren Maß als Autoren mit einem Internet-basierten Informationssystem arbeiten, ist die Two-Tier-Architektur ausreichend. So kann eine reduzierte Performance akzeptiert werden, da nicht mit allzu vielen gleichzeitigen Datenbankzugriffen gerechnet werden muß. Auch die explizite und ressourcenaufwendige Realisierung einer JDBC-Verbindung für jedes Applet ist vor diesem Hintergrund akzeptabel. Bei

kleinen Nutzergruppen kann zudem überschaut werden, in welcher Systemumgebung welche Browser auf dem Klienten installiert sind. Gegebenenfalls können diese Browser bzw. die darauf laufenden Applets so eingestellt werden, dass der z.T. erforderliche plattformabhängige Code keinen echten Nachteil der Two-Tier-Architektur darstellt. Schließlich müssen bei kleinen Nutzergruppen keine so hohen Sicherheitsanforderungen angesetzt werden, wie dies bei Systemen mit einer großen Nutzerzahl der Fall ist. Die Tatsache, dass die Two-Tier-Architektur für jeden Klienten ein Paßwort und Benutzernamen benötigt, stellt also keine echte Sicherheitslücke dar. Sollen jedoch Systeme auch von einer großen Zahl von Nutzern verwendet werden, bietet die Three-Tier-Architektur entscheidende Vorteile hinsichtlich Performance und Sicherheit.

Um Internet-basierte Informationssysteme für verschiedene Nutzergruppen effektiv nutzbar zu machen, ist es erforderlich, dass für ein und denselben Datenbestand nutzerspezifisch Metadaten verfügbar sind. Es ist geplant, diesen Aspekt in zukünftige Weiterentwicklungen des Systems zu integrieren. Ein weiterer wichtiger Aspekt ist die Möglichkeit zur Verwaltung mehrerer unabhängiger, nutzerspezifischer Dokumentbestände sowie der zugehörigen Metadaten auf einem Index Server. Das in diesem Beitrag beschriebene System kann nur für kleine Nutzergruppen effektiv eingesetzt werden. Ein Grund hierfür ist, dass der verfügbare Dokumentbestand in keiner Form strukturiert ist. Strukturierungen der Dokumente, die sich z.B. an organisatorischen Einheiten wie Abteilungen orientieren, sind jedoch notwendige Voraussetzung, um große Dokumentbestände aufzubauen, insbesondere wenn diese von zahlreichen verschiedenen Nutzern erweitert werden.

Derzeit sind nur Microsoft Office Dokumente und HTML-Dokumente recherchierbar. Für die Weiterentwicklung sind sowohl XML als auch PDF als Dokumentformate zu berücksichtigen.

Eine weitere technische Herausforderung besteht in der Integration von Dokumentbeständen über technische Systemgrenzen hinaus. Diese Aspekte werden

im Rahmen von dem vom Ministerium für Umwelt und Verkehr Baden-Württemberg initiierten Projekt DIWA (Dokumentenverwaltung im Web-Archiv von UVM und LfU) (Henning et al. 1999) weiter verfolgt und umgesetzt.

Literatur

AltaVista 1999: <http://altavista.afterfive.com/forum.htm>

BSCW 1999: <http://bscw.gmd.de/index.html>

Computer 1999: IEEE Computer, Special Issue „Digital Libraries“
Vol. 32, No. 2, 1999.

Dombeck, T., Tochtermann, K. 1999: Praktische Erfahrungen bei der Datener-schließung für eine Umweltdatenbank des Umweltbundesam-tes. Veröffentlicht in diesem Tagungsband.

DGB, 1999: Dokumentation der Datenbank Grunddaten Version 3.1. Dr. Lippke und Dr. Wagner GmbH, Berlin.

Hartwik, J., Einführung in JDBC: <http://home.fhtw-berlin.de/~s0109297/jdbc/jdbc.html>.

JDBC 1999: SUN, The JDBC(tm) Universal Data Access API:
<http://www.javasoft.com/products/jdbc/index.html>.

Hicks, D., Tochtermann, K., Kussmaul, A., Neils, S. 1999: Customization in Environmental Information Systems, *Tagungsband 13. Symposium Umweltinformatik*, Magdeburg, Metropolis-Verlag Marburg, 1999.

Schatz, B., Hsinchun, Ch. 1999: Digital Libraries: Technological Advances and Social Impacts; in: IEEE Computer 1999, p. 45-50.

Schmidt, A., Remote Method Invocation

<http://www.tzi.org/~ansu/papers/rmi/PowerPoint/index.htm>.

SUN, Java Remote Method Invocation: <http://www.javasoft.com/products/jdk/rmi/index.html>.

UDK, 1999: http://www.uvm.baden-wuerttemberg.de/www-udk/RPT_DATA/about_d.html.

Seggelke, J. und Mohaupt-Jahr, B.: "Der Verweis- und Kommunikationsservice des Umweltbundesamts – Ein Modellfall für das Umwelt-Intranet". In: W. Geiger, A. Jaeschke, D. Rentz, E. Simon, T. Spengler, L. Zilliox und T. Zundel (Hrsg.): Umweltinformatik '97. 11. Internationales Symposium, Straßburg, September 1997, Tagungsband. Metropolis-Verlag, Marburg, 1997.

Internetbasierte Dokumentenverwaltung heterogener Umweltdokumentenbestände

Inge Henning², Renate Ebel, Martina Tauber³, Klaus Tochtermann, Katrin Pursche, Andreas Kussmaul, Albrecht Schultze⁴

Abstract

Internet-based document management systems are often used to make available a large number of documents to a broad user community. This also applies to public administrations where environmental documents are often to be exchanged between different departments or within the same department. However, the documents to be handled in administrative settings are often large in size, different in format and heterogeneous in content and origin. In addition, public administrations often have to support heterogeneous technical infrastructures. Against this background commercial, off-the-shelf document management systems cannot meet the high requirements to manage effectively environmental documents in administrative settings. This paper presents the DIWA system, an Internet-based environmental document management system which is designed to meet the various different requirements modern public administrations have.

¹ I. Henning; Ministerium für Umwelt und Verkehr Baden-Württemberg, Postfach103439, D-70029 Stuttgart, E-Mail: Inge.Henning@uvm.bwl.de.

² R. Ebel, M. Tauber; Landesanstalt für Umweltschutz Baden-Württemberg, Postfach 210752, D-76157 Karlsruhe, E-Mail: Renate.Ebel | Martina.Tauber@x400.lfuka.um.bwl.de.

³ K. Tochtermann, K. Pursche, A. Kussmaul, A. Schultze; Forschungsinstitut für anwendungsorientierte Wissensverarbeitung, Postfach 2060, D-89081 Ulm, E-Mail: tochterm | pursche | kussmaul | schultze@faw.uni-ulm.de.

