

Seminar XML und Datenbanken

XML-basiertes Content-Management

Andreas Kümpel

Februar 2003

Gliederung

1. Content-Management-Systeme

- 1.1 Motivation
- 1.2 Begriffsbestimmung
- 1.3 Kennzeichen
- 1.4 Architektur

2. XML-Anfragesprachen

- 2.1 Anforderungen
- 2.2 XML-QL
- 2.3 XXL
- 2.4 XQL
- 2.5 XIRQL

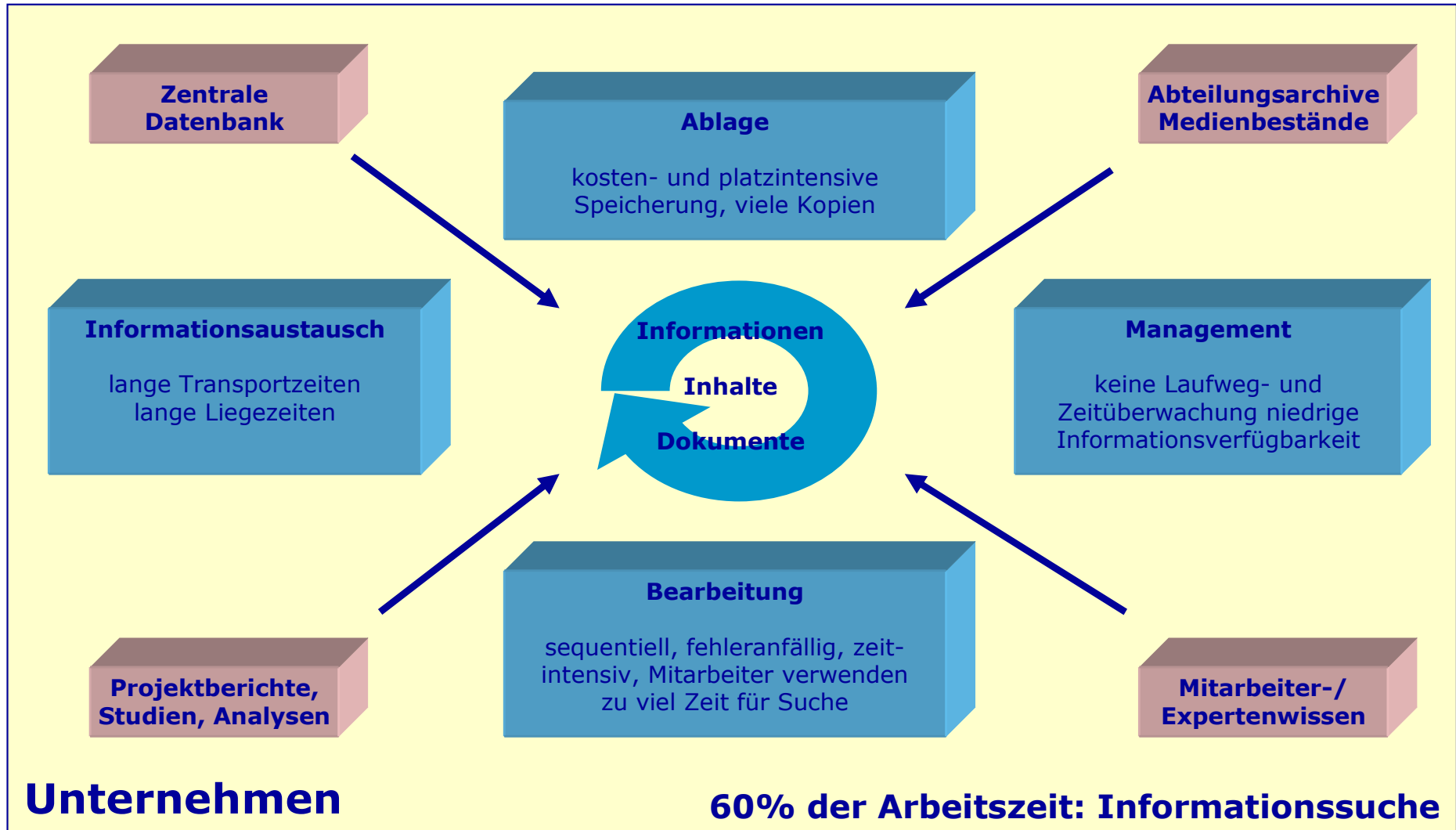
3. Produkte

- 3.1 MS Content
Management Server
- 3.2 Zope

4. Zusammenfassung / Ausblick

Content- Management-Systeme

„Motivation“



