

Zusammenfassung

Moderne Umweltverwaltungen setzen heutzutage in vielen Bereichen auf Intranet- und Internet-Technologien zur wirtschaftlichen Bereitstellung, Erschließung und Nutzung von multimedialen Umweltdokumenten. Während die Recherche nach und die Nutzung von verfügbar gemachten Umweltdokumenten von sehr vielen Mitarbeitern der Umweltverwaltungen wahrgenommen wird, haben nur wenige Anwender die Möglichkeit, in einer aktiven Rolle selbst Umweltdokumente bereitzustellen. Die Arbeitsproduktivität kann in Umweltverwaltungen jedoch entscheidend erhöht werden, wenn auch größere Anwenderkreise Umweltdokumente, die von allgemeiner Bedeutung sind, verwaltungsintern über ihre Bürokommunikationsumgebung in einem auf Internettechnik basierenden Umweltdokumenten-Archiv bereitstellen können.

Ziel des Projekts Dokumentenverwaltung im Web-Archiv von Ministerium für Umwelt und Verkehr und Landesanstalt für Umweltschutz Baden-Württemberg (DIWA) ist die Konzeption und Entwicklung eines Systems, das den Aufbau, die Pflege und die Nutzung eines Web-Archivs für multimediale Umweltdokumente erlaubt. In das Web-Archiv sollen in erster Linie solche Umweltdokumente systematisch eingestellt und recherchierbar gemacht werden, die in spezialisierten Recherche- und Redaktionssystemen, wie z.B. dem UDK (Umweltdatenkatalog) und dem HUDA-Werkzeugkasten (Hypermediatechnik für Umweltdaten), nicht berücksichtigt werden können, aber dennoch für einen großen Anwenderkreis von Interesse sind.

Die Dokumente, die mittels DIWA verfügbar gemacht werden, liegen physisch verteilt und in unterschiedlichen Dateiformaten (z.B. Microsoft Office, HTML, PDF, XML) im Intranet der Landesanstalt für Umweltschutz (LfU) und des Ministeriums für Umwelt und Verkehr (UVM) Baden-Württemberg unter verschiedenen Betriebssystemen (UNIX, Microsoft Windows NT) vor.

Für das Einstellen und die Recherche werden die Dokumente einer logischen virtuellen Ablagestruktur im Microsoft Internet Information Server zugeordnet. Diese Ablagestruktur ist frei definierbar und kann sich etwa an der internen Organisation einer Umweltverwaltung (Abteilungen, Referate etc.) oder an Umweltthemen (Luft, Wasser etc.) orientieren. Der zusätzlich verwendete Microsoft Index Server indiziert den Volltext und die Dokumenteigenschaften dieser Dokumente und erlaubt so eine effiziente Recherche. Eine redundante Haltung der Dokumente wird vermieden.

Die Autorenkomponente von DIWA erlaubt das Einstellen von Dokumenten in das Web-Archiv direkt aus der Bürokommunikationsumgebung des Autors heraus. In dem so aufgebauten Web-Archiv kann nun mittels einer Recherche-komponente nach Dokumenten gesucht werden. Die Recherche kann zum einen über Volltextsuche und Metainformationen mittels Suchformularen vorgenommen werden. Alternativ dazu sind die Dokumente über einen navigatorischen Zugang erreichbar, der den Zugriff auf die Dokumente entlang der Ablagestruktur unterstützt.

Mit DIWA können große Anwenderkreise an dem systematischen Aufbau eines Web-Archivs für Umweltdokumente beteiligt werden. Dadurch erhöht sich die Transparenz über die in einer Umweltverwaltung vorhandenen Umweltdokumente. Als Folge wird die Mehrfachnutzung dieser Dokumente angeregt und erleichtert, was in einem hohen Maße zur Wirtschaftlichkeit des Systems beiträgt.

Die über DIWA in einem Web-Archiv eingestellten Umweltdokumente können einem großen Anwenderkreis zur Verfügung gestellt werden. Da die Autoren der Dokumente diese Aufgabe selbst wahrnehmen können, ist es möglich, Dokumente ohne zeitliche Verzögerung bereitzustellen. Gleichzeitig führt dieser Aspekt zu einer Entlastung zentraler Stellen, die bislang für das Einstellen von Dokumenten in das verwaltungsinterne Intranet verantwortlich sind. Dies trägt

zu einer erheblichen Effizienzsteigerung beim Aufbau von Umweltdokumenten-Archiven bei.

Die einzelnen Komponenten von DIWA werden über Standard-Web-Browser zur Verfügung gestellt. Auf die Installation spezieller Software wird daher verzichtet. Dadurch entsteht nur ein geringer Einarbeitungsaufwand auf Seiten der Anwender, dem ein großes Nutzenpotential gegenübersteht.

Die Entwicklung ist so angelegt, dass DIWA in beliebigen Fachbereichen eingesetzt werden kann.

Der Prototyp von DIWA wurde im UVM und in der LfU installiert und wird dort in heterogenen Systemumgebungen mit verschiedenen Dokumentenbeständen der Institutionen getestet.

Das FAW wurde mit der Durchführung des Projekts DIWA durch das Ministerium für Umwelt und Verkehr Baden-Württemberg sowie die Landesanstalt für Umweltschutz Baden-Württemberg beauftragt. Die Stabsstelle für Verwaltungsreform im Innenministerium unterstützt dieses Projekt, welches Teil der Phase VI des F+E-Vorhabens GLOBUS (Globale Umweltsachdaten) ist.

Die Ergebnisse von DIWA sind Eigentum des Landes Baden-Württemberg. DIWA kann nach besonderer Vereinbarung und Regelung des Entgelts zur Nutzung ausserhalb der Zuständigkeitsbereiche der Eigentümer weitergegeben werden. Die Erkenntnisse aus dem F+E-Vorhaben werden interessierten Stellen zur Verfügung gestellt in der Hoffnung, weitere Partner zu gewinnen.

Inhaltsverzeichnis

ZUSAMMENFASSUNG	I
ABBILDUNGSVERZEICHNIS	4
1 RECHENSCHAFTSBERICHT	5
1.1 Projektorganisation	5
1.2 Aktivitätsübersicht.....	6
2 ANFORDERUNGEN AN DAS SYSTEM DIWA.....	9
2.1 Allgemeine Zielstellung von DIWA.....	9
2.2 Anforderungen und Rahmenbedingungen.....	10
2.2.1 Einstellen von Dokumenten in das Web-Archiv	11
2.2.2 Recherche nach Dokumenten im Web-Archiv	12
2.2.3 Technische Rahmenbedingungen.....	14
2.2.4 Offenheit gegenüber anderen Systemen	15
3 DIWA-ARCHITEKTUR UND REALISIERUNG	17
3.1 Systemarchitektur	17
3.2 Softwarearchitektur.....	21
3.2.1 Verwendete Software und Technologien.....	23
3.2.2 Softwarearchitektur für DIWA.....	35
4 DIWA-FUNKTIONALITÄT FÜR WEITERE SYSTEME	39
4.1 Anforderungen in FA-ZSV	40
4.2 FA-ZSV-Architektur und Realisierung.....	40

4.2.1	Verwendete Technologien.....	41
4.2.2	FA-ZSV-Softwarearchitektur.....	43
4.3	Stand der Entwicklungen.....	44
5	ANWENDER UND ERSTE ERFAHRUNGEN.....	47
6	VERGLEICH DER ANFORDERUNGEN MIT DEM ERREICHTEN PROJEKTSTAND.....	49
6.1	Erfüllung grundlegender Anforderungen.....	49
6.1.1	Einbindung von DIWA in die vertraute Bürokommunikationsumgebung.....	49
6.1.2	Berücksichtigte Dateiformate bei der Indizierung.....	50
6.1.3	Systematischer Zugang zum Web-Archiv.....	50
6.1.4	Einbinden bereits vorhandener Umweltdokumentenbestände in das Web-Archiv.....	51
6.1.5	Einbindung von DIWA in die vorhandenen Systemumgebungen...	51
6.1.6	Intranet-weiter Einsatz von DIWA.....	52
6.2	Einstellen von Dokumenten in das Web-Archiv.....	52
6.3	Recherche nach Dokumenten im Web-Archiv.....	53
6.4	FA-ZSV.....	55
6.5	Offenheit gegenüber anderen Systemen.....	56
7	AUSBLICK.....	57
	ABKÜRZUNGSVERZEICHNIS.....	59
	LITERATURVERZEICHNIS.....	61

ANHANG 1: INSTALLATION UND ADMINISTRATIVE AUFGABEN.....	63
ANHANG 2: BENUTZERHANDBUCH.....	121
INDEX	171
ANHANG 3: VERÖFFENTLICHUNGEN.....	175

Abbildungsverzeichnis

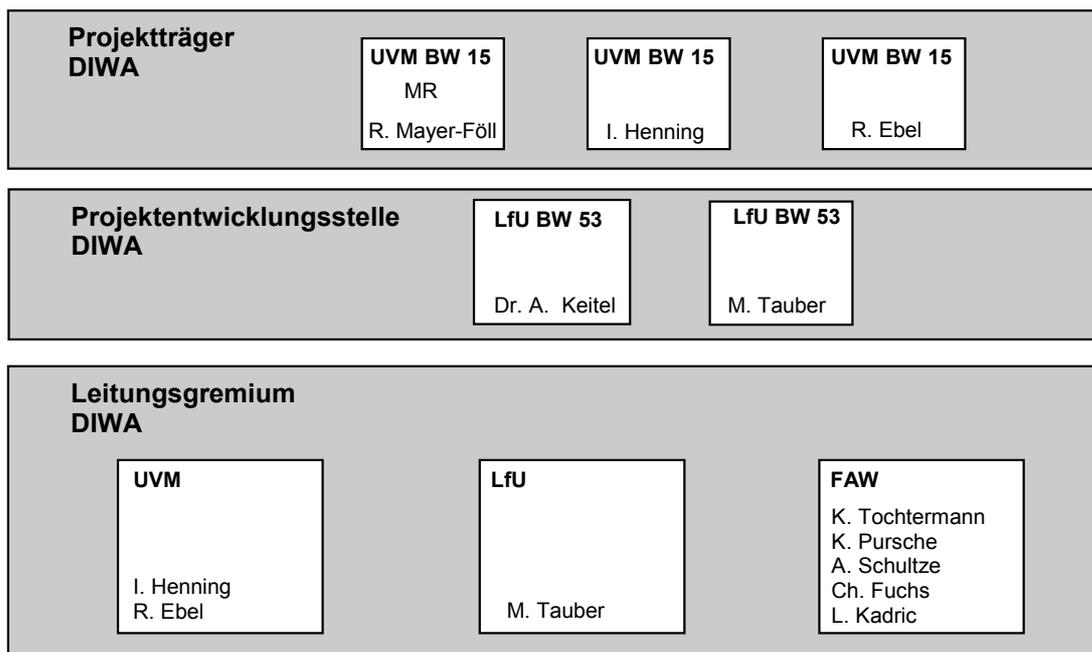
Abbildung 1: Übersicht über den Wirkungskreis von DIWA	11
Abbildung 2: Zusammenspiel zwischen einer physischen und einer virtuellen Ablagestruktur.....	18
Abbildung 3: Komponenten des Web-Archivs.....	19
Abbildung 4: Prinzipieller Ablauf der Autorenkomponente	20
Abbildung 5: Prinzipieller Ablauf einer Recherche	21
Abbildung 6: Aufbau der Verzeichnisse im WWW-Dienst.....	29
Abbildung 7: Verletzung der Zugriffsrechte entfernter Dokumentenbestände..	32
Abbildung 8: Two-Tier-Architektur	34
Abbildung 9: Three-Tier-Architektur	35
Abbildung 10: Software der Autorenkomponente.....	36
Abbildung 11: Software der Recherchekomponente.....	37
Abbildung 12: Softwarearchitektur DIWA	38
Abbildung 13: Systemarchitektur der FA-ZSV	41
Abbildung 14: Softwarearchitektur der FA-ZSV.....	43
Abbildung 15: Erfassen eines Erlasses im ersten Prototyp der FA-ZSV	44

1 Rechenschaftsbericht

Das Forschungs- und Entwicklungsvorhaben „Dokumentenverwaltung im Web-Archiv von UVM und LfU (DIWA)“ ist ein Projekt innerhalb des informationstechnischen Ressortplans des Ministeriums für Umwelt und Verkehr Baden-Württemberg.

1.1 Projektorganisation

Das Projekt „Dokumentenverwaltung im Web-Archiv von UVM und LfU (DIWA)“ wurde in folgender Projektorganisation (Stand 30.11.1999) durchgeführt.



Der Lenkungsausschuss GLOBUS steuert das Projekt fachübergreifend.

1.2 Aktivitätsübersicht

Im laufenden Projektjahr erfolgten die in diesem Abschnitt chronologisch aufgelisteten Aktivitäten.