1 Einführung

Gegenüber herkömmlichen Dokumentenverwaltungssystemen ermöglichen internetbasierte Dokumentenverwaltungssysteme die einfache Verfügbarmachung der zu verwaltenden Dokumente mit Hilfe moderner Internettechnologien bzw. setzen diese Technologien zur effizienten Dokumentenverwaltung ein. Wesentliche Charakteristika bekannter internetbasierter Dokumentenverwaltungssysteme, wie das im Digitalen Umweltatlas Hamburg verwendete System Lotus Notes (Blasius et al. 1997) oder im weitesten Sinne das System Hyper-Wave (Maurer 1996), bestehen darin, dass die Systeme in einer homogenen Systemumgebung eingesetzt werden und dass die zu verwaltenden Dokumente relativ homogen bezüglich ihres Formats, ihrer Größe, ihres Inhaltes und ihrer Herkunft sind. Umweltdokumente können sehr heterogen bezüglich dieser Eigenschaften sein. Hinzu kommt, dass in öffentlichen Verwaltungen sehr viele Stellen mit der Erstellung und Bereitstellung von Umweltdokumenten befaßt sind und häufig sehr heterogene technische Infrastrukturen anzutreffen sind. Vor diesem Hintergrund sind spezielle Anforderungen an internetbasierte Dokumentenverwaltungssysteme für Umweltdokumente zu stellen, die in der Regel nicht von herkömmlichen Systemen abgedeckt werden. Aus diesem Grund hat das Ministerium für Umwelt und Verkehr Baden-Württemberg (UVM) im Rahmen des Forschungs- und Entwicklungsvorhabens GLOBUS (Mayer-Föll et al. 1998) das Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW Ulm) mit der Entwicklung eines internetbasierten Dokumentenverwaltungssystems für heterogene Umweltdokumentenbestände beauftragt. Um bereits in der Konzeption möglichst viele unterschiedliche Anforderungen zu berücksichtigen, werden entsprechende Fachstellen sowohl aus dem UVM als auch aus der Landesanstalt für Umweltschutz Baden-Württemberg (LfU) frühzeitig in das Projekt DIWA (Dokumentenverwaltung im Web-Archiv von UVM und LfU) einbezogen.

2 Ziele und Anforderungen im Projekt DIWA

Die übergreifenden Ziele des Projektes sind der Aufbau, die Pflege und die Nutzung eines Web-Archivs für multimediale Umweltdokumentenbestände. Dabei sollen Anwender die ihnen vertraute Bürokommunikationsumgebung zur Erstellung von Umweltdokumenten verwenden können. Da die Dokumente hauptsächlich in den Dateiformaten Microsoft Office, HTML und PDF vorliegen, sind diese Formate im Web-Archiv zu unterstützen. Für die Zukunft sind auch Formate wie XML zu berücksichtigen.

Anwender sollen aus ihrer Bürokommunikationsumgebung heraus die von ihnen erstellten Umweltdokumente anderen Anwendern bzw. anderen Fachstellen verfügbar machen können, ohne dabei auf spezialisierte Berichtssysteme, wie das HUDA-Redaktionssystem (Tochtermann et al. 1998) oder das Alfa-Web-Autorensystem (Lautner et al. 1998), angewiesen zu sein. Durch die Verwendung eines einfach zu bedienenden, über den Web-Browser erreichbaren Werkzeugs sollen die notwendige Einarbeitungszeit minimiert und der Aufwand für das Installieren solcher Systeme auf dem lokalen Rechner vermieden werden.

Umweltdokumente, die nicht für Nachweissysteme, wie den Umweltdatenkatalog (UDK 1998), erschlossen und recherchierbar gemacht werden, sollen entweder durch die Anwender selbst oder durch entsprechende Fachstellen über das Internet oder über ein behördeninternes Intranet in einer thematischen Systematik zugänglich gemacht werden können. Für diesen Zweck sollen bei UVM und LfU bereits vorhandene Dokumentenablagestrukturen auf den entsprechenden Internetservern verfügbar gemacht und transparenter dargestellt werden. Derartige Ablagestrukturen lehnen sich an verwaltungsinterne, organisatorische Einheiten wie Arbeitsgruppen, Referate oder Abteilungen an.

Umweltdokumente aus bereits vorhandenen Umweltdokumentenbeständen sollen neben dem Web-Archiv recherchierbar sein, ohne dass Anwender ein

spezielles Werkzeug auf ihrem Rechner installiert haben und nutzen müssen (z.B. soll der in HUDA aufgebaute Dokumentenbestand recherchierbar und zugreifbar sein, ohne das entsprechende Datenverwaltungssystem verwenden zu müssen). Die Recherche nach Dokumenten soll nicht nur über eine Volltextsuche sondern auch über Dokumenteigenschaften möglich sein.

Eine Übersicht dieser zu erreichenden Dokumentenbestände im Intranet und im Internet wird in Abbildung 1 dargestellt.

Ziel ist es, im Rahmen des Projektes möglichst schnell einen ersten Prototypen in den praktischen Betrieb zu bringen. Im Betrieb des System sollen dann weitere von den Anwendern erkannte Anforderungen erhoben, priorisiert und anschließend umgesetzt werden.

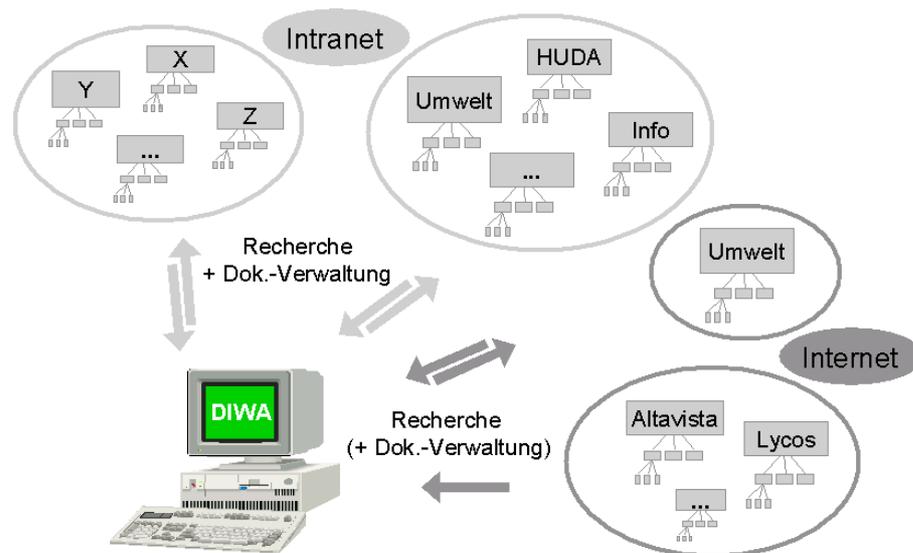


Abbildung 1: Erreichbare Dokumentenbestände in DIWA

3 Abgrenzung zu Redaktionssystemen

Der wesentliche Unterschied zu Redaktionssystemen wie HUDA (Tochtermann et al. 1998) und AlfaWeb (Lautner et al. 1998) ist, dass das geplante System

kein Redaktionssystem zur Erstellung von durchgängigen, in sich geschlossenen Berichten ist. Vielmehr wird aus der Standard-Bürokommunikationsumgebung heraus ein multimediales und heterogenes Umweltdokumentenarchiv aufgebaut und gepflegt, das die Wiederverwendbarkeit und damit die Mehrfachnutzung von Arbeitsergebnissen in Umweltverwaltungen ermöglicht. Diese Arbeitsergebnisse können zum einen grundsätzlich an jedem Arbeitsplatz erstellt und in das System eingestellt werden; zum anderen können sie prinzipiell an jedem Arbeitsplatz in einem Intranet recherchiert und abgerufen werden, sofern die erforderlichen Berechtigungen vorhanden sind. Vor diesem Hintergrund können die Arbeitsproduktivität erheblich erhöht und der Betreuungsaufwand trotz wachsender Dokumentbestände begrenzt werden, weshalb ein derartiges Projekt auch aus wirtschaftlicher Sicht von großer Bedeutung ist.