Begriffsbestimmung

Dokumenten Management System

- Verwaltung und Archivierung von bereits vorhandenen Dokumenten
- Versionverwaltung
- Suche über separate Meta-Daten
- „Dokumenten-Server“

Content Management System

- *Generierung*, Verwaltung und Archivierung von *Inhalten* zur Publikation
via Intranet oder Internet
- Organisation der Arbeitsschritte bei der Inhalts- und Dokumentenerstellung

Begriffsbestimmung (2)

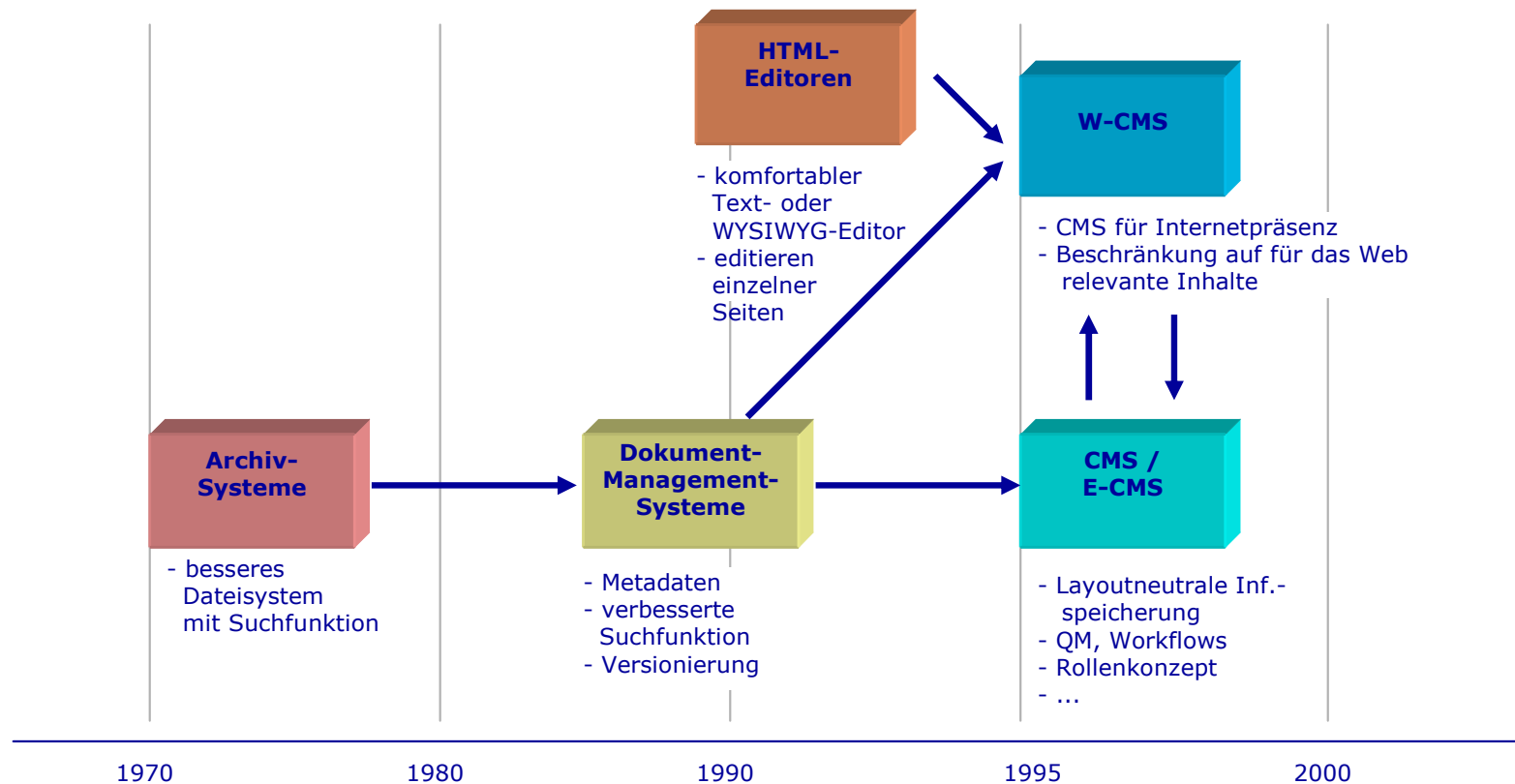
Web-Content Management System (W-CMS)