Datum	Aktivität	Arbeitsergebnisse
15.12.1998	1. Sitzung des LG DIWA, UVM Stuttgart	1. Bildung des Leitungsgremiums 2. Anforderungsanalyse 3. Definition der System-Umgebung
19.01.1999	2. Sitzung des LG DIWA, FAW Ulm	1. Präsentation des ersten Prototyps 2. Fortsetzung der Diskussion der Anforderungsanalyse
15.04.1999	Expertenrunde Umweltinformationssysteme, FAW Ulm	Vortrag über DIWA, Präsentation des DIWA-Funktionsprototyps
28.04.1999	Präsentation des DIWA-Prototyps, UVM Stuttgart	1. Einführung in und Vorführung von DIWA 2. Diskussion des Prototypen
04.05.1999	Workshop des Innenministeriums Baden-Württemberg, FAW Ulm	Vortrag über DIWA
10.06.1999	16. Sitzung des Arbeitskreises Umweltinformationssysteme, UVM Stuttgart	Präsentation DIWA
10.-11.06.1999	Workshop Umweltdatenbanken, Karlsruhe	Veröffentlichung und Vortrag über die technischen Grundlagen des ersten Prototyps von DIWA
15.06.1999	Präsentation des DIWA-Prototyps, MLR Stuttgart	1. Einführung 2. Vorführung des Prototyps 3. Diskussion der Anwendbarkeit im MLR
17.06.1999	10. Koordinierungsausschuss IuK im Geschäftsbereich des UVM, UVM Stuttgart	Präsentation DIWA
30.06.1999	3. Sitzung des LG DIWA, UVM Stuttgart	1. Einführung Ref.72 des UVM als Testnutzer 2. Sachstandsbericht

27.07.1999	Besuch von Vertretern des Landes Schleswig-Holstein, UVM Stuttgart	Präsentation DIWA
30.08.- 01.09.1999	Umweltinformatik-Tagung, Magdeburg	Veröffentlichung und Vortrag über DIWA
10.09.1999	1. Sitzung FA-ZSV, ZSV GAA Stuttgart	1. Diskussion Feinkonzept 2. Übergabe der Daten
24.09.1999	4. Sitzung des LG DIWA, UVM Stuttgart	1. Einführung der Akademie für Natur- und Umweltschutz des UVM als Testnutzer 2. Sachstandsbericht
11.10.1999	2. Sitzung FA-ZSV, ZSV GAA Stuttgart	1. Abstimmung der Tabellenstruktur 2. Abstimmung der Benutzeroberfläche
12.10.1999	Expertenrunde Umweltinformationssysteme, FAW Ulm	Vortrag über DIWA, Präsentation des DIWA-Funktionsprototypen
25.11.1999	GLOBUS-Abschlusspräsentation, Stuttgart	Präsentation des DIWA-Funktionsprototyps

Tabelle 1: Aktivitätsübersicht für das Projekt DIWA

2 Anforderungen an das System DIWA

2.1 Allgemeine Zielstellung von DIWA

Die übergreifenden Ziele des Projekts 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 Systeme, wie das HUDA-Redaktionssystem (Tochtermann et al. 1999) oder das AlfaWeb-Autorensystem (Weidemann 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 1999), erschlossen und recherchierbar gemacht werden, sollen weitgehend durch die Anwender selbst über das Internet oder 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 über das Web-Archiv recherchierbar sein, ohne dass Anwender ein spezielles Werkzeug auf ihrem Rechner installiert haben und nutzen müssen. Die Recherche nach Dokumenten soll nicht nur über eine Volltextsuche sondern auch über vorhandene Metainformationen in den Dokumenteigenschaften möglich sein. Schließlich soll DIWA so offen gestaltet werden, dass auch andere, vorwiegend auf einen bestimmten Themenbereich spezialisierte Systeme, wie etwa die Fachanwendung der Zentralen Stelle für die Vollzugsunterstützung beim Staatlichen Gewerbeaufsichtsamt Stuttgart (FA-ZSV), angebunden werden können.

2.2 Anforderungen und Rahmenbedingungen

Der Anwender soll ohne einen hohen Grad an technischem Hintergrundwissen in die Lage versetzt werden, Dokumente für andere bereitzustellen und selbst recherchieren zu können. Bereits vorhandene und sich entwickelnde Dokumentenbestände, die auf verschiedenen Servern eines Intranets liegen, sollen in ein Web-Archiv eingebunden werden. Ein Teil der Dokumentenbestände des Intranets wird auch für das Internet in Form eines Bürgerservices bereitgestellt. Berechtigte Autoren könnten auch hier die Möglichkeit zum Einstellen von Dokumenten erhalten. Darüber hinaus können mit DIWA Ressourcen aus dem Internet erschlossen werden. Abbildung 1 zeigt den Wirkungskreis von DIWA.

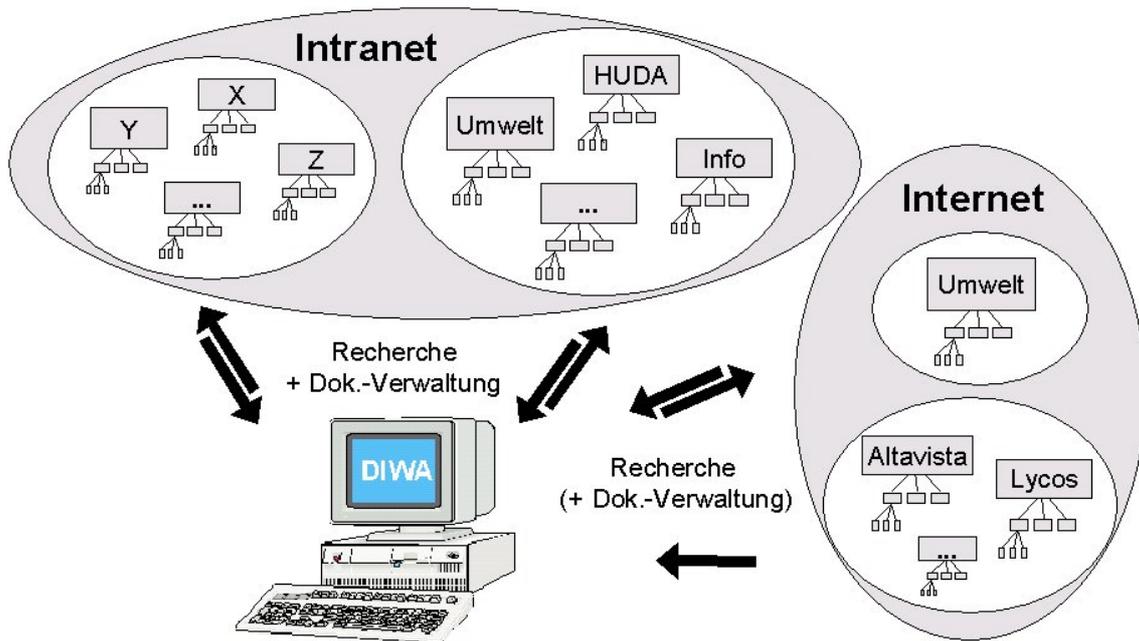


Abbildung 1: Übersicht über den Wirkungskreis von DIWA

Für die Realisierung von DIWA wurde die schnelle Entwicklung eines ersten Prototyps gefordert, um möglichst viele Anforderungen der Anwender bereits in dieser Phase des Projekts umsetzen zu können. Eine weitere Anforderung besteht darin, weitgehend kommerzielle Software einzusetzen. So können Standardfunktionalitäten, wie das Erstellen eines Volltextindex, von kommerziellen Produkten übernommen werden. Dies trägt entscheidend zu einer Erhöhung der Kosteneffizienz bei den Realisierungsarbeiten bei. Schließlich wird das zu entwickelnde System dadurch einfacher wartbar sowie anpassbar und trägt so der Tatsache der hohen Innovationsgeschwindigkeit im Bereich der Informations- und Kommunikationstechnologie Rechnung.

2.2.1 Einstellen von Dokumenten in das Web-Archiv

Das Einstellen von Dokumenten in das Web-Archiv soll Anwendern erlauben, in der Rolle eines Autors anderen Anwendern Umweltinformationen recherchier

bar und nutzbar zu machen. An die zu entwickelnde Autorenkomponente werden folgende Anforderungen gestellt:

- Autoren sollen ihre in der gewohnten Bürokommunikationsumgebung erstellten Dokumente auf Web Servern einstellen und somit einer Suchmaschine bekannt geben können. Des Weiteren sollen je nach Zugriffsrechten Verzeichnisse gelöscht sowie Dokumente heruntergeladen oder gelöscht werden können.
- Das Einstellen von Dokumenten soll über eine virtuelle Ablagestruktur ermöglicht werden. Durch dieses Vorgehen wird erreicht, dass der Anwender keine Kenntnis darüber besitzen muss, auf welchem Server des Intranets der entsprechende Dokumentenbestand vorliegt. Vielmehr orientiert sich der Autor an den angebotenen Themen in der Ablagestruktur. Alternativ dazu bleibt das Einstellen von Dokumenten über das Dateisystem (z.B. über Windows Explorer) möglich.
- Die Autorenkomponente soll für Anwender nur nutzbar sein, wenn sie über entsprechende Zugriffsrechte auf den Dokumentenbeständen des Web-Archivs verfügen. Damit wird vermieden, dass beliebige Anwender die Umweltdokumentenbestände verändern können.
- Die Autorenkomponente soll auf Internettechnologien basieren, in heterogenen Umgebungen lauffähig sein und über einen Standard-Web-Browser gestartet werden können. Dadurch erübrigt sich die lokale Installation der Autorenkomponente, und der Arbeitsaufwand verringert sich bei einem großen Anwenderkreis erheblich.

2.2.2 Recherche nach Dokumenten im Web-Archiv

Für Anwender des Systems soll die Recherche in den Dokumentenbeständen des Web-Archivs über gängige Web-Browser und nicht über Oberflächen spezieller Werkzeuge ermöglicht werden. Dennoch sollen weitestgehend auch Do

kumente, die in anderen Systemen bereits existieren (z.B. Dokumente im DRS (Document Retrieval System) der Abteilung 4 der LfU) über das System recherchierbar und zugreifbar gemacht werden. Über die zu entwickelnde Recherchekomponente soll es Anwendern möglich sein:

- einfach nach Dokumenten des Web-Archivs zu recherchieren und diese in der eigenen Bürokommunikationsumgebung einzusehen,
- 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 handeln als auch um benutzerdefinierte Informationen wie Zeit- oder Raumbezug,
- die Formulierung der Abfrage über einen vorgeschalteten Umweltthesaurus zu unterstützen. Damit ist es möglich, auch nach verwandten Begriffen, Ober-, Unterbegriffen etc. zu suchen,
- bei der Abfrage auszuwählen, ob sie nur in dem Web-Archiv von DIWA suchen möchten oder zusätzlich die Abfrage an weitere Umweltdokumentenbestände (z.B. HUDA) oder andere Suchmaschinen (z.B. AltaVista, Lycos) weiterleiten möchten. Dadurch kann die Suche je nach den gewünschten Ergebnissen erweitert oder eingeschränkt werden,
- die thematische Strukturierung des Web-Archivs einzusehen und damit die Recherche erleichtern und thematisch einschränken zu können,
- den Informationsgehalt der Ergebnisanzeige selbst zu bestimmen. Entsprechend der Einstellung sollen verschiedene Dateieigenschaften, eine Kurzfassung des Textes, die URL zum Dokument etc. angezeigt werden. Weiterhin soll die Sortierung der Ergebnisse nach verschiedenen Kriterien wie Trefferanzahl im Text, Titel, Dateigröße etc. festgelegt werden können,

- alternativ zur Recherche Dokumente über einen navigatorischen Zugang zum Web-Archiv erreichen zu können,
- als Rechercheergebnis alle zum Thema verfügbaren Dokumente angezeigt zu bekommen, unabhängig von den Zugriffsrechten des Anwenders. Erst beim Herunterladen des Dokuments auf den lokalen Rechner soll eine entsprechende Überprüfung stattfinden. Damit können auch Anwender, die keine Leserechte für bestimmte Dokumente besitzen, zumindest Informationen über deren Existenz erhalten.

2.2.3 Technische Rahmenbedingungen

DIWA soll nahtlos in die existierenden Systemumgebungen von UVM und LfU integrierbar sein. Da in beiden Einrichtungen unterschiedliche Systemumgebungen existieren, sind diese in der Entwicklung entsprechend zu berücksichtigen.

Im UVM soll DIWA in einer reinen Microsoft Windows NT-Umgebung lauffähig sein. D.h. sowohl der Server, auf dem DIWA installiert wird, als auch alle Server, auf denen sich relevante Dokumentenbestände befinden, arbeiten auf der Plattform Windows NT. Die Dokumente liegen in den Dateiformaten Microsoft Office (Word, Excel, PowerPoint), HTML, PDF und verschiedenen Grafikformaten vor. Weiterhin ist die „luK-Dienstvereinbarung UVM“ zu berücksichtigen. Mit ihr wird ein striktes System für die Zugriffsregelungen auf die Fächer der Ablagestruktur realisiert. Für die Vergabe und Einhaltung der Zugriffsrechte wird direkt die Benutzerverwaltung von Microsoft Windows NT genutzt. Diese Zugriffsregelungen sind unter DIWA sowohl beim Einstellen von als auch bei der Recherche nach Dokumenten unbedingt einzuhalten.

In der LfU findet man im Gegensatz zum UVM eine heterogene Infrastruktur vor. Die über DIWA zu erschließenden Dokumentenbestände sind in erster Linie auf Unix-Servern abgelegt. Es werden aber auch Microsoft Windows NT-

Server eingesetzt. Somit muss DIWA beide Plattformen bedienen können. Der Zugriff auf Dokumente unter Unix muss in jedem Fall gewährleistet werden. Da es sich um Dokumente handelt, die für das Intranet frei bereitgestellt werden, sind keine besonderen Sicherheitskonzepte zu beachten.

2.2.4 Offenheit gegenüber anderen Systemen

Die Entwicklung von DIWA soll so angelegt werden, dass eine gewisse Offenheit gegenüber anderen Systemen mit ähnlicher Funktionalität gewährleistet ist. Hierbei sind zweierlei Ausprägungen von Offenheit zu berücksichtigen:

- Es soll möglich sein, in DIWA verwendete Technologien und Funktionalitäten auch in anderen Systemen einzusetzen. Beispielsweise ist denkbar, dass nur die Recherchefunktionalität ohne Autorenfunktionalität nutzbar sein soll.
- DIWA ist so offen zu entwickeln, dass auch andere Systeme, die z.B. andere Technologien verwenden, an DIWA angebunden werden können, um so von einer einheitlichen Oberfläche für Anwender zugänglich zu sein. Als Beispiele sind hier etwa Nachweissysteme zu nennen, die selbst keine Dokumente, sondern nur Nachweise der Dokumente in relationalen Datenbanken verwalten. Zu berücksichtigen ist auch die Anbindung von Systemen, die ihre Dokumente in Datenbanken verwalten.

3 DIWA-Architektur und Realisierung

Die ersten Erfahrungen für die technische Realisierung von DIWA wurden im Vorfeld des Projekts während der Durchführung einer am FAW betreuten Diplomarbeit (Kussmaul 1998) gewonnen. Die Ergebnisse dieser Arbeit konnten direkt in das Projekt einfließen, sodass frühzeitig ein erster Prototyp bei den Anwendern installiert werden konnte. Auf Grund dieser Tatsache konnten neue Anforderungen und unterschiedliche technische Rahmenbedingungen laufend in das Projekt eingebracht und umgesetzt werden. Die folgenden Abschnitte stellen die Architektur des Systems und die für die Realisierung verwendeten Technologien mit dem Stand vor, der am Ende des laufenden Projektjahrs erreicht wurde.