4 Realisierung

Der erste entwickelte Systemprototyp basiert auf einem Index Server. Derartige Systeme sind stets an einen Web-Server gekoppelt, der die Dokumente verwaltet. Ein Index Server bereitet die auf dem zugehörigen Web-Server verwalteten Dokumente für eine Recherche auf, etwa in Form eines Index zur Volltextsuche. Der vorliegende Prototyp setzt auf dem Microsoft (MS) Index Server (MS IS 1999) auf, da dieser Index Server neben dem Dokumentinhalt auch Dokumenteigenschaften von Microsoft Office-, HTML- und ASCII-Dateien indizieren kann. Für das PDF-Format wurde ein Filter von Adobe in den Index Server integriert (Adobe 1999). Der MS Index Server ist in der Lage, die physischen Dokumentablagen inkrementell zu indizieren, d.h. er sucht in einem Indizierungszyklus nach neu eingestellten Dokumenten und indiziert nur diese. Dadurch verkürzt sich die Dauer eines solchen Zyklus.

Der MS Index Server setzt auf dem MS Internet Information Server (MS IIS 1999) als Web-Server auf. Im MS Internet Information Server können virtuelle Verzeichnisse angelegt werden, die logisch mit physischen Verzeichnissen verknüpft werden. Diese physischen Verzeichnisse können sich auf Servern verschiedener Systemumgebungen (Unix, Windows) befinden. Voraussetzung ist, dass die Verzeichnisse über einen Pfadnamen nach der Uniformed Naming Convention (UNC) angesprochen werden können. Da diese Bedingung nicht von allen Betriebssystemen erfüllt wird, ist das Verfügbarmachen von Dokumentenbeständen, die sich auf solchen Plattformen befinden, mit zusätzlichem Aufwand verbunden.

Ein Vorteil bei der Verwendung dieser Software besteht darin, dass der MS Index Server sowie der von ihm verwendete Web-Server MS Internet Information Server kostenlose Zusatzmodule der Windows NT Server-Software sind und das zu entwickelnde System auf dieser Plattform lauffähig sein soll.

Der auf der Standard-Software aufsetzende, Java-basierte Prototyp setzt sich aus den Modulen Administrator-, Autoren- und Recherchekomponente zusammen, die in Abhängigkeit von den Rechten und Anforderungen der Anwender kombiniert werden können. Die zentralen Eigenschaften der einzelnen Komponenten werden im folgenden beschrieben.

4.1 Administratorkomponente

In diesem Modul können die inkrementelle bzw. volle Indizierung sowie die Optimierung des Index manuell gestartet werden. Dazu werden im ersten Prototypen nur die Grundfunktionen des MS Index Server ohne Zusatzfunktionalität genutzt.

Weitere administrative Aufgaben werden in den jeweiligen Software-Standardoberflächen durchgeführt. Da der MS Internet Information Server direkt auf die Zugriffsrechte unter Windows NT zugreift, findet die Vergabe von

Nutzerrechten für die Verzeichnisse des Web-Archivs über den MS Explorer statt. Die Pflege der Struktur des Web-Archivs erfolgt über den Internet Service Manager des MS Internet Information Servers.

4.2 Autorenkomponente

Über die Autorenkomponente können Anwender – in der Rolle eines Autors – die von ihnen erstellten Umweltdokumente in den vom Index Server verwalteten Dokumentenbestand einstellen und somit für andere Anwender recherchierbar und nutzbar machen. Abbildung 2 zeigt die Oberfläche der Autorenkomponente. Die Autorenkomponente des Prototypen verfügt über folgende Funktionalitäten:

1. Autoren können ihre in der gewohnten Bürokommunikations-Umgebung erstellten Dokumente über eine FTP-Verbindung auf den Web-Server einstellen und somit dem Index Server bekanntgeben. Des weiteren können je nach Nutzerrechten Verzeichnisse gelöscht sowie Dokumente heruntergeladen oder gelöscht werden. Die von Anwendern häufig als relativ kompliziert empfundene Verwendung von FTP-Diensten wird durch ein intuitiv zu bedienendes Java Applet verborgen. Da beim Einstellen von Dokumenten auf einen Server auf die lokalen Rechnerressourcen des Autorenrechners zugegriffen wird, muß das Java Applet der Autorenkomponente signiert werden (Pursche et al. 1999).
2. Die Autorenkomponente basiert auf Java und kann über einen Standard-Web-Browser gestartet werden. Dadurch erübrigt sich die lokale Installation der Autorenkomponente, und der Arbeitsaufwand verringert sich bei einem großen Anwenderkreis erheblich.
3. Die Autorenkomponente ist für Anwender nur verfügbar, wenn sie über entsprechende Nutzerrechte auf dem Index Server verfügen. Damit wird ver-

mieden, dass beliebige Anwender die Umweltdokumentenbestände verändern können.

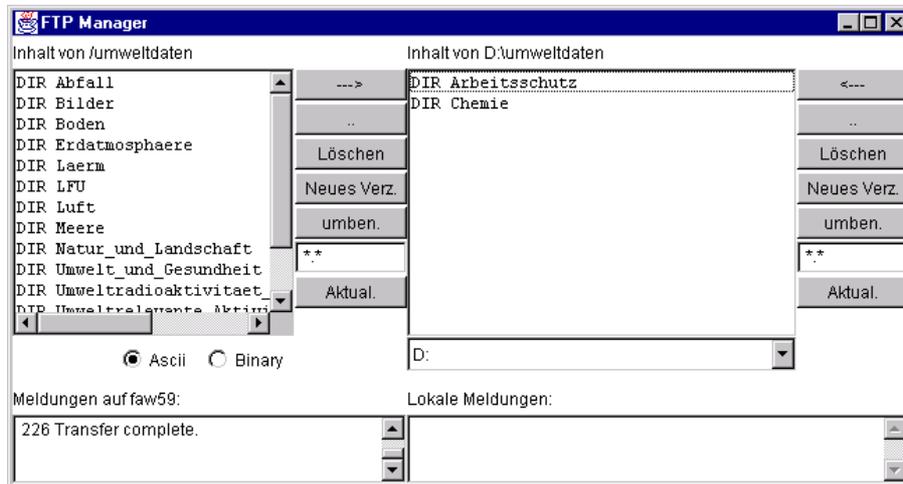


Abbildung 2: Oberfläche der Autorenkomponente des DIWA-Prototypen

4.3 Recherchekomponente

Für Anwender des Systems soll die Recherche auf dem Dokumentenbestand über gängige Web-Browser und nicht über Oberflächen spezieller Werkzeuge ermöglicht werden. Dennoch sollen weitestgehend auch Dokumente, die in anderen Systemen bereits existieren (z.B. mit dem HUDA-Redaktionssystem erstellte Umweltberichte) über das System recherchierbar und zugreifbar gemacht werden. Über die Recherchekomponente (vgl. Abbildung 4) ist es Anwendern möglich:

1. Dokumente einfach auf dem Index Server zu recherchieren und in der eigenen Bürokommunikations-Umgebung einzusehen,
2. eine Volltextrecherche bzw. eine Recherche über Dokumenteigenschaften durchzuführen. Bei den Dokumenteigenschaften kann es sich sowohl um von der Anwendung vordefinierte Eigenschaften wie Autor oder Titel han-

deln als auch um benutzerdefinierte Informationen wie Zeit- oder Raumbegzug.