- Konzentration auf für die Erstellung von Web-Präsentationen (Internetseiten, Portale) relevante Inhalte

Enterprise Content Management System (E-CMS)

- Begriffsverschiebung von „klassischen“ CMS zu E-CMS und von W-CMS zu CMS

Zeitliche Entwicklung von DMS/CMS



Kennzeichen eines CMS

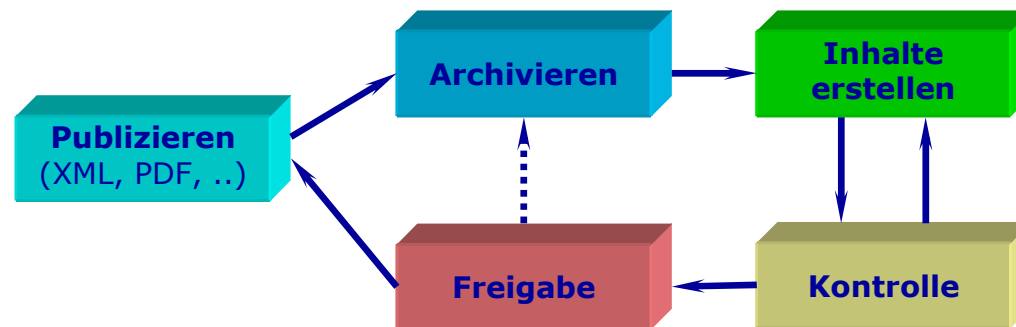
- Trennung von Inhalt und Layout

- XML
- einfache Wiederverwendung von Inhalten
- Erstellung von Publikationen (HTML, PDF) mittels Templates

Kennzeichen eines CMS (2)

- Unterstützung des „Content-Lifecycle“ / Workflow-Anbindung

- Unterstützung des gesamten Ablauf des Content-Lebenszyklus:



- Qualitätssicherung (ISO 9000)
- Erstellung eigener Workflows, d.h. Dokumente gehen während ihrer Entwicklung einen vordefinierten Weg „durch“ das Unternehmen

Kennzeichen eines CMS (3)

- Rollenkonzepte

- in Zusammenhang mit „Content-Lifecycle“
- Sicherheitsaspekte
- unterschiedliche Benutzer-Interfaces
- Kontexte (unterschiedliche Sichten auf Daten)

Kennzeichen eines CMS (4)