3.1 Systemarchitektur

Für die in Kapitel 2 beschriebenen Anforderungen bietet sich der Einsatz einer Suchmaschine in Form eines Index Servers an. Derartige Systeme sind meist an einen Web Server gekoppelt, über den die Dokumentenbestände zugänglich gemacht werden. Ein Index Server bereitet die auf dem zugehörigen Web Server verwalteten Dokumente für eine Recherche auf, etwa in Form eines Index zur Volltextsuche. Bei vielen kommerziellen Web Servern kann zwischen einer physischen Ablagestruktur der Dokumente und einer virtuellen Ablagestruktur unterschieden werden. Die physische Ablagestruktur legt fest, auf welchem Rechner in welchem Ordner sich ein Dokument befindet. Die virtuelle Ablagestruktur erlaubt es demgegenüber, Dokumentenbestände unabhängig von ihrer physischen Ablage zu strukturieren. Für Anwender ergibt sich daraus der Vorteil, dass sie sich nicht mehr in der physischen Netzstruktur auskennen müssen. Vielmehr können sie sich die Dokumentenbestände entlang einer vir

tuellen Ablagestruktur erschließen, die z.B. nach inhaltlichen Kriterien aufgebaut sein kann. Damit wird ein intuitiver Zugang weitgehend unabhängig von der Struktur innerhalb der einzelnen Dokumentenbestände erreicht, ohne diese verändern oder redundant halten zu müssen. Abbildung 2 zeigt die Zuordnung physischer Dokumentenbestände zu einer nach Umweltthemen strukturierten virtuellen Ablagestruktur. Über diese Struktur werden die Dokumente in das Web-Archiv eingestellt und auch recherchiert.

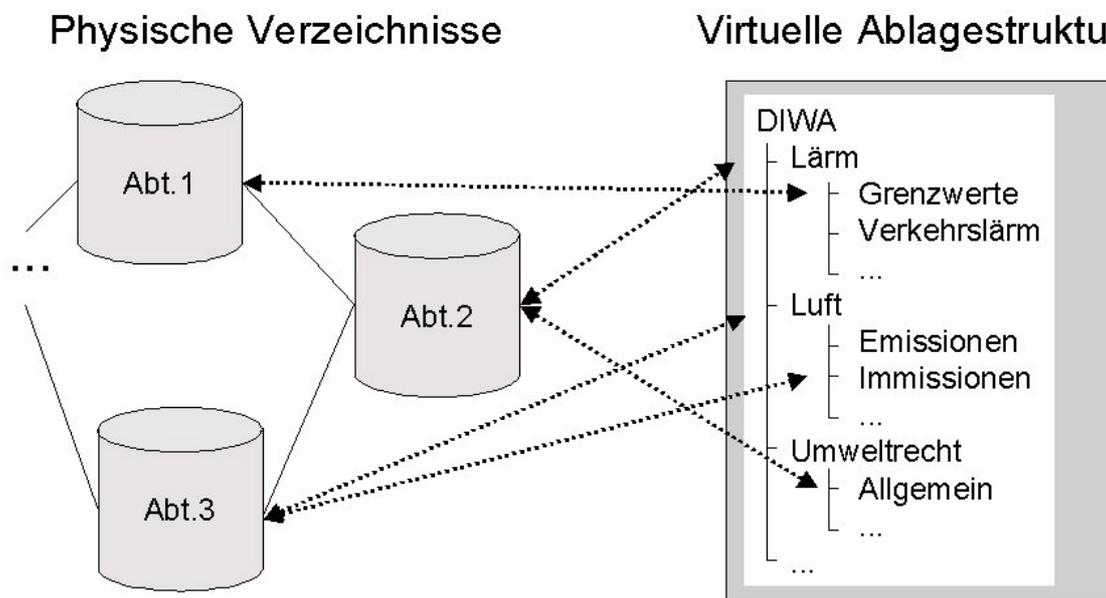


Abbildung 2: Zusammenspiel zwischen einer physischen und einer virtuellen Ablagestruktur

Beispielsweise befindet sich ein Dokumentenbestand zum Thema Luftimmissionen der Abteilung 3 (Abt. 3) physisch auf einem Rechner im Verzeichnis *D:\Luft\Immissionen*. Dieses freigegebene Netzlaufwerk kann nun direkt dem virtuellen Verzeichnis *DIWA/Luft/Immissionen* zugeordnet werden.

Bevor der prinzipielle Ablauf der Autoren- und der Recherchekomponente beschrieben wird, sollen die zentralen Komponenten des Systems kurz eingeführt

werden (Abbildung 3). Die Autoren- und die Recherchekomponente werden auf dem Klienten-Rechner des Anwenders ausgeführt. Auf dem Server befinden sich der Web Server, die Suchmaschine in Form eines Index Servers sowie der Umweltthesaurus. Der Index Server ist für die Indizierung der Dokumentenbestände des Web-Archivs, die auf verschiedenen Rechnern vorliegen können, verantwortlich.

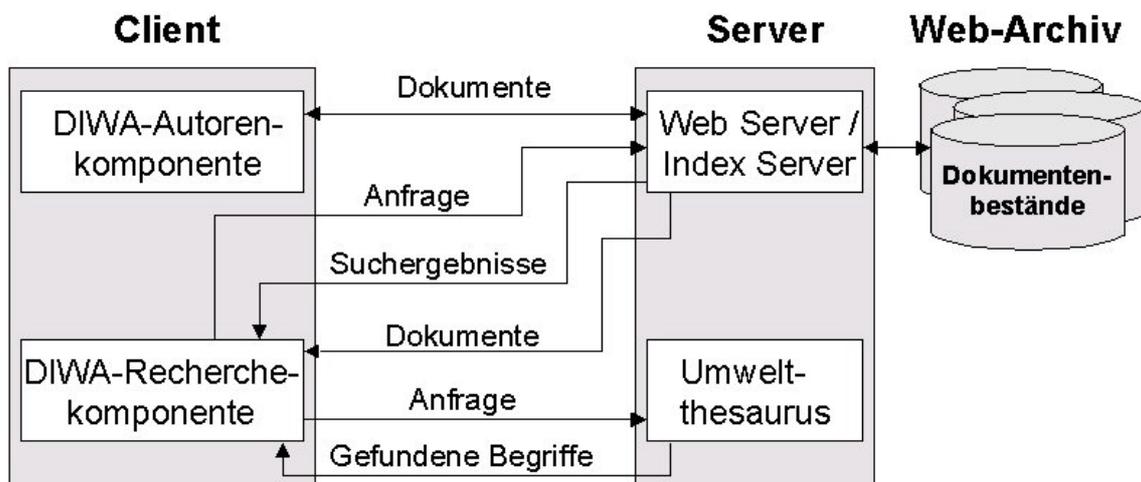
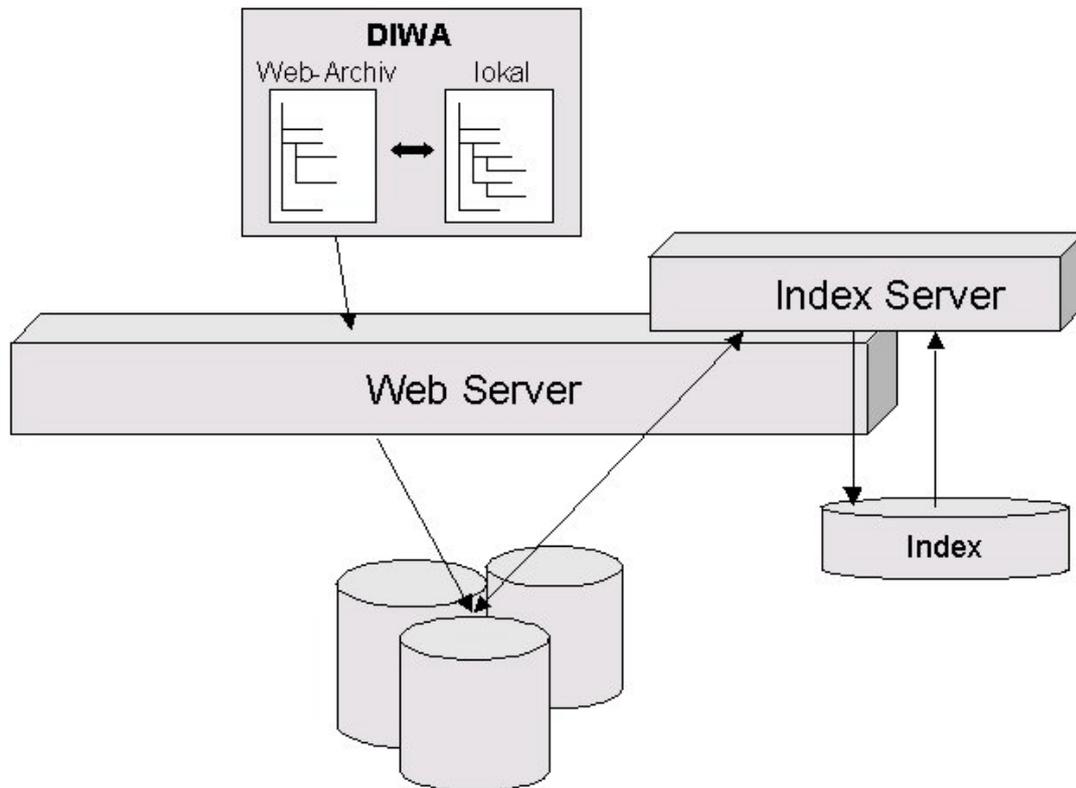


Abbildung 3: Komponenten des Web-Archivs

Über die Autorenkomponente kann der Autor in der virtuellen Ablagestruktur des Web Servers navigieren und Dokumente von seinem lokalen Rechner in den gewählten Dokumentenbestand einstellen bzw. Dokumente auf seinen lokalen Rechner herunterladen. In das Web-Archiv eingestellte Dokumente werden von der Suchmaschine in den Index aufgenommen. Abbildung 4 verdeutlicht diesen Vorgang.



Physische Dokumentenablage

Abbildung 4: Prinzipieller Ablauf der Autorenkomponente

In der Recherchekomponente formuliert der Anwender eine Abfrage, die an die Suchmaschine (den Index Server) übergeben wird. Diese sucht im Index nach den entsprechenden Ergebnissen und überträgt sie an den Anwender. Über den Link auf der Ergebnisseite kann das Dokument auf den Rechner des Anwenders heruntergeladen und dort geöffnet werden. Die Recherche kann sowohl in einer einfachen, reinen Volltextsuche als auch in einer erweiterten Suche über Metadaten erfolgen. Der beschriebene Ablauf wird in Abbildung 5 dargestellt.

Zusätzlich können zum Auffinden allgemein gültiger Umweltbegriffe für die Recherche Abfragen an einen Umweltthesaurus gestellt werden, der seine Ergebnisse wiederum dem Anwender übergibt. Bei dem verwendeten Umweltthesau

rus handelt es sich um den General European Multilingual Environment Thesaurus (GEMET) des Umweltbundesamtes.

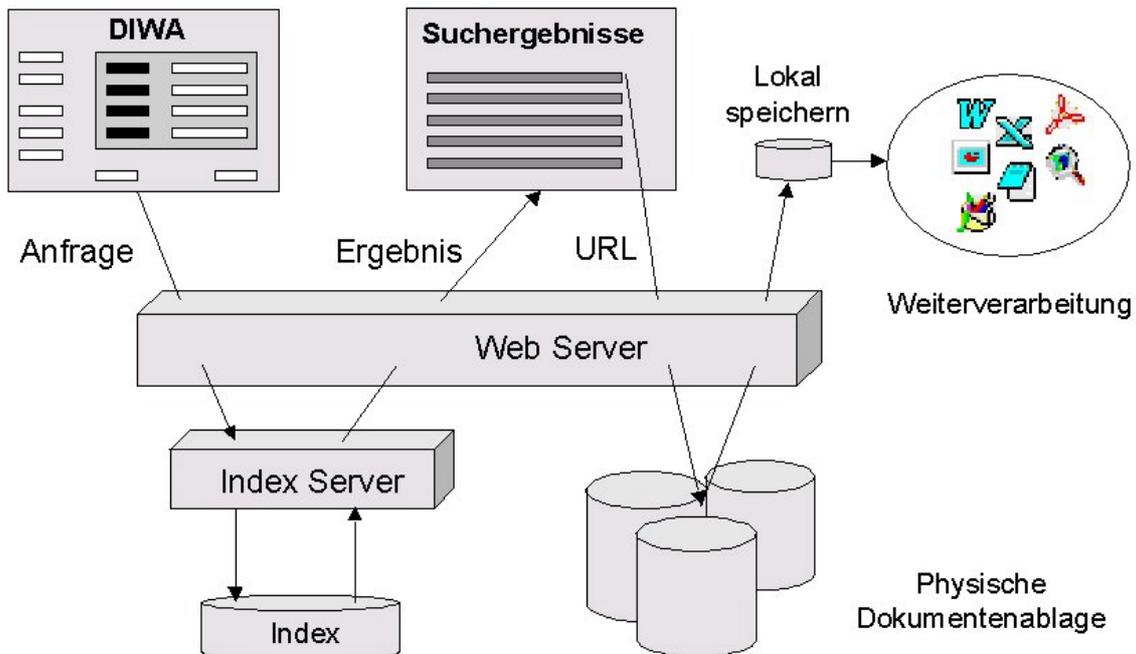


Abbildung 5: Prinzipieller Ablauf einer Recherche

3.2 Softwarearchitektur

Wie bereits in Kapitel 2 dargestellt, wird der weitgehende Einsatz kommerzieller Software unter den in Abschnitt 2.2.3 beschriebenen technischen Randbedingungen angestrebt. Auf der Server-Seite handelt es sich konkret um eine Suchmaschine in Verbindung mit einem Web Server; auf der Klienten-Seite um den Web-Browser. Zu Beginn des Projekts wurde eine umfassende Recherche nach geeigneten Suchmaschinen durchgeführt. Dabei waren Anforderungen wie Indizierung von Volltext und Dateieigenschaften, Lauffähigkeit unter den bestehenden Rahmenbedingungen, einfache Wartung, Kosten etc. zu berücksichtigen. Unter ca. 20 betrachteten Suchmaschinen (z.B. Swish-E, Excite,

Ultraseek Server, Excalibur etc.) wurde schließlich der Microsoft Index Server (MS IS) ausgewählt. Er setzt auf den Web Server Microsoft Internet Information Server (MS IIS) auf, der bereits im UVM eingesetzt wird. Diese Standardsoftware wird in den Abschnitten 3.2.1.1 und 3.2.1.2 näher betrachtet.

Da der MS IIS die Zugriffsrechte von Dokumenten nicht ausreichend berücksichtigt, wird auf Server-Seite Common Gateway Interface (CGI)- bzw. Internet Server Application Programming Interface (ISAPI)-Technologie zur Überprüfung und Einhaltung der Zugriffsrechte eingesetzt. Informationen dazu finden sich in den Abschnitten 3.2.1.2 und 3.2.1.3.