3. die Formulierung der Anfrage über einen vorgeschalteten Umweltthesaurus zu unterstützen. Damit ist es möglich, auch nach verwandten Begriffen, Ober-, Unterbegriffen etc. zu suchen. Diese Begriffe werden dann in einer hierarchischen Struktur dargestellt (vgl. Abbildung 3).
4. bei der Anfrage auszuwählen, ob sie nur in dem Web-Archiv von DIWA suchen möchten oder zusätzlich die Anfrage an weitere Umweltdokumentenbestände (z.B. HUDA) oder andere Suchmaschinen (z.B. AltaVista, Lycos) weiterleiten möchten (vgl. Abbildung 1). Dadurch kann die Suche je nach den gewünschten Ergebnissen erweitert oder eingeschränkt werden. Die Ergebnisse werden für jede Suchmaschine auf einer gesonderten Browser-Seite ausgegeben.
5. Um die thematische Strukturierung des Dokumentenbestandes einzusehen und damit die Recherche erleichtern und thematisch einschränken zu können, wird über ein signiertes Java Applet eine FTP-Verbindung mit dem Dokumentbestand des Index Servers hergestellt. Da der Aufbau von FTP-Verbindungen in die Sicherheitsbeschränkungen von Java fällt, muß auch die Recherchekomponente signiert werden.
6. unterschiedliche Formatierungen für die Ergebnisanzeige auszuwählen. Es können verschiedene Dateieigenschaften, eine Kurzfassung des Textes, die URL zum Dokument etc. angezeigt werden. Weiterhin kann die Sortierung der Ergebnisse nach verschiedenen Kriterien wie Trefferanzahl im Text, Titel, Dateigröße etc. festgelegt werden.

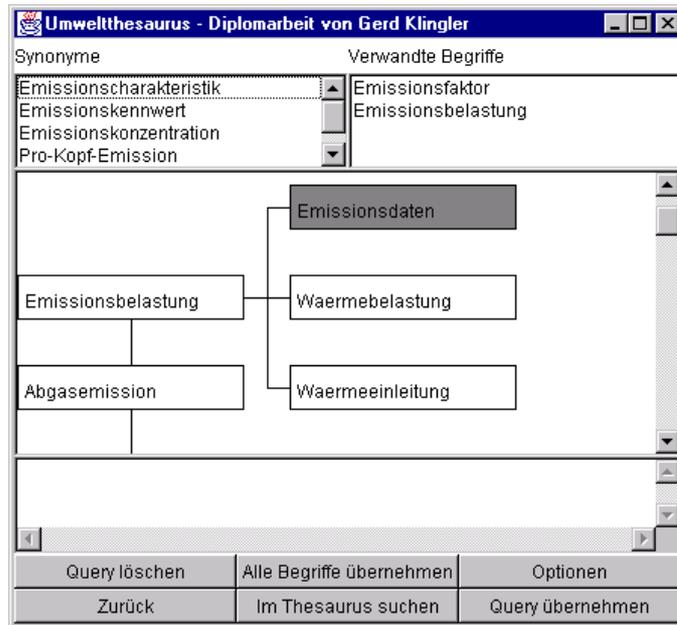


Abbildung 3: Oberfläche des in der Recherche verwendeten Umweltthesaurus

Abbildung 3 zeigt die Oberfläche des in der Recherche verwendeten Umweltthesaurus. Es werden die Unterbegriffe von „Emissionsbelastung“ sowie die Synonyme und verwandten Begriffe von „Emissionsdaten“ angezeigt.

Abbildung 4 zeigt die Rechercheoberfläche des DIWA-Prototypen. In der Abbildung wird eine Suche nach Dokumenten durchgeführt, die den Begriff Abfall beinhalten, vom Autor UBA erzeugt wurden, das MS Word-Format besitzen, im Themengebiet Abfall abgelegt sind, den Raumbezug BRD sowie den Zeitbezug 1998 als Metainformationen besitzen. Das grau hinterlegte Suchformular kann in Abhängigkeit von dem zu durchsuchenden Dokumentenbestandes verändert werden, d.h. es können jeweils dem Thema entsprechende Metadaten verwendet werden.

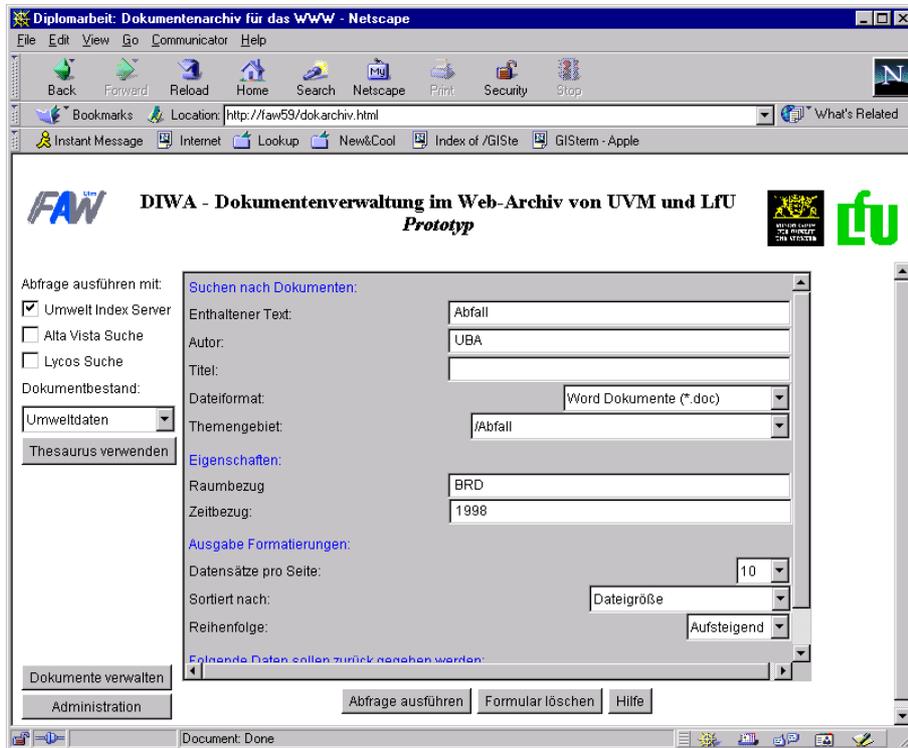


Abbildung 4: Oberfläche der Recherchekomponente des Prototypen

Die vom Anwender festgelegte Anfrage wird über eine MS Index Server-spezifische Internet-Datenabfragedatei (.idq) an den Index Server übergeben. Auf der Basis der angegebenen Kriterien wird im Index des Web-Archivs bzw. in weiteren Dokumentenbeständen gesucht. Die gefundenen Dokumente werden dann mit Hilfe einer speziellen HTML-Erweiterungs-Datei (.htx) und der vom Nutzer festgelegten Ausgabeformatierung dynamisch aufbereitet und in einer neuen Browser-Seite angezeigt.

Über die in der Ergebnisseite enthaltenen URL's können die Dokumente nun zur Weiterverarbeitung auf dem lokalen Rechner heruntergeladen werden. Erst in diesem Schritt wird physisch auf das Dokument zugegriffen.

Die Ergebnisliste enthält nur Dokumente, auf die der Nutzer Zugriffsrechte besitzt. Die Existenz von Dokumenten bleibt dem nicht berechtigten Nutzer also verborgen.

Der beschriebene Ablauf einer Recherche wird in Abbildung 5 dargestellt.