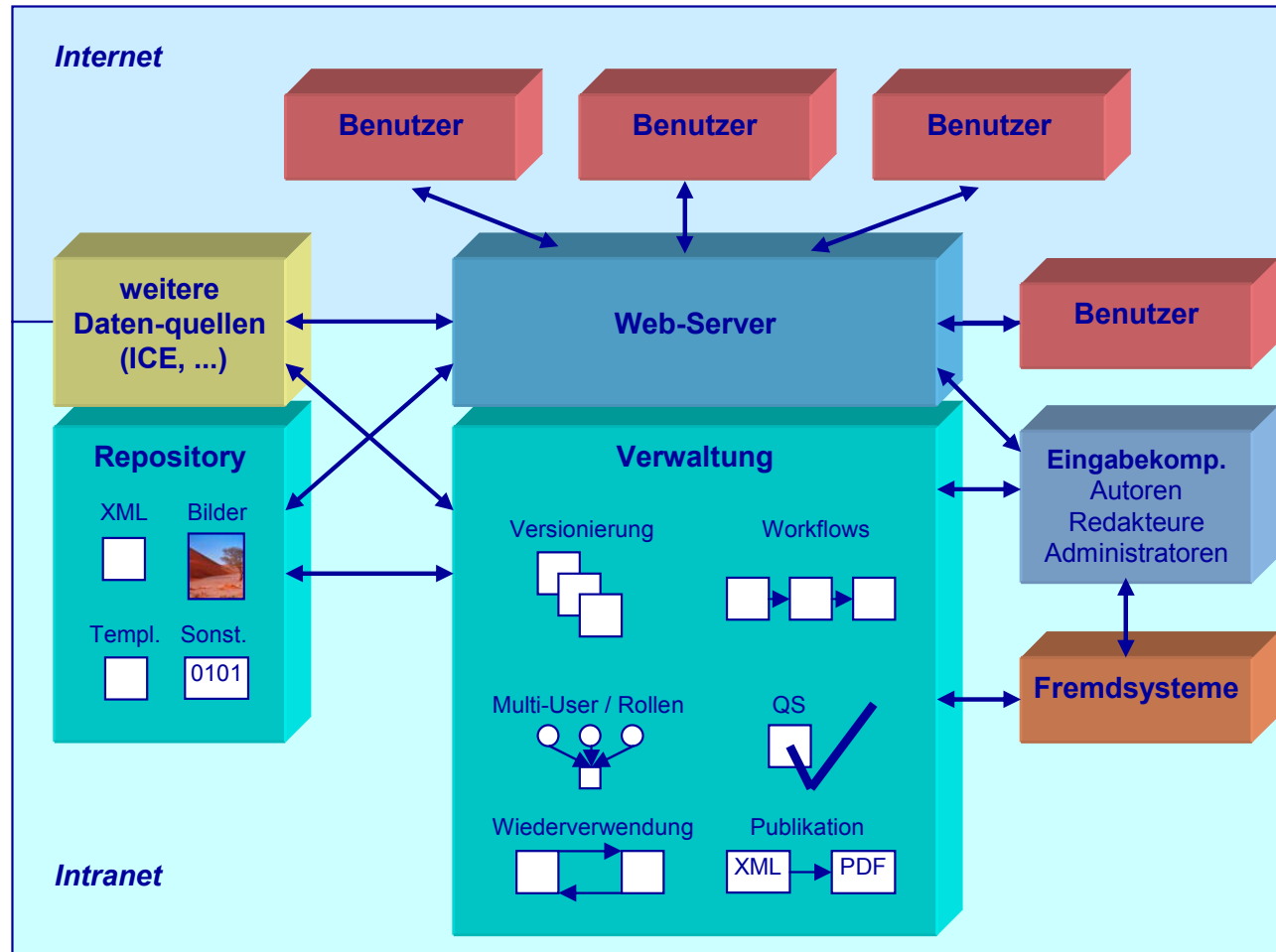
- Versionierung und Configuration-Management

- „Konsistente“ Dokumente / Inhalte
- Lange Transaktionen bei paralleler Bearbeitung