Eine zweite wesentliche Anforderung ist der Einsatz von Internettechnologien in DIWA, um das System unabhängig von lokalen Installationen (ausgenommen ein Standard-Web-Browser) nutzen zu können. Da DIWA sowohl unter dem Microsoft Internet Explorer als auch unter dem Netscape Navigator als Web-Browser lauffähig sein soll, bieten sich hier zwei Alternativen. Zum einen kommt der Einsatz von HTML-Oberflächen in Verbindung mit CGI in Betracht. Diese Variante hat den Vorteil schneller Übertragungsraten. Allerdings musste frühzeitig festgestellt werden, dass die technischen Möglichkeiten zum Entwickeln der DIWA-Oberfläche und –funktionalität beschränkt sind. So konnte z.B. das gewünschte Wechseln der Suchformulare in derselben Benutzeroberfläche nicht realisiert werden. Als Alternative dazu bietet sich die Internet-Programmiersprache Java an. Auf Grund der Möglichkeiten, komplexe Funktionen beliebig hinter der Benutzeroberfläche zu verbergen, wurde Java für die Realisierung von DIWA gewählt. Grundlagen, Vor- und Nachteile von Java werden in Abschnitt 3.2.1.4 beschrieben.

Abschließend wird die Software-Architektur von DIWA aufgezeichnet und erläutert.

3.2.1 Verwendete Software und Technologien

In diesem Kapitel werden die in DIWA eingesetzte Software MS IIS und MS IS sowie die verwendeten Technologien CGI und Java beschrieben.

3.2.1.1 Microsoft Index Server

Wie bereits erwähnt, wurden bei der Recherche nach Suchmaschinen verschiedene Aspekte berücksichtigt. Die Wahl des MS IS erfolgte aus mehreren Gründen:

- Der MS IS ist als einzige der untersuchten Suchmaschinen in der Lage, Dateieigenschaften von Microsoft Office- und PDF-Dokumenten zu indizieren. Da diese Eigenschaften in DIWA als Metainformationen genutzt werden sollen, ist bereits dieses Kriterium entscheidend. Die Indizierung von HTML-Meta-Tags ist auch in anderen Suchmaschinen möglich.
- Neben der Indizierung von Volltext und Dateieigenschaften werden auch die Zugriffsrechte der einzelnen Dokumente in den Index aufgenommen. Diese Möglichkeit besteht real allerdings nur für Dokumentenbestände, die auf demselben Rechner liegen wie der MS IS.
- Das Einbinden von Dokumentenbeständen, die auf Unix-Rechnern vorliegen, ist möglich, wenn sie über das UNC (Universal Naming Convention)-Format verfügbar gemacht werden.
- Der MS IS ist in der Lage, Dokumentenbestände inkrementell zu indizieren. D.h. er indiziert nicht alle Dokumente in festgelegten Zyklen, sondern nur solche Dokumente, die verändert, gelöscht oder neu eingestellt wurden. Hinzu kommt, dass der MS IS in einer reinen Windows NT-Umgebung eine Information vom Betriebssystem bekommt, wenn ein Dokument verändert wurde. Insgesamt wird dadurch einerseits der Server entlastet und zum anderen der Index fast zeitgleich aktualisiert.

- Die Wartung des MS IS ist relativ einfach. Nach der ersten Einrichtung fallen kaum Arbeiten an. Ausserdem wird eine komfortable Benutzeroberfläche angeboten, während andere Suchmaschinen häufig Kommandozeilen-orientiert arbeiten.
- Der MS IS wird zusammen mit dem MS IIS im Windows NT Option Pack angeboten. Da in DIWA als Server-Plattform Windows NT verwendet werden soll, ist der Einsatz von MS IS und MS IIS optimal. Es kann davon ausgegangen werden, dass die einzelnen Komponenten auch bei neuen Versionen untereinander kompatibel sein werden. Das Windows NT Option Pack ist kostenlos erhältlich, womit sich der finanzielle Aufwand beim Einsatz von DIWA wesentlich reduziert.

Andere Suchmaschinen bieten ebenfalls hohe Funktionalität, verfügen aber nicht über diese Summe an für den Einsatz in DIWA notwendigen oder vorteilhaften Eigenschaften. So kann Swish-E (Swish-E 1999) nicht inkrementell indizieren, was für die Auslastung des Servers bei dynamischen Dokumentenbeständen eine große Rolle spielt. Der Ultraseek Server (Infoseek 1999) kann zwar eine umfangreichere Menge an Dokumentformaten indizieren, aber nur über den Volltext und nicht über Dokumenteigenschaften. Die Suchmaschinen ASTAWare Searchkey Pro (ASTAWare 1999) und Alkaline (Alkaline 1999) indizieren zwar Dokumenteigenschaften, aber nur in HTML-Dokumenten, was dem Anspruch der Verfügbarmachung heterogener Dokumentenbestände widerspricht.

Der MS IS ist die Lösung von Microsoft, um vom MS IIS verwaltete Inhalte zu durchsuchen und zu indizieren. Klienten können Abfragen formulieren, indem sie über einen beliebigen Web-Browser die Felder eines einfachen Abfrageformulars ausfüllen. Der Web Server leitet das Abfrageformular an das Abfragemodul weiter, welches die zugehörigen Dokumente ermittelt und das Suchergebnis in Form einer Web-Seite an den Klienten zurückgibt.

Abfrageformulare

Anwender übermitteln ihre Abfragen, indem sie Felder in einem Formular ausfüllen. Solche Formulare können von Administratoren im Standard-HTML-Format, in JavaScript oder in Java erstellt werden. Das Formular kann so angepasst werden, dass Anwender nach Inhalten oder/und Dokumenteigenschaften suchen können.

Eine Abfrage umfasst folgende Grundmerkmale:

- Gültigkeitsbereich,
- Einschränkung,
- Ergebnismenge.

Gültigkeitsbereich

Der Gültigkeitsbereich einer Abfrage bezeichnet eine zu durchsuchende Teilmenge des Dokumentenbestands. In der Regel werden Gültigkeitsbereiche durch einen virtuellen Verzeichnispfad angegeben. Der MS IS indiziert Dokumente auf der Grundlage ihrer Zugehörigkeit zur virtuellen Ablagestruktur. Der Administrator kann alle virtuellen Ablageverzeichnisse auf einem Server indizieren oder lediglich bestimmte virtuelle Ablageverzeichnisse für die Indizierung auswählen. Abfragen können mehrere oder auch einzelne virtuelle Verzeichnisse umfassen.

Einschränkung

Folgende Dokumenttypen können durchsucht werden:

- HTML,
- Microsoft Word,
- Microsoft Excel,
- Microsoft PowerPoint,

- PDF (Filter wird von Adobe angeboten),
- XML (Suche nach Dateieigenschaften derzeit noch nicht möglich),
- reine Textdokumente.

Die Liste der unterstützten Dokumenttypen kann durch einen neuen „Inhaltsfilter“ erweitert werden, welcher ein herstellereigenes Dokumentformat liest und Wörter als reinen Text zurückgibt, der vom Index Server indiziert wird.

Möglichkeiten zur Präzisierung der Recherche sind:

- Suchen nach Begriffen, Teilbegriffen (Verwendung von Platzhaltern), Wortstämmen im Volltext und den Dateieigenschaften,
- Volltextabfragen, bei denen Wortfolgen oder ganze Sätze angegeben werden können,
- Verknüpfung der Begriffe mit Booleschen Operatoren,
- Verwendung von Wichtungen für einzelne Begriffe.

Unterstützung mehrerer Sprachen

Die meisten Web-Seiten werden heute in Englisch verfasst. Da der MS IIS auch Dokumente mehrerer Sprachen verwalten kann, ist die mehrsprachige Indizierung und Abfrage ein Standard-Leistungsmerkmal des MS IS. Beim Entwurf des Abfragesystems wurden verschiedenste Sprachen berücksichtigt. Es ist modular aufgebaut und kann dynamisch sprachspezifische Dienstprogramme hinzuladen und wieder entfernen. Bei diesen sprachbezogenen Komponenten handelt es sich um Module für die Worttrennung und die Ermittlung von Wortstämmen.

Ergebnismengen

Ergebnismengen können formatiert dargestellt werden:

- Festlegung der maximalen Anzahl anzuzeigender Treffer,

- Festlegung der Anzahl der Treffer, welche pro Seite angezeigt werden sollen,
- Sortieren der Treffer nach bestimmten Kriterien (Datum, Größe, Autor etc.).

Erstellen von Abfragen

Das Durchführen von Abfragen über den MS IS ist ein komplexer Vorgang, der eng an den MS IIS gekoppelt ist. Der Internet Database Connector (IDC), eine Komponente des MS IIS, wandelt eine Abfrage aus einem Formular in ein Format um, das für eine Suche über Open Database Connectivity (ODBC) verwendet werden kann. Die Abfrage erfolgt in ODBC, da der Index intern im MS IS in Form einer Datenbank abgelegt ist. Der Index verarbeitet eine Abfrage und gibt ein Ergebnis zurück, welches vom IDC in eine HTML-Seite umgewandelt und am Bildschirm des Anwenders angezeigt wird. Die Ergebnisseite enthält im Allgemeinen Uniform Resource Locators (URLs) zu den gefundenen Dokumenten sowie Eigenschaften dieser Dokumente (z.B. Dateigröße, Autor, Kurzfassung des Inhalts usw.).

Eine technische Beschreibung der Entwicklung spezieller Dateien zur Abfrage und Aufbereitung der Ergebnisse ist im Anhang 1 Abschnitt 5.2 zu finden.

3.2.1.2 Microsoft Internet Information Server

Im Folgenden wird der Microsoft Internet Information Server (MS IIS) vorgestellt, wobei in erster Linie auf die Dienste eingegangen wird, die in DIWA genutzt werden.

Der MS IIS ist ein Web Server, mit dem Informationen in einem Intranet oder im Internet veröffentlicht werden können. Der MS IIS überträgt Daten mit Hilfe des Protokolls Hypertext Transfer Protocol (HTTP). Darüber hinaus läßt sich der MS IIS für die Bereitstellung der Dienste File Transfer Protocol (FTP) und Gopher konfigurieren. Der FTP-Dienst erlaubt Benutzern die Dateiübertragung von

und zu einem Web Server. Der Gopher-Dienst verwendet ein menügesteuertes Protokoll für das Auffinden von Dokumenten. Dieser Dienst spielt für DIWA jedoch keine Rolle, sodass hierauf nicht weiter eingegangen wird.

Der WWW-Dienst

Im WWW-Dienst (World Wide Web) werden in sogenannten Informationsverzeichnissen die Informationen gespeichert, welche über das Internet/Intranet abgerufen werden können. Die Art der Informationen spielt dabei kaum eine Rolle; statische HTML-Dateien, Skripte, ausführbare Programme oder eben die Dokumente des Web-Archivs werden hier verwaltet.

Der MS IIS unterscheidet dabei zwei Arten von Informationsverzeichnissen:

- Das sogenannte Basisverzeichnis wird einem physischen Verzeichnis auf dem Server zugeordnet. Alle in dem angegebenen Verzeichnis sowie in sämtlichen Unterverzeichnissen gespeicherten Informationen stehen den Klienten zur Verfügung.
- Sogenannte virtuelle Verzeichnisse verweisen auf Verzeichnisse, welche sich entweder auf einem lokalen Laufwerk des Servers befinden oder über das Netzwerk verteilt sind und mit einem UNC-Pfadnamen sowie einem Benutzernamen und einem Kennwort für die Zugriffsberechtigung angegeben werden. An die virtuellen Verzeichnissen wird ein Alias vergeben. Der Alias ist der Name, der von den Klienten verwendet wird, um auf Informationen in dem virtuellen Verzeichnis zuzugreifen.

Zur Vereinfachung für den Anwender stellen die Dienste des MS IIS die gesamten virtuellen und physischen Verzeichnisse nach aussen als eine einzige Verzeichnisstruktur dar. Das Basisverzeichnis ist das Stammverzeichnis dieser virtuellen Verzeichnisstruktur. Jedes virtuelle Verzeichnis wird so adressiert, als handle es sich um ein Unterverzeichnis zum Basisverzeichnis (Abbildung 6).

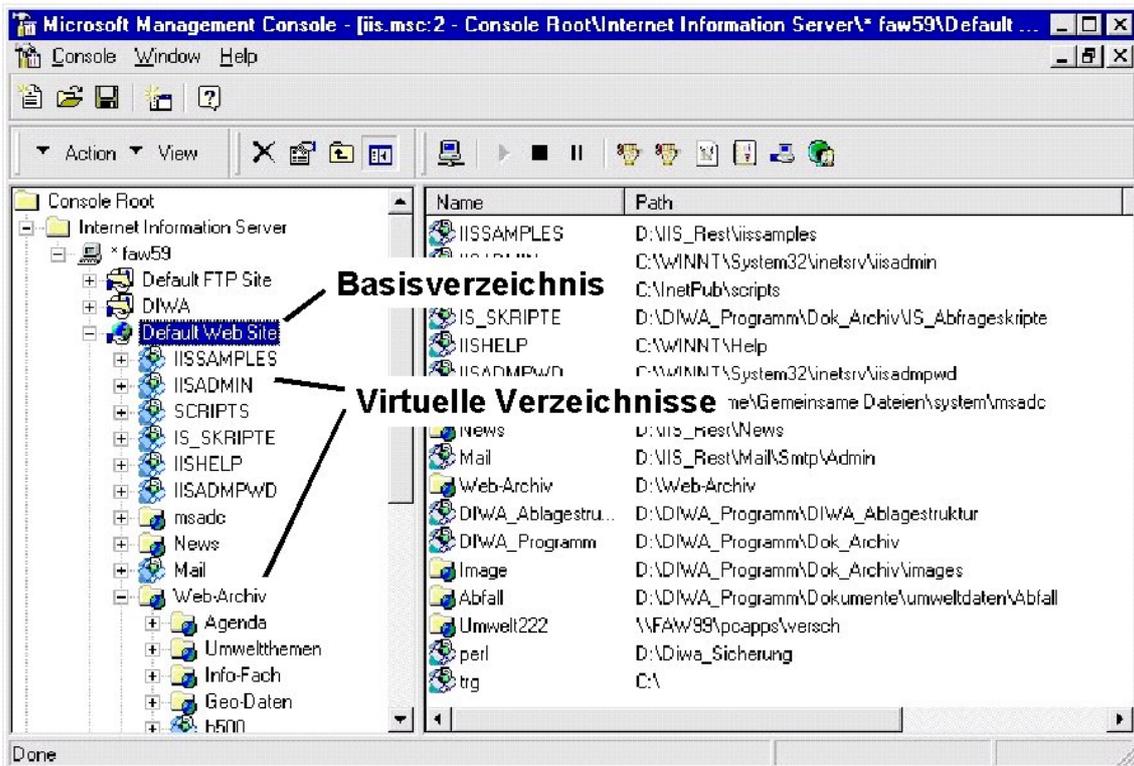


Abbildung 6: Aufbau der Verzeichnisse im WWW-Dienst

Virtuelle Verzeichnisse erscheinen nicht in Verzeichnislisten (für den WWW-Dienst auch Verzeichnissuche genannt). Um auf ein virtuelles Verzeichnis zugreifen zu können, müssen die Anwender den Alias des virtuellen Verzeichnisses kennen und die Internet-Adresse in den Browser eingeben. Für den WWW-Dienst können aber auch Verknüpfungen in HTML-Seiten erstellt werden.