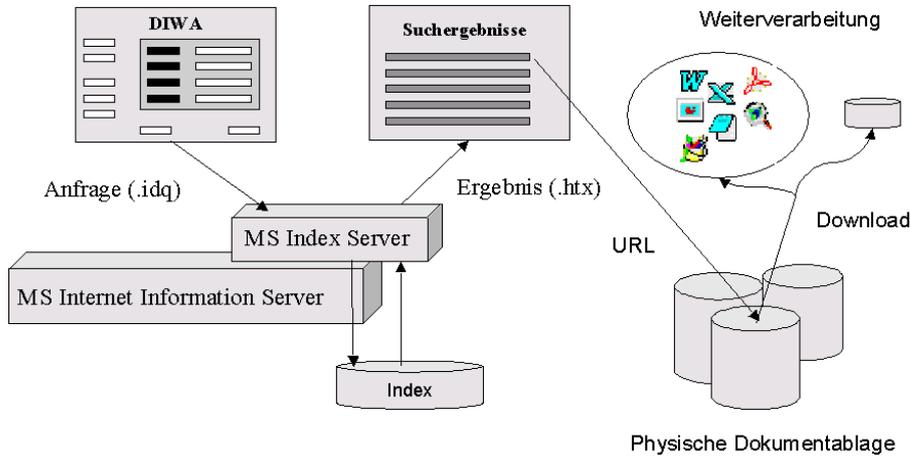


Abbildung 5: Ablauf der Recherche sowie der Weiterverarbeitung von Dokumenten

5 Anforderungen für die weitere Realisierung von DIWA

Der im vorigen Kapitel beschriebene Prototyp wurde im UVM installiert und steht über das Intranet auch der LfU zu Testzwecken zur Verfügung. Bereits jetzt wurden aufgrund erster Erfahrungen mit dem Prototypen im Projektteam neue Anforderungen erarbeitet. Zentrales Thema ist dabei der Aufbau einer virtuellen Ablagestruktur für das Web-Archiv. Ziel einer virtuellen Ablagestruktur ist die thematische Untergliederung bzw. Zusammenfassung von Dokumenten unabhängig von ihrem physischen Ablageplatz auf einem Server. In Verwaltungen werden Dokumente häufig entsprechend der organisatorischen Struktur, z.B. nach Abteilungen, abgelegt. Für eine Recherche sind diese Untergliederungen allerdings wenig sinnvoll. Hier bietet sich z.B. eine Zuordnung zu Umweltthemen an. Der MS Internet Information Server bietet die Möglichkeit, eine virtuelle Ablagestruktur aufzubauen und den virtuellen Verzeichnissen physische Verzeichnispfade zuzuordnen (Abbildung 6). In diesen Prozeß können alle

Server eines Intranets einbezogen werden, solange sie nicht außerhalb einer Fire Wall liegen. Dieses Vorgehen hat den Vorteil, dass für die Recherche eine sinnvolle Gliederung geschaffen wird, ohne die meist umfangreiche physische Dokumentenablage verändern zu müssen.

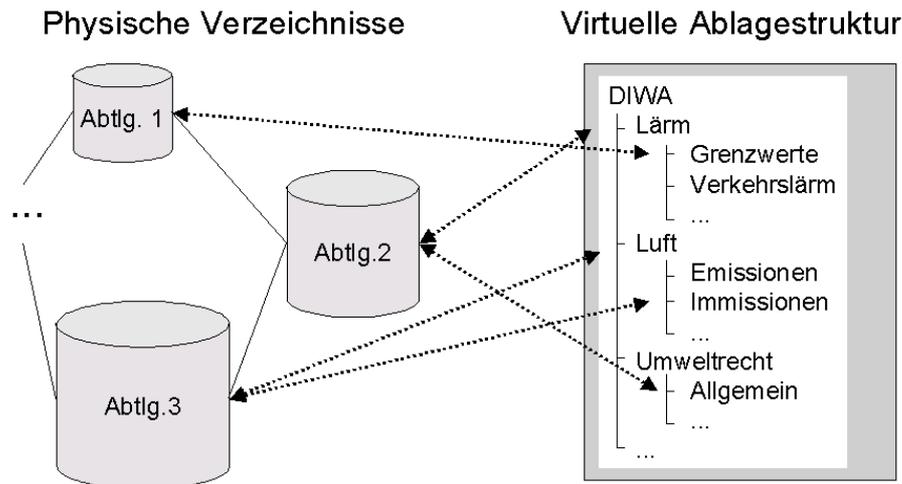


Abbildung 6: Zuordnung physischer Verzeichnisse zu einer virtuellen Ablagestruktur

Eine weitere Anforderung betrifft die Einbindung von Dokumentenbeständen auf Servern, die unter Unix betrieben werden. Der MS Internet Information Server sowie der MS Index Server können nur auf Verzeichnisse zugreifen, die über einen Pfadnamen nach der UNC verfügen. Auf Unix-Servern ist das nicht automatisch der Fall. Eine Lösung ist die Installation eines Fileservers auf den jeweiligen Servern, der die entsprechenden UNC-Pfadnamen erzeugen kann (z.B. Samba Server).

Weitere Anforderungen werden im folgenden, den einzelnen Komponenten zugeordnet, dargestellt.

5.1 Administratorkomponente

Die Administratorkomponente soll die Vergabe von Nutzerrechten für die einzelnen Komponenten von DIWA erlauben. Dafür sollen Nutzergruppen festgelegt werden, die Zugriffsrechte entweder nur auf die Recherchekomponente oder zusätzlich auf die Autorenkomponente erhalten. Neue Anwender werden dann einer dieser Gruppen zugeordnet. Diese Nutzergruppen müssen für das Einstellen von Dokumenten über die Autorenkomponente weiter unterteilt werden, da Autoren Dokumente nur in für sie relevante Verzeichnisse einstellen dürfen.

Alle weiteren administrativen Aufgaben werden über die Standardoberflächen der entsprechenden Software ausgeführt.

5.2 Autorenkomponente

Für das Einstellen von Dokumenten in das Web-Archiv sind verschiedene Ansätze zu untersuchen. Zum einen könnten Dokumente ganz normal auf einem Server gespeichert werden, d.h. der Autor nutzt die Standardmöglichkeiten der Bürokommunikations-Umgebung für das Einstellen von Dokumenten in ein physisches Verzeichnis, welches einem virtuellen Verzeichnis des Web-Archivs zugeordnet ist. Problematisch ist hier, dass der Autor Kenntnis über diese Zuordnung von Verzeichnissen im Web-Archiv besitzen muß. Er muß wissen, ob ein bestimmtes Verzeichnis auf seinem Server indiziert und somit im Web-Archiv recherchierbar ist oder nicht. Weiterhin benötigt der Autor die Information, welchem Thema das Verzeichnis in der virtuellen Ablagestruktur zugeordnet ist, um eine unkorrekte Zuordnung zu vermeiden.

Die zweite Variante besteht im Einstellen von Dokumenten über die virtuelle Ablagestruktur, d.h. der Autor ordnet über die Autorenkomponente von DIWA ein Dokument einem virtuellen Verzeichnis mit einem bestimmten Thema zu. Über die Zuordnungen der virtuellen zur physischen Ablagestruktur wird das

zugehörige physische Verzeichnis ermittelt und das Dokument dort gespeichert. Dies ist allerdings nur dann möglich, wenn der Autor die entsprechenden Zugriffsrechte für dieses physische Verzeichnis besitzt.

Für die endgültige Lösung dieses Problems werden die Testergebnisse potentieller Nutzer ausschlaggebend sein.