- Anbindung an Fremdsysteme

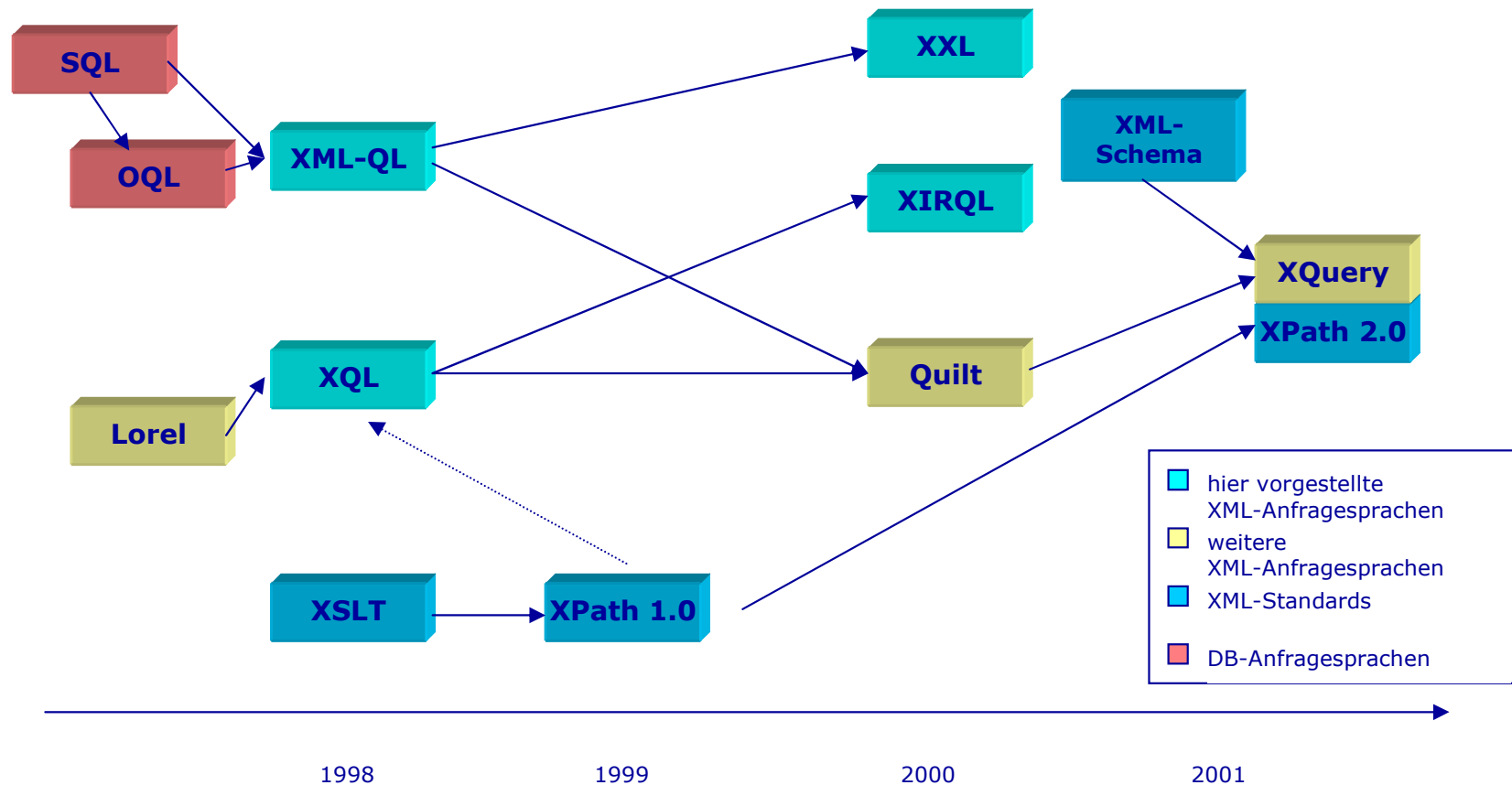
- Import der Daten von Fremdsystemen

Architektur eines CMS



XML- Anfragesprachen

Historische Entwicklung von XML-Anfragesprachen



Aufgaben von XML-Anfragesprachen

- Suche in XML-Dokumenten

(select ?? from abc.xml where „<titel>...</titel>“ = „ABC“ ???)

- Suche in „verteilten“ XML-Dokumenten

Datenbanken, Dateisystemen, Internet, ...