Der FTP-Dienst

FTP (FTP 1999) war eines der ersten in TCP/IP-Netzwerken und dem Internet verwendeten Protokolle. Es dient dazu, innerhalb eines Netzwerks Dateien von einem Computer auf einen anderen zu übertragen. Dabei eignet sich FTP besonders zur Dateiübertragung zwischen unterschiedlichen Plattformen, z.B. zwischen Unix und MS Windows.

Unter den Internet-Diensten ist FTP das einzige Protokoll, mit dem Dateien direkt von einem Klienten auf einen Server kopiert werden können. Auf der anderen Seite können den Benutzern mit Hilfe von FTP Dateien jedes Formats zur Verfügung gestellt werden.

ISAPI

ISAPI (Internet Server Application Programming Interface) (ISAPI 1999) wurde von Microsoft als Alternative zur CGI-Technologie (siehe Abschnitt 3.2.1.3) für die Anwendung im MS IIS entwickelt. Sogenannte ISAPI-Extensions können wie CGI-Skripte von Klienten-Seite auf dem Server gestartet und dort ausgeführt werden. Eine besondere Funktionalität bietet sich durch die ISAPI-Filter. Diese Filter werden vom MS IIS selbst gestartet, sobald bestimmte Ereignisse eintreten. Bei diesen Ereignissen kann es sich z.B. um eine Anfrage an den Web Server, um das Senden von Daten, um das Beenden einer Sitzung etc. handeln. Somit kann mit Hilfe eines ISAPI-Filters auf bestimmte Ereignisse reagiert bzw. können diese ausgewertet werden. Auch der MS IS arbeitet intern mit ISAPI-Technologie.

3.2.1.3 CGI und Perl

Bei CGI-Programmen (Common Gateway Interface) (CGI 1999) handelt es sich um Programme, die auf Grund eines HTTP-Aufrufs vom Klienten auf dem Server ausgeführt werden und die Ergebnisse dieses Programms über HTTP zum Klienten übertragen. Für die Realisierung dieser Server-Programme wird keine bestimmte Programmiersprache vorgeschrieben. Häufig wird die Skriptsprache Perl verwendet (Perl 1999). Vorteilhaft für den Einsatz in DIWA ist vor allem, dass Perl zum einen sehr gut mit Betriebssystemen kommunizieren kann und zum anderen für die Ver- und Bearbeitung von Zeichenketten prädestiniert ist. Wie in Abschnitt 3.2.1.2 beschrieben, verwendet der MS IIS für den Zugriff auf Dokumentenbestände, die nicht auf demselben Server liegen (im Folgenden als *entfernte Dokumentenbestände* bezeichnet), einen bestimmten Benutzer-

Account. Dadurch werden alle Anwender in die Lage versetzt, Dokumente über diesen Account zu erreichen - unabhängig davon, ob sie persönlich Leserechte besitzen oder nicht. In Abbildung 7 wird dieser Sachverhalt dargestellt. Der Anwender mit dem Account 99 besitzt eigentlich keine Zugriffsrechte auf den Dokumentenbestand B. Wenn er aber mit der Internet-Adresse des Dokuments über den MS IIS das Dokument herunterladen möchte, ist dies möglich, da der MS IIS mit dem leseberechtigten Account 1 auf das Dokument zugreift.

Ein ähnliches Problem tritt bei der Nutzung des FTP-Dienstes des MS IIS auf. Auch dort kann es bei der Anbindung von entfernten Dokumentenbeständen dazu kommen, dass Anwender unbefugt auf Daten zugreifen können.

Zur Entwicklung eines „Wächters“ in DIWA eignet sich Perl, da es in der Lage ist, die Zugriffsrechte von Dokumenten zu lesen und mit denen des Anwenders abzugleichen.

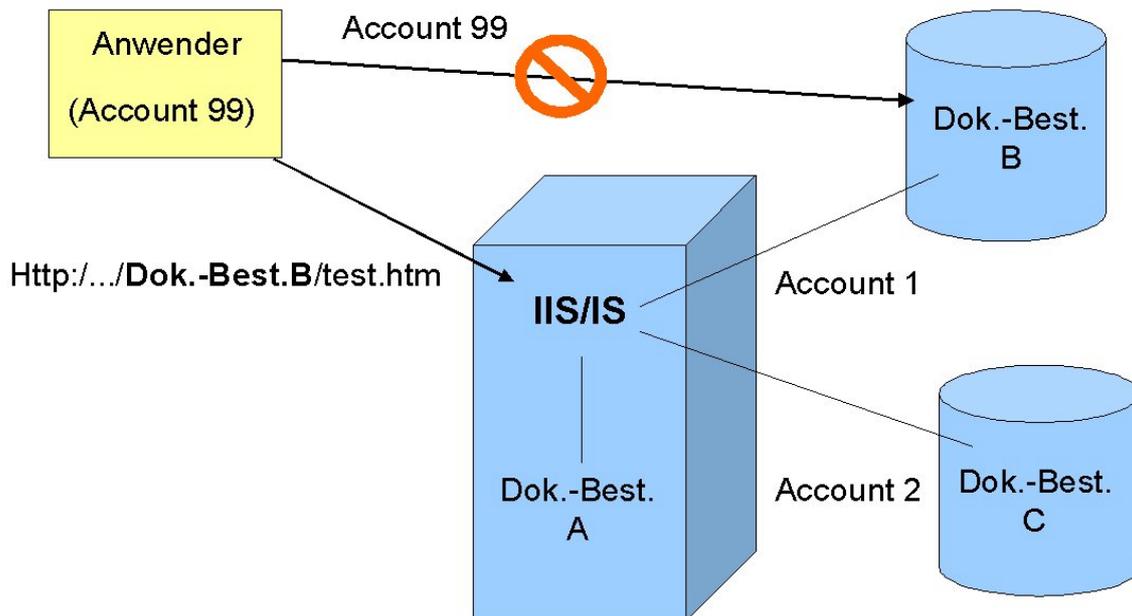


Abbildung 7: Verletzung der Zugriffsrechte entfernter Dokumentenbestände

Zu Beginn des Projekts wurde FTP für das Einstellen von Dokumenten in das Web-Archiv verwendet. Auf Grund der Verletzung der Zugriffsrechte wird auch die Autorenkomponente auf CGI umgestellt.

Um auch den Zugriff auf Dokumente über eine URL ausserhalb von DIWA zu verhindern, muss ein ISAPI-Filter (siehe Abschnitt 3.2.1.2) entwickelt werden, der bei jedem Aufruf einer Web-Seite prüft, ob das Web-Archiv betroffen ist oder nicht.

3.2.1.4 Java

Mit dem Aufkommen und der rasanten Verbreitung des WWW im Internet erkannte man sehr bald, dass eine architekturunabhängige Sprache wie Java ideal für die Programmierung im Internet geeignet ist, da ein Programm auf allen am Internet angeschlossenen Rechnern lauffähig sein würde.

Inzwischen haben Java-Entwicklungen nicht nur bei der Firma Sun, sondern auch bei Lizenznehmern wie IBM, SGI und Oracle einen hohen und weiter zunehmenden Stellenwert erreicht.

Mit Java können eigenständige Programme entwickelt werden oder auch sogenannte Applets. Bei Applets handelt es sich um Programmfragmente, die über das Internet geladen, in einem Standard-Web-Browser interpretiert und auf dem lokalen Rechner ausgeführt werden. Genau diese Technik ist für den Einsatz in DIWA geeignet, da das Programm von einem Server auf beliebige Anwender-Rechner geladen werden kann und somit der Anforderung der Verwendung von Internettechnologien genügt.

Bei der Verwendung von Java Applets werden Server und Netz entlastet, indem z.B. statt serverseitig erstellter Grafiken lediglich die erforderlichen Java-Applets und Daten auf den lokalen Rechner übertragen und Grafiken lokal aktualisiert werden.

Als problematisch bei der Nutzung von Java Applets stellen sich allerdings die Sicherheitsrestriktionen in Java dar, die auch als Sandbox-Prinzip bezeichnet werden. Auf Grund dieses Prinzips sind Java Applets u.a. folgende Funktionen auf dem Rechner des Anwenders nicht gestattet:

- lokale Dateien lesen,
- lokale Dateien schreiben,
- drucken,
- Netzwerkverbindungen können nur zu dem Server aufgebaut werden, von dem das Java Applet gestartet wurde, nicht aber zu anderen Servern,
- andere Anwendungen starten,
- anderen Anwendungen Nachrichten senden,

- Benutzer-Eigenschaften abfragen.

Besonders das Lesen lokaler Dateien und das Schreiben von Dateien auf verschiedenen Servern werden für das Einstellen von Dokumenten in das Web-Archiv, also für die Autorenkomponente, benötigt. Einen Ausweg aus dieser Situation bietet das Signieren der Java Applets (Pursche et al. 1999).

Java bietet mit der Java Database Connectivity (JDBC) Möglichkeiten, auf beliebige relationale Datenbanken zuzugreifen. Dieser Aspekt spielt für die Verfügbarmachung des Umweltthesaurus eine entscheidende Rolle, da alle verwendbaren Thesauri in Form von Datenbanken vorliegen. JDBC setzt auf dem X/OPEN SQL-Call-Level-Interface (CLI) auf und besitzt damit die gleiche Basis wie die ODBC-Schnittstelle.

Problematisch ist, dass mit der reinen Anwendung von JDBC auf Grund des Sandbox-Prinzips nur auf den Web Server zugegriffen werden kann, von dem das Java Applet, welches auf die Datenbank zugreift, gestartet wurde (Two-Tier-Architektur, Abbildung 8).

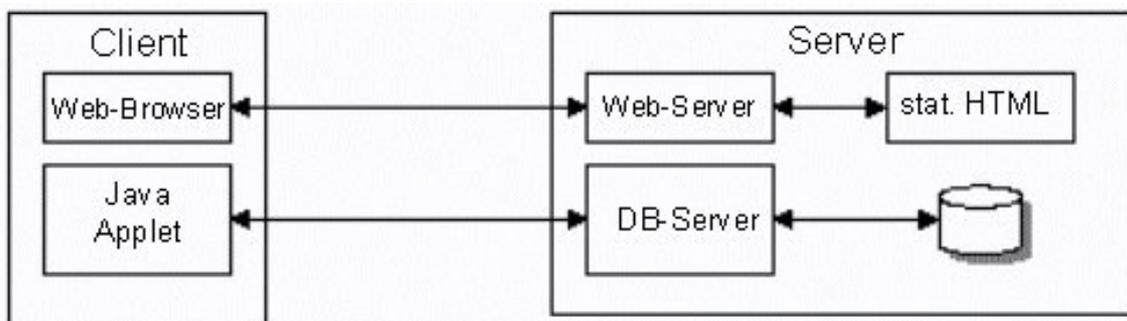


Abbildung 8: Two-Tier-Architektur

Werden Web- und Datenbank-Server auf verschiedenen Maschinen betrieben, kann das Java Applet nicht mehr ohne Signierung auf die Datenbank, im konkreten Fall den Umweltthesaurus, zugreifen. Auch hierfür bietet Java eine Lösung in Form der Remote Method Invocation (RMI) an, welche Java die Kom

munikation in verteilten Systemen erlaubt. Mittels dieser Technologie kann auf dem Web Server ein Gateway eingerichtet werden, welches als Datenbank-Klient die Abfrage der Datenbank stellvertretend für das Java Applet übernimmt und die Ergebnisse anschließend zurücksendet (Three-Tier-Architektur, Abbildung 9).

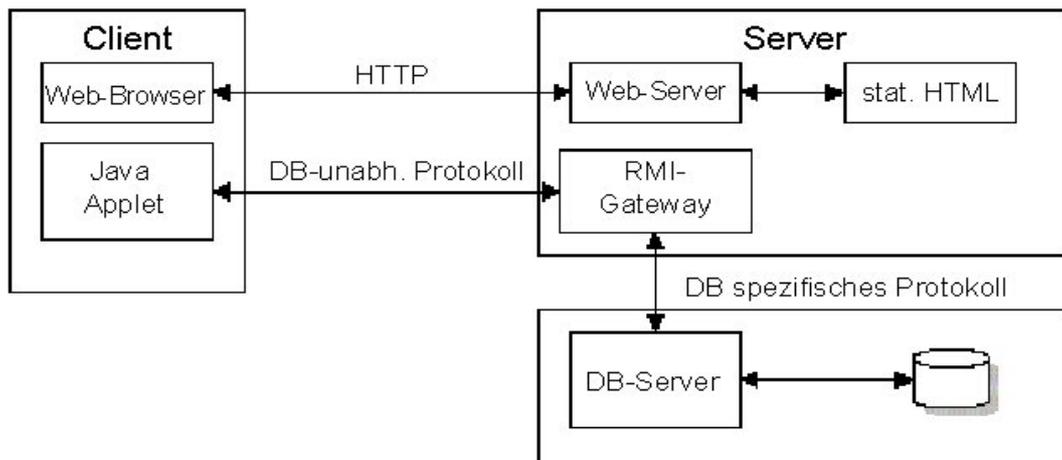


Abbildung 9: Three-Tier-Architektur

3.2.2 Softwarearchitektur für DIWA

Zentrales Element von DIWA ist die Basissoftware MS IIS / MS IS. Über den MS IIS erfolgen sämtliche Rechercheabfragen an den MS IS, die Rückgabe der Ergebnisse, der Zugriff auf die Dokumente und das Einstellen von Dokumenten in das Web-Archiv.

Die Autorenkomponente, als Java Applet realisiert, erlaubt das Einstellen von Dokumenten vom lokalen Rechner des Autors über die virtuelle Ablagestruktur in das Web-Archiv. Die Zuordnung der virtuellen Verzeichnisse zu den zugehörigen physischen Verzeichnissen erfolgt über den MS IIS. Neu eingestellte Dokumente werden dann vom MS IS in den Index aufgenommen. Die Vorgänge sind in Abbildung 10 dargestellt.