5.3 Recherchekomponente

Für die Recherchekomponente wurden konkrete Anforderungen auf verschiedenen Gebieten identifiziert. Ein Hauptthema ist dabei die Festlegung von Metainformationen als Suchkriterien. Diese Metainformationen werden z.B. in Office-Dokumenten als benutzerdefinierte Eigenschaften oder in HTML-Dokumenten als Meta-Tags festgelegt. Diese Eigenschaften müssen von den Autoren gefüllt werden, um sie für die Recherche verfügbar zu machen. Um eine hohe Akzeptanz bei den Autoren zu erreichen, muß sich der Umfang bzw. die Anzahl dieser Informationen in Grenzen halten. Andererseits sollen Dokumente so eindeutig wie möglich über Eigenschaften identifiziert werden können. Für die Lösung dieses Spagats werden Metainformationen anderer Systeme (z.B. Datenbank Grunddaten) und die Erfahrungen der zukünftigen Anwender genutzt. Um Autoren das Befüllen der Metainformationen zu erleichtern, werden in einem ersten Schritt Dokumentvorlagen für Office-Dokumente erstellt, in welchen die entsprechenden Dateieigenschaften vordefiniert sind.

Um die Recherche besser an die Bedürfnisse der Anwender anzupassen, wird auch daran gedacht, Nutzerprofile anzubieten. Somit könnten Anwender ein persönliches Interessenprofil anlegen, aufgrund dessen sie genau in dem Teil des Bestandes recherchieren können, der auch thematisch für sie von Interesse ist. Nutzerprofile könnten mit den unabhängigen Indizes harmonisiert werden.

Schließlich wurde die Anforderung formuliert, dass bei einer Recherche auch die Dokumente berücksichtigt werden, auf die ein Nutzer keinen Zugriff hat. In der Darstellung des Rechercheergebnisses würden dann derartige Dokumente zwar angezeigt werden, sie könnten aber, da der Nutzer keine Zugriffsrechte hat, von diesem nicht eingesehen werden. Allerdings bestünde die Möglichkeit, sich bei den Verantwortlichen Informationen über das Dokument und ggf. auch das Dokument selbst zu besorgen.

6 Ausblick

Aufgrund der verschiedenen zu berücksichtigenden Systemumgebungen zum einen und der heterogenen Umweltdokumentenbestände zum anderen ist derzeit die Erwartungshaltung, dass man am ehesten mit der Erstellung eines flexiblen Werkzeugkastens die derzeit bekannten anwenderspezifischen als auch die noch genauer zu definierenden technischen Anforderungen erfüllen kann. Je nach Bedürfnissen können einzelne Werkzeuge des Werkzeugkastens den verschiedenen Nutzergruppen bereitgestellt werden. Bei der Entwicklung der Werkzeuge wird in höchst möglichem Maße auf existierende Systemkomponenten zurückgegriffen. Die softwaretechnische Herausforderung wird daher in der Integration dieser Systemkomponenten bestehen. Dass dieser Ansatz sehr erfolgversprechend ist, wurde mit der Realisierung des HUDA-Werkzeugkastens eindrucksvoll unter Beweis gestellt.

Ein wesentliches Ziel für zukünftige Entwicklungen in DIWA ist die Integration von Komponenten zur sicheren Datenübertragung. Dieser Aspekt kommt vor allem dann zum Tragen, wenn Dokumente auch über das Internet externen Nutzern zur Verfügung gestellt werden sollen.

Mit DIWA ist es bei der Recherche über Metainformationen derzeit nur möglich, nach Dokumenten zu suchen, die einheitlich inhaltlich erschlossen wurden. Soll eine Suche nach Dokumenten ermöglicht werden, denen unterschiedliche Me-

tainformationen zugrundeliegen, so müßten Brokerkonzepte verwendet werden, wie sie z.B. aus der Konzeption zu VKS-Umwelt (Seggelke et al. 1997) und GEIN 2000 (German Environmental Information Network) bekannt sind.

Da sich DIWA derzeit noch in Entwicklung befindet, wird die Gelegenheit genutzt, um frühzeitig zu prüfen, welche Aspekte für die Integration eines E-Commerce-Systems zu berücksichtigen sind. Die Entwicklung soll so ausgelegt sein, dass DIWA auch von anderen Bereichen, z.B. allgemeine Verwaltung und Verkehr, genutzt werden kann.

Literatur

Adobe (1999): <http://www.adobe.com/supportservice/custsupport/LIBRARY/4e06.htm>

Blasius, K.; Maier, K. (1997): Digitaler Umweltatlas (DUA); MEMO Nr. 95 des Lehrstuhls für Softwaretechnologie der Universität Dortmund, Workshopbericht "Hypermediatechnik für Umweltdaten", (Hrsg.) Greve, K.; Keitel, A.; Seggelke, J.; Schütz, Th.; Sobottka, H.-G..

Lautner, P.; Weidemann, R.; Witt-Hock J. (1998): "Anforderungen an ein hypermediales Umwelt-Fachinformationssystem". Tagungsband des 1. GI-Workshops "Hypermedia im Umweltschutz", Ulm, Metropolis Verlag 1998.

Maurer, H. (1996): HyperWave - The Next Generation Web Solution. Addison Wesley.

Mayer-Föll, R., Jaeschke, A. (1998): Projekt GLOBUS: Multimediales Recherchieren und Verarbeiten von globalen Umweltsachdaten im Umweltinformationssystem Baden-Württemberg Phase V - 1998.

MS IIS (1999): <http://www.microsoft.com/ntserver/web/default.asp>.

MS IS (1999): [http://www.microsoft.com/ntserver/web/techdetails/overview/
IndexServer.asp](http://www.microsoft.com/ntserver/web/techdetails/overview/IndexServer.asp).

Pursche, K.; Tochtermann, K.; Fuchs, Ch. (1999): "Zertifizierung und Signierung von Java Applets für deren Nutzung in Internet-basierten Umweltinformationssystemen"; Tagungsband des 2. GI-Workshop Hypermedia im Umweltschutz, Nürnberg, Metropolis Verlag 1999.

Seggelke, J.; Mohaupt-Jahr, B. (1997): "Der Verweis- und Kommunikationsservice des Umweltbundesamts – Ein Modellfall für das Umwelt-Intranet". In: W. Geiger, A. Jaeschke, D. Rentz, E. Simon, T. Spengler, L. Zilliox und T. Zundel (Hrsg.): Umweltinformatik '97. 11. Internationales Symposium, Straßburg, September 1997, Tagungsband. Metropolis-Verlag, Marburg, 1997.

Tochtermann, K.; Riekert, W.-F.; Kadric, L.; Kramer, R.; Schmidt, R.; Geiger, W.; Peter, N.; Reissfelder, M.; Doberkat, E.-E.; Sobottka, H.-G.; Keitel, A.; Zitzmann, W.; Schuetz, Th.; Burkhardt, J. (1998): "HUDA: A toolbox for environmental report production". Tagungsband des 12. Internationalen Symposiums "Informatik im Umweltschutz", Bremen, Metropolis Verlag 1998.

UDK (1998): http://www.uvm.baden-wuerttemberg.de/www-udk/RPT_DATA/about_d.html.