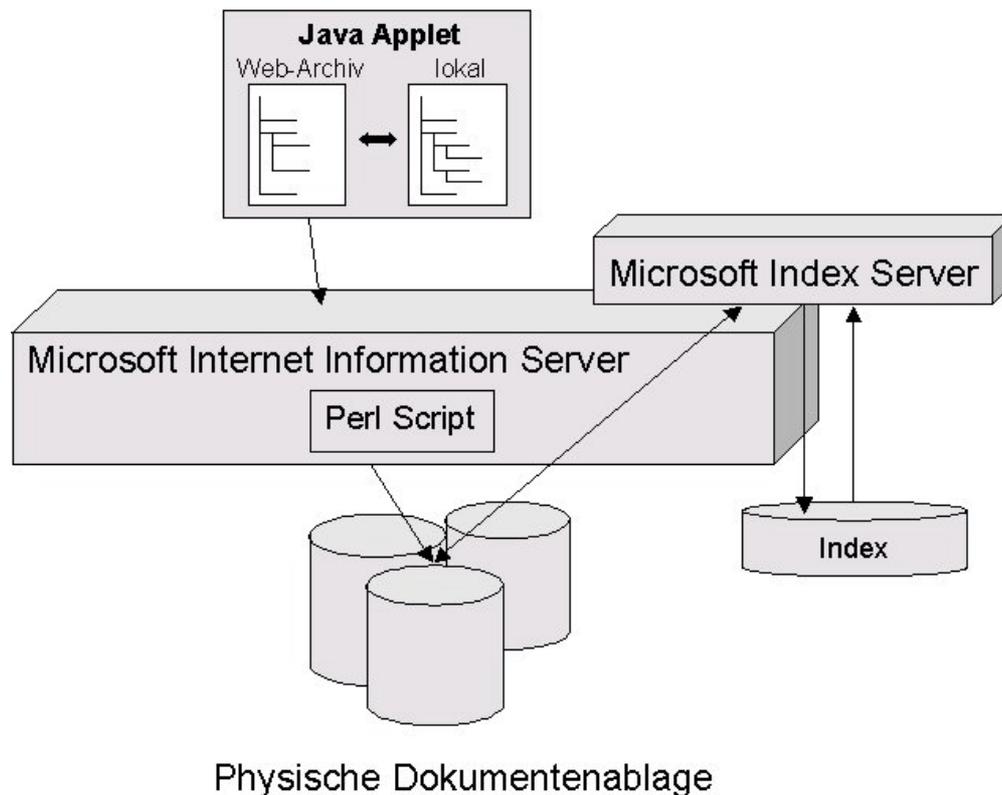


Abbildung 10: Software der Autorenkomponente

Die Recherchekomponente, ebenfalls als Java Applet realisiert, stellt eine komfortable Benutzeroberfläche für die Formulierung und automatische Generierung von Suchabfragen dar. Sie ist zusätzlich für die Festlegung der Parameter zur Formatierung der Ergebnisseiten zuständig. Die Suche im Index selbst wird vom MS IS ausgeführt. Der Ablauf ist in Abbildung 11 aufgezeichnet.

Zusätzlich wird in der Recherchekomponente ein Umweltthesaurus angeboten, der in Form einer Datenbank vorliegt. Der Zugriff auf den Thesaurus erfolgt über einen RMI-Server, der mittels JDBC Abfragen an die Datenbank stellt. Mit Hilfe des RMI-Servers ist es möglich, sowohl eine Two-Tier- als auch eine Three-Tier-Architektur zu realisieren. D.h. die Thesaurus-Datenbank kann auf demselben, aber auch einem anderen Server liegen wie das DIWA-Programm.

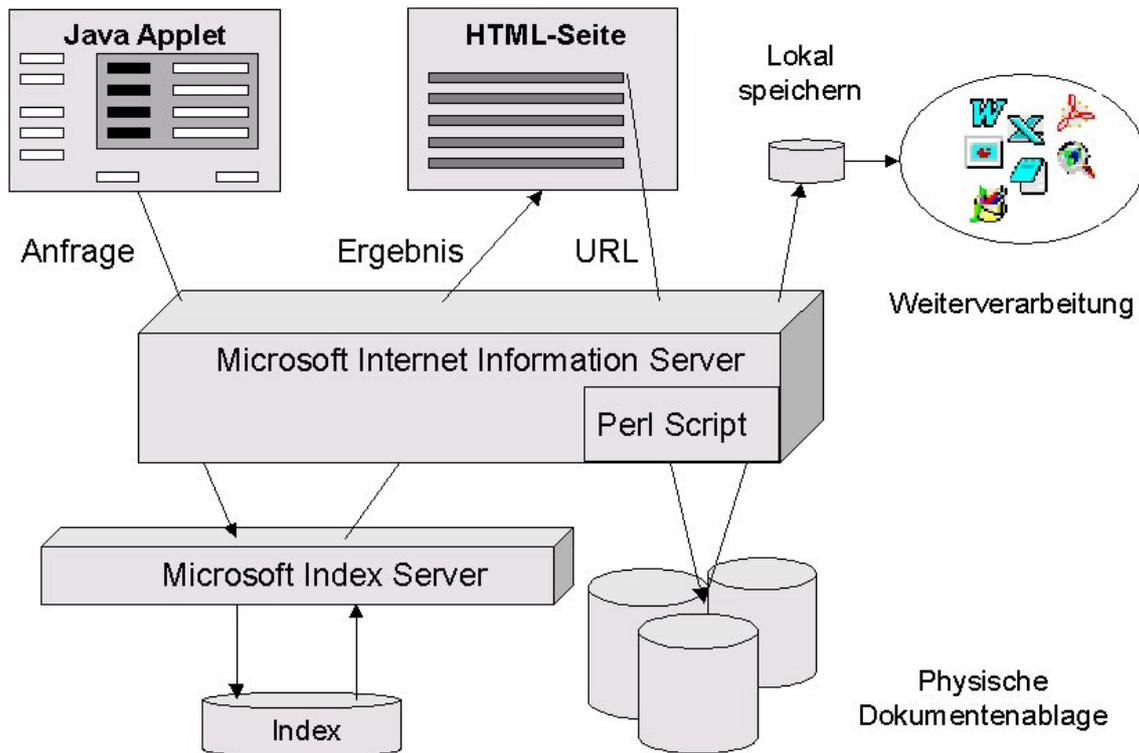


Abbildung 11: Software der Recherche Komponente

Auf Grund der in Abschnitt 3.2.1.3 beschriebenen Probleme werden beim Zugriff auf Dokumente bzw. beim Einstellen von Dokumenten die Zugriffsrechte über „Wächter“ in Form von Perl-Skripten überprüft. Zusätzlich wurde ein ISAPI-Filter entwickelt, der bei jedem Aufruf einer Web-Seite prüft, ob Dokumente des Web-Archivs angesprochen werden. Diese Aufrufe werden direkt abgewiesen. Dokumente des Web-Archivs dürfen nur über die Perl-Skripte angesprochen werden, wie es in DIWA automatisch geschieht. Nur durch dieses Zusammenspiel kann die Sicherheit der Dokumente des Web-Archivs gewährleistet werden.

Die Gesamtarchitektur von DIWA wird in Abbildung 12 dargestellt.

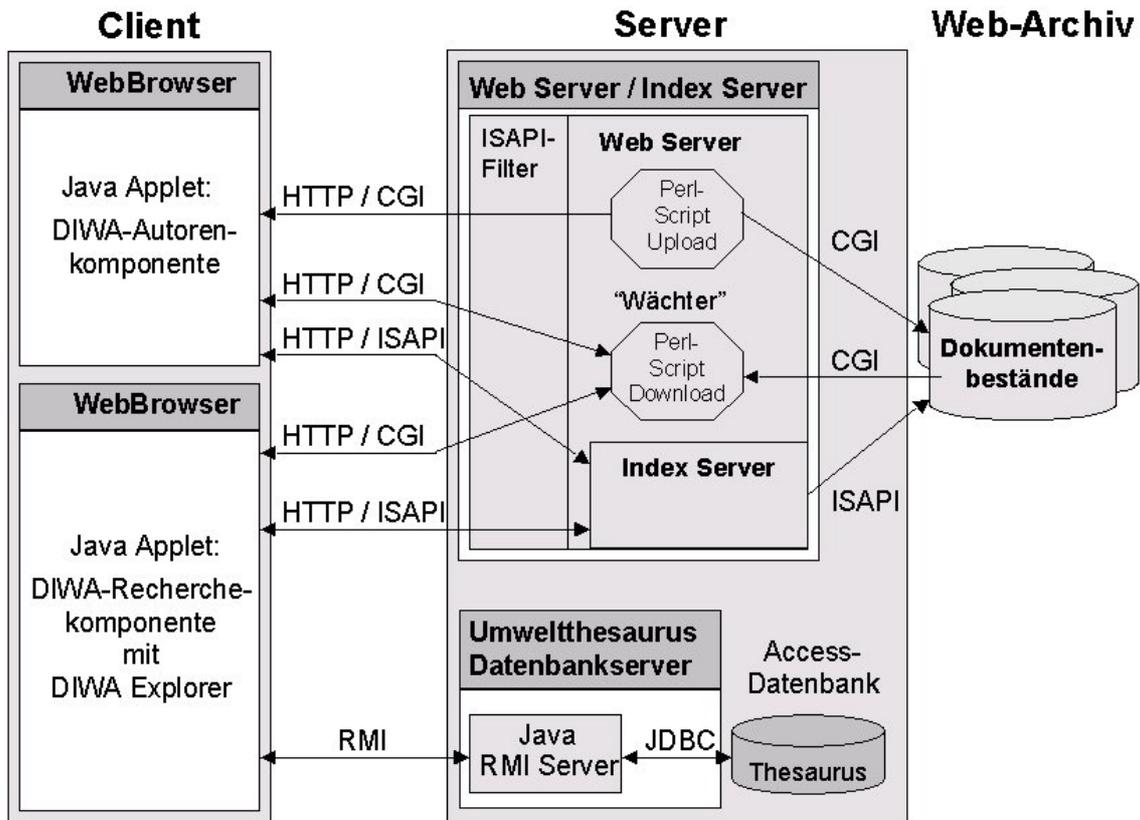


Abbildung 12: Softwarearchitektur DIWA

4 DIWA-Funktionalität für weitere Systeme

Die Funktionalität von DIWA wird auch in anderen Systemen nutzbar gemacht. Beispiele solcher Systeme sind die Fachanwendung der Zentralen Stelle für die Vollzugsunterstützung beim Staatlichen Gewerbeaufsichtsamt Stuttgart (FA-ZSV) und der Zentrale Fachdienst Wasser, Boden, Abfall, Altlasten der LfU (ZFD) mit dem System DRS (Document Retrieval System).

Im Rahmen von DIWA wird die FA-ZSV erneuert, welche die Pflege der Vorschriftensammlung der Gewerbeaufsicht (VSGA) beinhaltet. Dabei handelt es sich um ein Datenbanksystem, in welchem die Fundstellen und Änderungen von Vorschriften verwaltet werden. Ein Teil der Vorschriften liegt in digitaler Form, ein anderer als Lose-Blatt-Sammlung vor. Da das bisher verwendete System nicht Jahr-2000-fähig ist, entsteht in der ersten Phase der Realisierung ein hoher Zeitdruck, um den laufenden Betrieb der Datenpflege ohne Unterbrechung gewährleisten zu können. Die Kopplung an DIWA wird schrittweise so durchgeführt, dass zuerst die digital vorliegenden Dokumente über die Recherchekomponente verfügbar gemacht werden. Weiterhin kann eine Autorenkomponente eingesetzt werden, um neue digitale Dokumente in den Bestand einzutragen.

4.1 Anforderungen in FA-ZSV

Im Teilprojekt FA-ZSV soll ein mehrplatzfähiges System entwickelt werden, mit dem die Fundstellen der Vorschriften und Erlasse der VSGA verwaltet und weiter verarbeitet werden können. Für das Teilprojekt FA-ZSV existieren folgende Anforderungen:

- In der ersten Phase sind die Daten, die bisher in einer nicht Jahr-2000-fähigen Datenbank vorliegen, zu sichern. Zielsystem ist das Datenbankmanagementsystem Oracle.
- Im Zielsystem Oracle sind ein vorgegebenes Datenbank-Schema anzulegen und die übernommenen Daten anzupassen.
- Das System soll mit Hilfe moderner Internettechnologien mehrplatzfähig gestaltet werden.
- Mit dem System soll die weitgehend automatisierte Erstellung von Arbeitsanweisungen sowie Monats- und Jahresberichten (Vorgangsbearbeitung) ermöglicht werden. Für die Gestaltung der Berichte sind vorgegebene Formate einzuhalten.
- Das Eintragen von Daten in die Datenbank sowie das Recherchieren nach Datensätzen in der Datenbank sollen von denselben Benutzeroberflächen aus möglich sein.
- Digital vorliegende Dokumente sollen über eine Recherchefunktionalität verfügbar gemacht werden.

4.2 FA-ZSV-Architektur und Realisierung

Die Daten der FA-ZSV liegen in einer Datenbank auf demselben Server vor, von dem aus die Anwendung gestartet wird (Two-Tier-Architektur). Für das Eintragen von und die Recherche nach Datensätzen wird ein und dieselbe Be

nutzeroberfläche verwendet. Die Überprüfung von Schreib- und Leserechten erfolgt direkt in der Datenbank. Auf Grund dieser Voraussetzungen ergibt sich eine einfache Systemarchitektur wie in Abbildung 13 dargestellt.

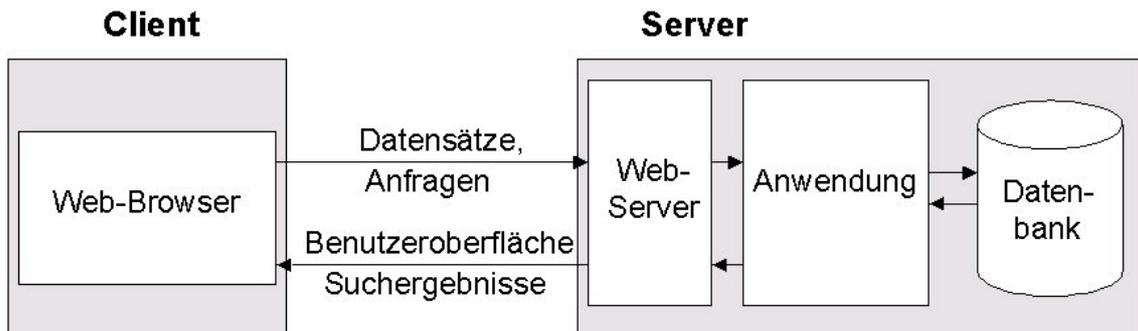


Abbildung 13: Systemarchitektur der FA-ZSV

Der Anwender startet die Anwendung in seinem Standard-Web-Browser. Für den Aufruf der Anwendung wird ein Web Server benötigt, der die entsprechende Oberfläche an den Anwender sendet. Alle Abfragen an die Datenbank bzw. in die Datenbank einzutragende Daten werden über den Web Server an die Anwendung gesendet. Die Anwendung kommuniziert direkt mit der Datenbank. Ergebnisdaten werden auf dem umgekehrten Weg an den Anwender übertragen.

4.2.1 Verwendete Technologien

Für die Übernahme der Daten aus dem bisherigen Clipper-Programm wird das Datenbanksystem MS Access genutzt, da die Datenübertragung zwischen diesen beiden Systemen problemlos möglich ist. Des Weiteren lässt sich die vorgegebene Datenstruktur technisch einfacher in MS Access umsetzen als im Zielsystem Oracle. Die Übernahme der Datenstruktur mit den Daten von MS Access zu Oracle ist problemlos möglich.

Für die Benutzerschnittstelle zum Einstellen von und Recherchieren nach Datensätzen in die/der Datenbank werden HTML und CGI (Perl) eingesetzt. Beide

Technologien werden in den folgenden Abschnitten hinsichtlich ihrer Anwendung in der FA-ZSV untersucht.

4.2.1.1 HTML

Für die Gestaltung der Benutzeroberfläche der FA-ZSV wird HTML (Hypertext Markup Language) (HTML 1999) eingesetzt. HTML wurde als eine Spezialisierung von SGML (Standard Generalized Markup Language) entwickelt und ist für die Gestaltung von Web-Seiten im Internet zu einem quasi-Standard geworden. Jeder Standard-Web-Browser ist in der Lage, HTML zu interpretieren. HTML wurde für dieses Teilprojekt der Sprache Java aus mehreren Gründen vorgezogen:

- Die umfassende Funktionalität von Java wird hier im Gegensatz zum eigentlichen DIWA-System nicht benötigt.
- Die Gestaltung von Benutzeroberflächen ist wesentlich weniger zeintensiv als in Java.
- Die Übertragungsraten im Zusammenhang mit CGI-Technologie sind schneller als bei Java.

4.2.1.2 CGI und Perl

Wie bereits in Abschnitt 3.2.1.3 beschrieben, handelt es sich bei CGI (Common Gateway Interface) –Programmen um Programme, die auf Grund eines HTTP-Aufrufes vom Klienten auf dem Server ausgeführt werden und die Ergebnisse dieses Programmes über HTTP zum Klienten übertragen.

In FA-ZSV wird CGI für das Einstellen von und Recherchieren nach Dokumentenbeständen in die/der Datenbank verwendet. Technisch wird die Script-Sprache Perl eingesetzt. Ausführlichere Informationen zu Perl finden Sie in Abschnitt 3.2.1.3.

4.2.2 FA-ZSV-Softwarearchitektur

Auf Grund der Verwendung der in Abschnitt 4.2.1 vorgestellten Technologien ergibt sich für FA-ZSV die folgende Softwarearchitektur (Abbildung 14):

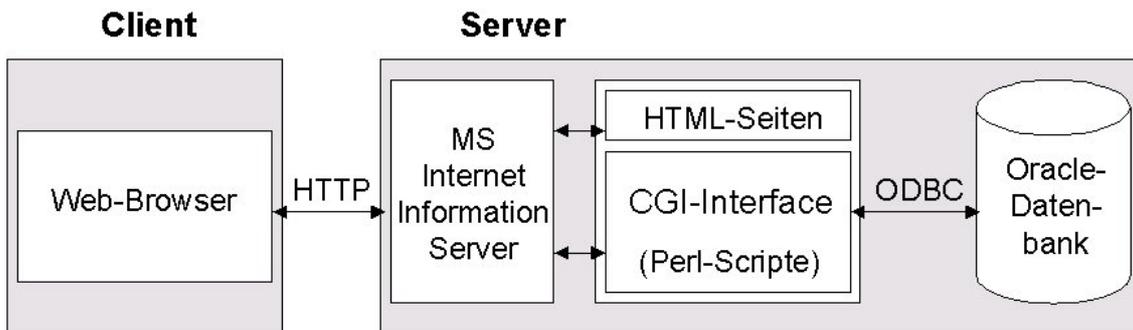


Abbildung 14: Softwarearchitektur der FA-ZSV

Als Web Server für die Kommunikation zwischen dem Web-Browser des Anwenders und der eigentlichen Anwendung wird der bereits eingesetzte MS Internet Information Server genutzt. Als Protokoll zwischen Web-Browser und Web Server wird HTTP verwendet.

Die eigentliche Anwendung setzt sich aus den HTML-Seiten für die Benutzeroberfläche und der CGI-Schnittstelle in Form von Perl-Skripten zusammen. Die CGI-Schnittstelle greift über eine ODBC-Schnittstelle auf die Datenbank zu (Two-Tier-Architektur).

Wie der Vergleich zur Softwarearchitektur von DIWA (Abbildung 12) zeigt, ist die Architektur für FA-ZSV wesentlich weniger komplex und konnte daher relativ schnell aus der DIWA-Architektur abgeleitet werden. Somit wurde der Forderung nach Offenheit gegenüber anderen Systemen (vgl. Abschnitt 2.2.4) Rechnung getragen.

4.3 Stand der Entwicklungen

Zentraler Punkt bei der Entwicklung war die Sicherung der Daten aus der nicht Jahr-2000-fähigen Datenbank. Die Daten sind bereits in die vorgegebene Datenstruktur unter MS Access übernommen worden. Die entstandene Datenbankstruktur wird nun in das Zielsystem Oracle übernommen.

Der Prototyp der Anwendung FA-ZSV befindet sich in der Entwicklung. Mitte November 1999 wurde ein erster Prototyp bei den Gewerbeaufsichtsämtern installiert, um rechtzeitig Rückmeldungen der Anwender einarbeiten zu können. Eine Benutzeroberfläche für die Erfassung neuer Erlasse ist in Abbildung 15 dargestellt.

The screenshot shows a web browser window titled "FA-ZSV - Microsoft Internet Explorer". The main content area is titled "Fachanwendung der ZSV" and contains a form titled "Erlass erfassen". The form is organized into several sections:

- Header:** Fields for "Nr.", "Datum", "Stelle" (with a dropdown menu showing "UVM"), and "Aktenzeichen".
- Betreff:** A text input field with a small vertical scroll bar.
- Fundstelle:** Fields for "Gesetzblatt" (dropdown menu showing "Abl. EG L"), "Gbl.Nr.", "Jahr", "Seite", and "veröffentlicht am".
- Sachgebiet:** Fields for "Sachgebiet" (dropdown menu showing "AbfR"), "Ordnungszahl", and "Jahresbericht" (dropdown menu showing "Nicht im Jahresbericht").
- Vorgang erforderlich ?** Radio buttons for "ja" and "nein" (selected), followed by a "Vorgangsnummer" field.
- Schlusssatz:** Fields for "Schlagwort" (dropdown menu showing "Schlagwort 1") and "Bemerkungen" (text input field).

On the left side of the form, there is a vertical navigation menu with the following items: "Erlass", "Vorschrift", "Artikelgesetz", "Arbeitsanweisung", "Vorgang", "Abgleichliste", "Monatsbericht", "Mail versenden", and "Jahresbericht". On the right side, there are three icons: a floppy disk (save), a red 'X' (cancel), and a printer icon.

Abbildung 15: Erfassen eines Erlasses im ersten Prototyp der FA-ZSV

Bis Mitte Februar 2000 wurden alle benötigten Funktionalitäten in der neuen Anwendung realisiert.

5 Anwender und erste Erfahrungen

Das System DIWA wurde im UVM und in der LfU installiert. Als Testanwender im UVM konnten das Referat 72 und die Akademie für Natur- und Umweltschutz Baden-Württemberg gewonnen werden. Das Referat 72 des UVM nutzt den ca. 750 Dokumente umfassenden Bestand des Projekts Kernreaktor-Fernüberwachung (KFÜ) zum Testen. Insbesondere sollen dabei alle wichtigen Dokumente der laufenden KFÜ-Erneuerung in einem zentralen Bestand gehalten werden. Bisher angefallener Zeitaufwand für das Führen von Handakten soll eingespart und das Kopieren vermieden werden. Hierzu werden alle Dokumente mit den Projektpartnern mittels E-Mail ausgetauscht. Das bisher durch die Ablagestruktur bedingte, teilweise schwierige Wiederauffinden von alten Dokumenten wird durch DIWA erleichtert. Es wurde eine Liste mit Metainformationen für die Verschlagwortung der KFÜ-Dokumente aufgestellt. Die Dokumente werden entsprechend ihrer Wichtigkeit und Gültigkeitsdauer kontinuierlich verschlagwortet. Dabei stellt sich heraus, dass diese Verschlagwortung „per Hand“, also ohne softwareseitige Unterstützung, sehr aufwendig ist.

In der LfU wurden Teile des Intranet-Angebotes an Umweltdaten und –informationen für den Test von DIWA eingesetzt. Hier wurden zwei Probleme sichtbar. Zum einen war die Einbindung der Dokumentenbestände auf Unix-Plattformen über das Novell-Netzwerk entgegen der Dokumentation des MS IIS bisher nicht möglich. Als Lösung wurde der Zugang über einen Samba-Server realisiert. Eine zweite Hürde stellte der Aufbau der virtuellen Ablagestruktur dar. Da die physischen Verzeichnisse für die Web-Präsentation im Intranet strukturiert sind, ist eine eindeutige Zuordnung zu Umweltthemen nicht immer möglich. Ausserdem sind in den Web-Verzeichnissen nicht nur HTML-Dokumente, sondern auch Programme für die dynamische Seitenerzeugung abgelegt, die aber für

einen Anwender nicht sichtbar sein sollen. Als Lösung wäre hier ein Filter für die Anzeige im DIWA-Explorer denkbar.

6 Vergleich der Anforderungen mit dem erreichten Projektstand

Die frühzeitige Einführung eines Prototyps von DIWA in UVM und LfU hat sich auf Grund der Rückmeldungen und der Mitarbeit der Anwender bereits in der ersten Phase als positiv erwiesen. Daraus ergaben sich auch zusätzliche Aspekte zu den am Anfang gestellten Anforderungen, die bereits jetzt umgesetzt werden konnten.

Insgesamt konnten die generellen Anforderungen an DIWA und die von Anwendern im Laufe des Projekts vorgeschlagenen neuen Funktionalitäten erfüllt werden. Ein Ausbau der einzelnen Komponenten unter Berücksichtigung weiterer Anwenderanforderungen ist jedoch erforderlich, um zu einer robusten, anwenderfreundlichen und einfach wartbaren Betriebsversion von DIWA zu gelangen (vgl. auch Kapitel 7).

In den folgenden Abschnitten werden die zu Beginn des Projekts gestellten Anforderungen (siehe Kapitel 2) mit dem erreichten Projektstand abgeglichen.

6.1 Erfüllung grundlegender Anforderungen

6.1.1 Einbindung von DIWA in die vertraute Bürokommunikationsumgebung

DIWA wird über einen Standard-Web-Browser auf dem Rechner des Anwenders gestartet. Für die Gestaltung der einzelnen Komponenten von DIWA wurde weitgehend auf das Layout von Microsoft Windows zurückgegriffen. Die Installation zusätzlicher Software auf dem Rechner des Anwenders ist nicht notwendig.

6.1.2 Berücksichtigte Dateiformate bei der Indizierung

Der MS IS kann standardmäßig folgende Dateiformate verarbeiten:

- Microsoft Word, Excel, PowerPoint
- HTML
- Reine Textdokumente (plain text)

Für die Indizierung von PDF-Dokumenten wird der von Adobe angebotene Filter eingesetzt.

Mit Hilfe des HTML-Filters können auch XML-Dokumente in den Volltextindex aufgenommen werden. Für die Indizierung von Eigenschaften (Tags) der XML-Dokumente steht von kommerzieller Seite noch kein Filter zur Verfügung. Die Eigenentwicklung eines solchen Filters erfordert einen sehr hohen Aufwand. Auf Grund der zunehmenden Bedeutung von XML ist allerdings mit einem baldigen Angebot von kommerzieller Seite zu rechnen.

6.1.3 Systematischer Zugang zum Web-Archiv

Der geforderte systematische Zugang zum Web-Archiv wurde über die virtuelle Ablagestruktur realisiert. Über eine Struktur werden somit mehrere Dokumentenbestände verfügbar gemacht. Die Anwender müssen nun keine Kenntnis darüber besitzen, wo sich die einzelnen Dokumentenbestände befinden, sondern können sich an Hand der virtuellen Ablagestruktur durch eine thematische Sicht bewegen. Dieser Zugang wird sowohl in der Autorenkomponente zum Einstellen der Dokumente als auch in der Recherchekomponente zur Einschränkung des Suchraums sowie im navigatorischen Zugang in einer übersichtlichen Baumstruktur verwendet.

6.1.4 Einbinden bereits vorhandener Umweltdokumentenbestände in das Web-Archiv

Im UVM wurden bisher das UVM-weite INFO-Fach, die Dokumente des Projekts Kernreaktor-Fernüberwachung im Referat 72 sowie einzelne Dokumentenbestände der Akademie für Natur- und Umweltschutz in das Web-Archiv von DIWA eingebunden. Alle genannten Dokumentenbestände liegen auf Microsoft Windows NT-Rechnern vor.

In der LfU wurden Dokumente des Intranet-Angebots in das Web-Archiv von DIWA aufgenommen. Diese Dokumente liegen auf einem Unix-Server vor. Um diese Dokumente für DIWA zugänglich zu machen, musste auf dem Unix-Server der Samba-Server installiert werden, der einen Zugriff auf die Platten des Unix-Rechners ermöglicht. Als problematisch erwies sich die Tatsache, daß keine einzelnen Dokumente, sondern nur vollständige physikalische Verzeichnisse eingebunden werden können, d.h. es werden auch Dateien eingebunden und vom MS IS indiziert, die auf dem LfU-Intranet-Server nicht über Links angeboten werden. Das hat zur Folge, dass bei Recherchen nicht in das Intranetangebot aufgenommene Dateien als Suchergebnisse auftauchen können. Auch im DIWA-Explorer erscheinen die kompletten Inhalte der eingebundenen physischen Verzeichnisse.

Insgesamt gestaltet sich das Einbinden vorhandener Dokumentenbestände als technisch unproblematisch. Allerdings erfordert die Verschlagwortung dieser Dokumente mit Metainformationen einigen Aufwand auf Seiten der Anwender.

6.1.5 Einbindung von DIWA in die vorhandenen Systemumgebungen

UVM

Das Einbinden von DIWA in die reine Microsoft Windows NT-Umgebung des UVM wurde realisiert. Die strikten Regelungen der Zugriffsrechte konnten mit

Hilfe der Entwicklung von Perl-Skripten als sogenannte Wächter eingehalten werden. Die Sicherheitsrestriktionen des MS IIS allein reichten hierfür nicht aus (siehe Abschnitt 3.2.1.3).