Intranet-basierte Verwaltung und Verfügbarmachung multimedialer Umweltdokumentenbestände

Inge Henning, Renate Ebel¹, Martina Tauber², Katrin Pursche, Christian Fuchs,
Klaus Tochtermann³

1 Einführung

Das Internet bzw. ein Intranet werden in der Umweltverwaltung bislang im wesentlichen in zwei Bereichen eingesetzt. Zum einen werden aufbereitete Umweltdaten und –berichte dem Bürger über das Internet zur Verfügung gestellt. Auf der anderen Seite wird über das Intranet bzw. das Internet die Recherche nach umweltrelevanten Dokumenten und Arbeitsmaterialien über spezielle Zugangssysteme wie die XfaWeb-Familie (Weidemann et al. 1998) oder den UDK (Umwelt-Daten-Katalog) (Nikolai et al.1999) ermöglicht. Neben dieser Menge ausgewählter Dokumente existiert aber eine weit größere Anzahl an Dokumenten, die ebenfalls für einen grossen Nutzerkreis von Interesse sind. Diese Dokumente entstehen im Arbeitsalltag und werden dort für die Verrichtung der Tagesarbeit genutzt. Sie enthalten oft detaillierte Informationen und werden zum Teil mit hohem Aufwand erstellt (z.B. thematische Karten). Bisher sind die

¹ I. Henning, R.Ebel; Ministerium für Umwelt und Verkehr Baden-Württemberg,, Postfach103439, D-70029 Stuttgart, E-Mail: Inge.Henning | Renate.Ebel@uvm.bwl.de.

² M. Tauber; Landesanstalt für Umweltschutz Baden-Württemberg, Postfach 210752, D-76157 Karlsruhe, E-Mail: Martina.Tauber@lfuka.lfu.bwl.de.

³ K. Pursche, Ch. Fuchs, K. Tochtermann; Forschungsinstitut für anwendungsorientierte Wissensverarbeitung, Postfach 2060, D-89081 Ulm,
E-Mail: pursche | fuchs | tochterm@faw.uni-ulm.de.

Möglichkeiten, solche Dokumente anderen Anwendern verfügbar zu machen bzw. nach ihnen recherchieren zu können, eingeschränkt.

Mit der Realisierung des im Folgenden beschriebenen Systems können multimediale Dokumentenbestände auf unterschiedlichen Plattformen über eine einheitliche Sicht für einen großen Nutzerkreis verfügbar gemacht werden.

Dieser Beitrag stellt die Ziele und die Realisierung des Projektes DIWA (Dokumentenverwaltung im Web-Archiv von UVM und LfU) vor, aber auch bereits gewonnene Erfahrungen mit dem Prototypen und zukünftige Aufgaben des Systems.

2 Ausgangssituation und Ziele

In der täglichen Arbeit der Umweltverwaltungen wird eine Menge von Dokumenten erzeugt, die nur zu einem geringen Teil in speziellen Systemen nachgewiesen und somit recherchierbar gemacht werden. Alle anderen Dokumente werden in meist organisatorisch strukturierten Ablagefächern im Intranet, in gemeinsamen Dateiablagen oder im Mailbereich des Bürokommunikationssystems abgelegt und in der Regel nicht verschlagwortet. Eine gezielte und strukturierte Recherche nach bestimmten Informationen ist für einen Anwender praktisch unmöglich.

Ziel des Projektes DIWA (Henning et al. 1999) ist es nun, multimediale Dokumentenbestände unter einer einheitlichen Sicht und unabhängig von ihrer physischen Ablage in einem Web-Archiv verfügbar zu machen. Aufgrund einer einfachen, aber effektiven Recherchemöglichkeit soll die Mehrfachnutzung von Dokumenten erleichtert werden. Gleichzeitig sollen die Anwender des Systems angeregt werden, eigenes Wissen in Form von Dokumenten in das Web-Archiv einzustellen. Erst durch diesen zweiten Schritt kann zusätzlich zur Archivierung und Recherchierbarkeit von Dokumenten ein Informationsaustausch erreicht werden.

Da mit dem zu entwickelnden System eine große Anzahl von Anwendern erreicht werden soll, werden moderne Internettechnologien bei der Realisierung eingesetzt. Damit erübrigt sich die Installation von Software auf den Client-Rechnern, und der Anwender kann in seiner gewohnten Bürokommunikationsumgebung arbeiten sowie über den Standard-Web-Browser die DIWA-Oberfläche bedienen.

Neben den bereits genannten Dokumenten sind auch spezielle, bereits verschlagwortete Bestände verfügbar zu machen, ohne dass der Anwender über dafür spezialisierte Recherchesysteme verfügen oder diese erlernen muss.

3 DIWA – Ein Überblick

3.1 Der Zugang zu multimedialen Dokumentenbeständen

Den zentralen Baustein des Systems DIWA stellt die virtuelle Ablagestruktur dar. Die in einem Intranet auf verschiedenen Servern vorliegenden Dokumentenbestände sind jeweils in sich nach verschiedenen, meist organisatorischen Gesichtspunkten gegliedert. Für eine Verfügbarmachung aller Dokumentenbestände in einem System muss nun eine einheitliche thematische Sicht geschaffen werden. Da die Dokumentenbestände nicht umstrukturiert oder redundant gehalten werden sollen, bietet sich der Aufbau einer virtuellen, übergeordneten Struktur, z.B. nach Umweltthemen, an. Diese virtuelle Ablagestruktur wird in der von DIWA verwendeten Basissoftware, dem Microsoft Internet Information Server (MS IIS) (MS IIS 1999), abgebildet. Die Verbindung der physischen Dokumentenbestände mit der virtuellen Ablagestruktur erfolgt über logische Verknüpfungen.

Der mit dem MS IIS zusammen arbeitende Microsoft Index Server (MS IS) (MS IIS 1999) indiziert die über die virtuelle Ablagestruktur erreichbaren Dokumen-

tenbestände und legt einen übergreifenden Index an. An diesen Index können nun Suchanfragen gestellt werden. Die Recherche erlaubt nicht nur die Suche im Volltext der Dokumente, sondern auch nach Dokumenteigenschaften. In DIWA werden Dokumenteigenschaften als Metainformationen genutzt. Diese Dokumenteigenschaften werden nicht wie gewohnt getrennt von den Dokumenten gehalten, sondern in diesen abgelegt. Praktisch können die Dateieigenschaften von Microsoft Office- und PDF-Dokumenten sowie HTML-Meta-Tags genutzt werden. Problematisch ist die Festlegung einer Menge von Metainformationen für die Verschlagwortung der einzelnen Dokumentenbestände. Sie muss auf der einen Seite umfassend genug sein, um eine effektive Recherche zu erlauben. Auf der anderen Seite darf die Befüllung dieser Metainformationen für den Autor keinen hohen Aufwand bedeuten, da dadurch die Akzeptanz gegenüber DIWA sinken würde.

3.2 Recherchekomponente

Für die Suche nach Dokumenten werden in DIWA verschiedene Komponenten angeboten. Zum einen kann die Recherche über Suchformulare durchgeführt werden. Hierbei wird zwischen einer einfachen Volltextsuche und der erweiterten Suche unterschieden. Die erweiterte Suche bezieht neben der Volltextrecherche auch die Suche nach speziellen Metainformationen ein. Die Zusammenstellung der Metainformationen ist abhängig von dem Dokumentenbestand, der recherchiert werden soll. Für die festgelegten Gruppen von Metainformationen werden deshalb jeweils angepasste Formulare angeboten. Der Anwender kann in der Recherche den Suchbereich auf einzelne Themen der virtuellen Ablagestruktur einschränken. Unterstützung bei der Begriffsfindung wird durch einen Umweltthesaurus gewährleistet.

Alternativ zur Recherche über Suchformulare wird ein navigatorischer Zugang zu den Dokumentenbeständen in Form des DIWA Explorers angeboten. Im DIWA Explorer wird die virtuelle Ablagestruktur abgebildet. Für die einzelnen

Themen kann sich der Anwender nun alle vorhandenen Dokumente anzeigen lassen.

Beide Komponenten wurden als Java Applets realisiert. Damit entfällt die Installation zusätzlicher Software auf dem Client-Rechner. Der Anwender startet DIWA aus seinem Standard-Web-Browser heraus.

3.3 Autorenkomponente

Die Autorenkomponente dient ausschließlich dem Einstellen und dem Herunterladen von multimedialen Dokumenten in ihrem originalen Dateiformat in bzw. aus dem Web-Archiv. Der Zugang zum Web-Archiv erfolgt auch hier über die virtuelle Ablagestruktur. Der Nutzer muss also keine Kenntnis darüber besitzen, auf welchem Server des Intranets der Dokumentenbestand liegt, in den er sein Dokument einstellen möchte. Er orientiert sich lediglich an der thematischen Ordnung.

Auch die Autorenkomponente wurde als Java Applet implementiert. Da sie auf den lokalen Rechner des Autors zugreift, ist eine Signierung notwendig (Pur-sche et al. 1999).

4 Projektstand und Erfahrungen

DIWA wurde bereits im UVM (Ministerium für Umwelt und Verkehr Baden-Württemberg) und der LfU (Landesanstalt für Umweltschutz Baden-Württemberg) installiert. Neben allgemein zugänglichen Daten der Institutionen wurden Dokumentenbestände von Testnutzern eingebunden. Die erste Aufgabe der Testnutzer lag in der Festlegung einer virtuellen Ablagestruktur. Als zweiter Schritt wurde eine für ihre Dokumentenbestände optimale Menge an Metainformationen evaluiert. Dabei stellte sich heraus, dass zumeist auf Standardeigenschaften wie Kategorie, Stichworte und Kommentare zurückgegriffen wurde. Für diese Metainformationen wurden dann Listen mit zu verwendenden

Inhalten (Werten) entwickelt. Als aufwendig stellt sich nun die Verschlagwortung bereits vorhandener Dokumentenbestände dar. Bisher wird so vorgegangen, dass die Verschlagwortung kontinuierlich nach Bedeutung und zeitlicher Gültigkeit der Dokumente vorgenommen wird. Diskussionen mit Anwendern ergaben, dass für die Unterstützung der Autoren bei dieser Arbeit zwei Wege existieren. Zum einen könnten entsprechende Komponenten bereits in die Software zur Erstellung der Dokumente eingebaut werden (z.B. als Makro in MS Office). Zum anderen könnte direkt in der Autorenkomponente überprüft werden, ob ein Dokument, das ein Autor in das Web-Archiv einstellen möchte, für den entsprechenden Dokumentenbestand ausreichend verschlagwortet wurde. Ist das nicht der Fall, wird dem Autor Unterstützung angeboten. Erst dann kann das Dokument in das Web-Archiv eingetragen werden. Die zweite Variante wäre unabhängig von der zur Erstellung des Dokumentes verwendeten Software. Gleichzeitig würde gesichert werden, dass nur verschlagwortete Dokumente in das Web-Archiv eingestellt werden. Gerade der zweite Aspekt ist von Bedeutung: Eine Recherche über Metainformationen ist nur dann sinnvoll, wenn auch alle Dokumente vollständig erschlossen wurden. In diesem Zusammenhang sind bereits vorhandene Verschlagwortungsassistenten wie in GEIN2000 (GEIN 1999) zu untersuchen.

5 Ausblick

Aufgrund der Rückmeldungen der Anwender können die folgenden Anforderungen an zukünftige Arbeiten gestellt werden. So wird die Einbindung von Datenbankabfragen in die Recherchekomponente angestrebt, um weitere Informationsquellen einbinden zu können. Darüber hinaus wird überlegt, weitere spezielle Fachinformationsdienste in DIWA zu integrieren, um das Informationsspektrum zu erweitern.

In Verbindung mit der Recherchekomponente sollen schließlich Suchanfragen bzw. Suchergebnisse gespeichert werden. Mit dieser Funktionalität wird es Anwendern ermöglicht, die für die Durchführung einer Aufgabe relevanten Dokumente stets schnell im Zugriff zu haben. Es entfällt damit der Aufwand, für jede Sitzung Suchanfragen neu zusammenzustellen.

Beim aktuellen DIWA-Prototypen müssen alle virtuellen Verzeichnisse an zentraler Stelle manuell eingerichtet, einem physischen Verzeichnis zugeordnet und dort gewartet werden. Die zentrale Administration trifft in analoger Weise bei der Einrichtung von Suchformularen zu, die auf Dokumentenbestände einzelner Anwender zugeschnitten sind. Wollen Anwender z.B. nur den Namen eines Verzeichnisses ändern, so müssen sie sich für die Durchführung dieser Aufgabe an diese zentrale Stelle wenden. Da in der Fortsetzung weitere Anwender für die Nutzung von DIWA gewonnen werden sollen, ist bereits jetzt absehbar, dass hier ein Engpass entstehen wird. Vor diesem Hintergrund soll ein Werkzeug entwickelt werden, mit dem jeder Anwender die für seine Dokumentenbestände benötigte virtuelle Verzeichnisstruktur selbst administrieren kann. Dieses Werkzeug umfasst Funktionalitäten zum Einrichten, Ändern und Erweitern der virtuellen Ablagestruktur. Zudem soll Anwendern ein Werkzeug angeboten werden, über das sie in Analogie zur Verwaltung von virtuellen Ablagestrukturen ihre eigenen Suchformulare gestalten können.

Literatur

GEIN (1999): <http://www.gein.de>.

Henning, I.; Ebel, R.; Tauber, M.; Tochtermann, K.; Pursche, K.; Kussmaul, A.; Schultze, A.: Internetbasierte Dokumentenverwaltung heterogener Umweltdokumentenbestände; Tagungsband Umweltinformatik '99, Metropolis-Verlag, Marburg, 1999.

MS IIS (1999): <http://www.microsoft.com/ntserver/web/default.asp>.

MS IS (1999): [http://www.microsoft.com/ntserver/web/techdetails/overview/
IndexServer.asp](http://www.microsoft.com/ntserver/web/techdetails/overview/IndexServer.asp).

Nikolai, R.; Kazakos, W.; Kramer, R.; Behrens, S.; Swoboda, W.; Kruse, F. (1999): WWW-UDK 4.0: Die neue Generation eines Web-Portals zu deutschen und österreichischen Umweltdaten; Tagungsband Umweltinformatik '99; Metropolis-Verlag; Marburg.

Pursche, K.; Tochtermann, K.; Fuchs, Ch. (1999): "Zertifizierung und Signierung von Java Applets für deren Nutzung in Internet-basierten Umweltinformationssystemen"; Tagungsband des 2. GI-Workshop Hypermedia im Umweltschutz, Nürnberg, Metropolis Verlag 1999.

Weidemann, R.; Geiger, W.; Reißfelder, M.; Zilly, G. (1998): Von AlfaWeb zur XfaWeb-Systemfamilie; Projekt GLOBUS: Multimediales Recherchieren und Verarbeiten von globalen Umweltsachdaten im Umweltinformationssystem Baden-Württemberg; Phase V 1998; Wissenschaftliche Berichte; FZKA 6250.