LfU

In der LfU wurde DIWA auf einem Microsoft Windows NT-Server installiert. Die bisher eingebundenen Dokumentenbestände befinden sich aber auf Unix-Rechnern. Die Einbindung der Dokumentenbestände war bisher nicht über das bereits vorhandene Novell-Netzwerk zu realisieren, da Kompatibilitätsprobleme zwischen MS Windows NT und Novell auftraten. Bei der jetzigen Lösung musste ein Samba-Server auf dem Unix-Rechner installiert werden. Dieser stellt die UNC-Pfadnamen sowie die Zugriffsberechtigungen der einzubindenden Verzeichnisse in einem von Windows NT verarbeitbaren Format zur Verfügung.

6.1.6 Intranet-weiter Einsatz von DIWA

Bisher wurde DIWA im UVM und in der LfU installiert und getestet. Dabei wurden jeweils nur die eigenen Dokumentenbestände in das Web-Archiv des UVM bzw. das Web-Archiv der LfU aufgenommen. Ein über beide Institutionen übergreifendes Web-Archiv wurde auf Grund der bestehenden Sicherheitsrestriktionen beim gegenseitigen Zugriff bisher nicht realisiert.

6.2 Einstellen von Dokumenten in das Web-Archiv

Autoren können die Autorenkomponente aus der DIWA-Rechercheoberfläche heraus in einem Standard-Web-Browser starten. Damit wird ihnen ermöglicht, in der gewohnten Bürokommunikationsumgebung erstellte Dokumente auf Web Servern einzustellen. Bereits mit dem Einstellen eines Dokuments wird es automatisch der Suchmaschine bekannt gegeben. Beim Einstellen von Doku

menten bzw. beim Löschen von Dokumenten oder Verzeichnissen werden die auf Betriebssystemebene festgelegten Zugriffsrechte eingehalten.

Das Einstellen von Dokumenten wird über die virtuelle Ablagestruktur ermöglicht, die in einer übersichtlichen und intuitiv bedienbaren Baumstruktur dargestellt wird.

Die Autorenkomponente wurde in Java realisiert und basiert somit auf Internettechnologien. Eine lokale Installation beim Anwender erübrigt sich daher.

6.3 Recherche nach Dokumenten im Web-Archiv

Für Anwender von DIWA wird die Recherche in den Dokumentenbeständen des Web-Archivs über Standard-Web-Browser und nicht über Oberflächen spezieller Werkzeuge ermöglicht.

Die Recherche im Web-Archiv kann sowohl über eine Volltextrecherche als auch über eine Recherche nach Dokumenteigenschaften durchgeführt werden. Bei den Dokumenteigenschaften kann es sich sowohl um von der Anwendung vordefinierte Eigenschaften der MS Office- oder PDF-Dokumente wie Autor oder Titel handeln als auch um benutzerdefinierte Informationen wie Zeit- oder Raumbezug.

Für die Formulierung der Abfrage können sich die Anwender von einem vorgehalteten Umweltthesaurus unterstützen lassen. Damit ist es möglich, auch nach verwandten Begriffen, Ober-, Unterbegriffen etc. zu suchen.

Die Anwender können bei der Abfrage auswählen, ob sie nur in dem Web-Archiv von DIWA suchen möchten oder ihre Abfrage zusätzlich an andere Suchmaschinen weiterleiten möchten. Bisher können die Suchmaschinen AltaVista und Lycos verwendet werden. Spezielle Umweltdokumentenbestände (z.B. HUDA, DRS, XfaWeb) werden noch nicht über DIWA verfügbar gemacht,

da in erster Linie solche Bestände eingebunden werden sollten, für die noch keine Recherchemöglichkeit besteht.

Die thematische Strukturierung des Web-Archivs kann über die virtuelle Ablagestruktur, die im DIWA-Explorer in Baumstruktur dargestellt wird, eingesehen werden. Über diese virtuelle Ablagestruktur können die Anwender auch die Suche auf bestimmte Themen einschränken.

Die Formatierung der Suchergebnisse kann nach unterschiedlichen Kriterien vorgenommen werden. Es kann zwischen verschiedenen Mengen von Ausgabedaten, die u.a. verschiedene Dateieigenschaften, eine Kurzfassung des Textes etc. beinhalten, ausgewählt werden. Weiterhin kann die Sortierung der Ergebnisse nach verschiedenen Kriterien wie Trefferanzahl im Text, Titel, Dateigröße etc. festgelegt werden.

Die Suchergebnisse einer Abfrage enthalten generell alle zum Thema verfügbaren Dokumente unabhängig von den Zugriffsrechten des Anwenders. Erst beim Herunterladen des Dokuments auf den lokalen Rechner wird eine entsprechende Überprüfung durchgeführt. Also können auch Anwender, die keine Leserechte für bestimmte Dokumente besitzen, zumindest Informationen über deren Existenz erhalten.

Eine während der Entwicklungsarbeiten von den Anwendern gestellte Anforderung wurde mit dem DIWA-Explorer als navigatorischem Zugang zum Web-Archiv realisiert. Hiermit ist es möglich, dass Anwender sich einen Überblick über alle verfügbaren Dokumente machen können. Da eine Recherche in der Regel immer nur einen Teil des Gesamtdokumentenbestands liefert, ist der DIWA-Explorer eine wichtige Ergänzung zur DIWA-Recherchekomponente.

6.4 FA-ZSV

Die Daten der bisherigen Fachanwendung der Zentralen Stelle für die Vollzugsunterstützung beim Staatlichen Gewerbeaufsichtsamt Stuttgart (FA-ZSV), die in einer nicht Jahr-2000-fähigen Datenbank vorlagen, wurden gesichert und in Microsoft Access übernommen. Von dort werden sie in das Zielsystem Oracle übertragen.

Mit Hilfe der Internettechnologien HTML und CGI konnte bereits ein Prototyp der mehrplatzfähigen Anwendung FA-ZSV gestaltet werden.

Das Eintragen von Daten in die Datenbank sowie das Recherchieren nach Datensätzen in der Datenbank ist von denselben Benutzeroberflächen aus möglich. Die vollständigen Funktionalitäten zum Einstellen, Ändern und Recherchieren nach Erlassen und Vorschriften wurden fertiggestellt.

Mit dem System wird die weitgehend automatisierte Erstellung von Arbeitsanweisungen sowie Monats- und Jahresberichten (Vorgangsbearbeitung) ermöglicht.

Digital vorliegende Dokumente werden nach Abschluss der Arbeiten an der eigentlichen Anwendung über eine Recherchefunktionalität (ähnlich DIWA) verfügbar gemacht werden.

Aus dieser Teilanwendung ergeben sich neue Aspekte für die weitere Arbeit von DIWA, wie z.B. der Zugriff auf Datenbanken aus der Recherchekomponente heraus.

6.5 Offenheit gegenüber anderen Systemen

Eine Anforderung bzgl. der Offenheit von DIWA war, in DIWA verwendete Technologien und Funktionalitäten auch in anderen Systemen einzusetzen. Neu aufgenommen in das Anforderungsspektrum von DIWA wurde das Teilprojekt FA-ZSV. Hier konnten bereits mit dem DIWA-Prototypen gemachte Erfahrungen sowie entwickelte Komponenten sinnvoll genutzt werden. In Zukunft ist denkbar, die Recherchefunktionalität von DIWA in FA-ZSV für digital vorliegende Dokumente in FA-ZSV einzusetzen.

Bisher wurden in DIWA noch keine anderen Recherchesysteme, die z.B. andere Technologien verwenden, angebunden, um so von einer einheitlichen Oberfläche für Anwender zugänglich zu sein. Technisch ist dies aber möglich. Zu berücksichtigen ist dann auch die Anbindung von Systemen, die ihre Dokumente in Datenbanken und nicht auf Web Servern verwalten.

7 Ausblick

Im Rahmen von DIWA wird die Fachanwendung der Zentralen Stelle für die Vollzugsunterstützung beim Staatlichen Gewerbeaufsichtsamt Stuttgart realisiert, die die Pflege der Vorschriftensammlung der Gewerbeaufsicht beinhaltet. Die Kopplung von FA-ZSV an DIWA wird schrittweise so durchgeführt werden, dass zuerst die digital vorliegenden Dokumente über die Recherchekomponente verfügbar gemacht werden. In diesem Rahmen wird auch die Einbindung von Datenbankabfragen in die Recherchekomponente angestrebt, um weitere Informationsquellen, wie die Datenbank der FA-ZSV, einbinden zu können. Weiterhin wird überlegt, weitere spezielle Recherchewerkzeuge (XfaWeb (Weidemann et al. 1998), DRS (Document Retrieval System) usw.) in DIWA zu integrieren, das Informationsspektrum zu erweitern.

In Verbindung mit der Recherchekomponente könnten schließlich Suchabfragen bzw. Suchergebnisse gespeichert werden. Mit dieser Funktionalität würde 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 bestimmte Suchabfragen neu zusammenzustellen.

Beim aktuellen DIWA-Prototypen müssen alle virtuellen Verzeichnisse an zentraler Stelle eingerichtet und von dort aus gewartet werden. Dies trifft in analoger Weise für die 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. In Anbetracht der steigenden Zahl von DIWA-Nutzern ist bereits jetzt absehbar, dass hier ein Engpass entstehen wird. Vor diesem Hintergrund sollte ein Werkzeug entwickelt werden, mit dem jeder Anwender die für seine Dokumentenbestände benötigte virtuelle Ablagestruktur selbst administrieren kann. Dieses Werkzeug könnte Funktiona

litäten zum Einrichten, Ändern und Erweitern der virtuellen Ablagestruktur umfassen. Zudem sollte Anwendern ein Werkzeug angeboten werden, über das sie in Analogie zur Verwaltung von virtuellen Ablagestrukturen ihre eigenen Suchformulare gestalten können.

Als aufwendig stellt sich die Verschlagwortung bereits vorhandener Dokumentenbestände dar. Bisher wird so vorgegangen, dass die Verschlagwortung kontinuierlich nach Bedeutung und zeitlicher Gültigkeit der Dokumente vorgenommen wird. Zur Unterstützung der Autoren bei dieser Arbeit existieren zwei Wege. Zum einen könnten entsprechende Komponenten bereits in die Software zur Erstellung der Dokumente eingebaut werden. 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 Dokuments verwendeten Software. Gleichzeitig würde sichergestellt, dass nur verschlagwortete Dokumente in das Web-Archiv eingestellt werden. Gerade der zweite Aspekt ist von Bedeutung: Eine Recherche über Metainformationen ist umso erfolgreicher, je mehr Dokumente vollständig erschlossen wurden. Hier sind die Entwicklungen im Projekt GEIN (GEIN 2000) zu beobachten, wo ein spezielles Werkzeug für die Unterstützung bei der Verschlagwortung („Verschlagwortungsassistent“) geplant ist.

Abkürzungsverzeichnis

CGI	Common Gateway Interface
CLI	Call-Level-Interface
DIWA	Dokumentenverwaltung im Web-Archiv von Ministerium für Umwelt und Verkehr und Landesanstalt für Umweltschutz Baden-Württemberg
DRS	Document Retrieval System
FA-ZSV	Fachanwendung der Zentralen Stelle für die Vollzugsunterstützung beim Staatlichen Gewerbeaufsichtsamt Stuttgart
FTP	File Transfer Protocol
GAA	Staatliches Gewerbeaufsichtsamt
GEMET	General European Multilingual Environment Thesaurus
GLOBUS	Globale Umweltsachdaten
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
Htx	HTML-Extension
HUDA	Hypermediatechnik für Umweltdaten
IDC	Internet Database Connector
Idq	Internet Data Query
IP	Internet Protocol
ISAPI	Internet Server Application Programming Interface
JDBC	Java Database Connectivity

JDK	Java Developers Kit
KFÜ	Kernreaktor-Fernüberwachung
LAN	Local Area Network
LfU	Landesanstalt für Umweltschutz Baden-Württemberg
MS IIS	Microsoft Internet Information Server
MS IS	Microsoft Index Server
ODBC	Open Database Connectivity
PDF	Portable Document Format
RMI	Remote Method Invocation
SGML	Standard Generalized Markup Language
TCP/IP	Transmission Control Protocol/Internet Protocol
UDK	Umweltdatenkatalog
UNC	Universal Naming Convention
URL	Uniform Resource Locator
UVM	Ministerium für Umwelt und Verkehr Baden-Württemberg
VSGA	Vorschriftensammlung der Gewerbeaufsicht
WWW	World Wide Web
XML	Extensible Markup Language

Literaturverzeichnis

- Adobe (1999): <http://www.adobe.com/support/downloads/acwin.htm>.
- Alkaline (1999): <http://searchtools.com/tools/alkaline.html>.
- ASTAWare (1999): <http://searchtools.com/tools/searchkey.html>.
- CGI (1999): <http://hoohoo.ncsa.uiuc.edu/cgi/overview.html>.
- Flanagan, D. (1998): Java in a Nutshell, O'Reilly Verlag.
- FTP (1999): <http://rzdspc77.informatik.uni-hamburg.de/Internet/FTP/einfuehrung.html>.
- GEIN (2999): <http://gein.de>.
- HTML (1999): http://www.hiz.de/html_frame/index.html.
- Infoseek (1999): <http://software.infoseek.com/products/ultraseek/ultratop.htm>.
- ISAPI (1999): http://msdn.microsoft.com/library/devprods/vs6/visualc/vccore/_core_isapi.3a_extending_web_servers.htm
- Kussmaul, A. (1999): Multimediales Dokumentenarchiv für das World Wide Web, Diplomarbeit, Fachhochschule Ulm, 1999.
- Perl (1999): <http://www.activestate.com/ActivePerl/>.
- 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.
- Samba (1999): <http://de.samba.org/samba/about.html>.
- Swish-E (1999): <http://sunsite.berkeley.edu/SWISH-E/Manual/>.
- Tochtermann, K.; Kadric, L.; Rainbold, E.; Pursche, K.; Kramer, R.; Schmidt, R.; Jaeschke, A.; Geiger, W.; Wiesel, J.; Hofmann, C.; Keitel, A.; Reimann, A.; Schütz, T.; Burkhardt, J. (1999): Erstellung

von Umweltberichten mit dem HUDA-Werkzeugkasten, Tagungsband Umweltinformatik `99, Metropolis Verlag, 1999.

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

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; Abschlussbericht Phase V 1998; Hrsg. Mayer-Föll, R., Jaeschke, A.; Wissenschaftliche Berichte; FZKA 6250.