- Suche auf XML-Dokumenten mit und ohne Schema / DTD

- Einfluss auf die Struktur der Ausgabe

- Teilbaum eines XML-Dokumentes
- Neue XML-Datenstruktur als Ergebnis

XML-QL

- **XML-Query Language**
- im Rahmen eines AT&T Labs-Research-Projektes entwickelt

Ziele

- Daten-Extrahierung aus großen XML-Dokumenten
- Transformation von XML nach XML (unterschiedliche Schemata)
- Suche über mehrere XML-Dokumente

Syntax

- WHERE - CONSTRUCT-Klausel
- WHERE: Vergleich von Pfadausdrücken, Bindung an Variablen, Filterung
- CONSTRUCT: Ergebnisse als neues XML-Dokument liefern

XML-QL - Beispiel

```

<bib>
<book year="1995">
  <title> An Introduction to Database Systems </title>
  <author> <lastname> Weber </lastname> </author>
  <publisher> <name> Addison-Wesley </name > </publisher>
</book>

<book year="1998">
  <title> Foundation for Object/Relational Databases: The Third Manifesto </title>
  <author> <lastname> Weber </lastname> </author>
  <author> <lastname> Darwen </lastname> </author>
  <publisher> <name> Addison-Wesley </name > </publisher>
</book>
</bib>

```

Beispiel-Anfrage

```

WHERE <book year=$y>
  <publisher><name>Addison-Wesley</></>
  <title> $t</>
  <author> $a</>
</> IN "www.ab.de/ab.xml", $y > 1990
CONSTRUCT <result>
  <author> $a</>
  <title> $t</>
</>

```

Ergebnis

```

<result>
  <author> <lastname> Weber </lastname> </author>
  <title> An Introduction to Database Systems</title>
</result>
<result>
  <author> <lastname> Weber </lastname> </author>
  <title> Foundation for Object/Relational Databases:
The Third Manifesto </title>
</result>
<result>
  <author> <lastname> Darwen </lastname> </author>
  <title> Foundation for Object/Relational Databases:
The Third Manifesto </title>
</result>

```

XML-QL - Beispiel (2)

```

<bib>
<book year="1995">
  <title> An Introduction to Database Systems </title>
  <author> <lastname> Weber </lastname> </author>
  <publisher> <name> Addison-Wesley </name > </publisher>
</book>

<book year="1998">
  <title> Foundation for Object/Relational Databases: The Third Manifesto </title>
  <author> <lastname> Weber </lastname> </author>
  <author> <lastname> Darwen </lastname> </author>
  <publisher> <name> Addison-Wesley </name > </publisher>
</book>
</bib>

```

Beispiel-Anfrage

```

WHERE <book>
  <publisher><name>Addison-Wesley</></>
  <title> $t</>
  <author> $a</>
</> CONTENT AS $p IN "www.ab.de/ab.xml"
CONSTRUCT <result><title> $t </>
  WHERE <author> $a </> IN $p
  CONSTRUCT <author> $a </>
</>

```

Ergebnis

```

<result>
  <title>An Introduction to Database Systems</title>
  <author> <lastname> Weber </lastname> </author>
</result>

<result>
  <title> Foundation for Object/Relational Databases:
The Third Manifesto </title>
  <author> <lastname> Weber </lastname> </author>
  <author> <lastname> Darwen </lastname> </author>
</result>

```

XXL

- Flexible **X**ML Search **L**anguage
- XML-Anfragesprache mit „Ranked Retrieval“
- Begriffshierarchie als Ontologiegerüst für die Suchmaschine

Struktur der Anfrage

Select ...

From ...

Where ...

- Vergleiche mit Ähnlichkeitsoperator „~“ auf Element-Inhalte und - Namen

XXL - Beispiel

```
<region>Europa
<ort>Sylt
<lage>Deutschland</lage>
<beschreibung>Eine .... Insel.</beschreibung>
<aktivitäten>Dünenwanderung, Surfen</aktivitäten>
<saison>Sommer</saison>
</ort>
</region>
```

```
<region>Übersee
<sehenswürdigkeit>Townsville
<land>Australien</land>
<attraktion>Korallenriff</attraktion>
<aktivitäten>Tauchen, Schnorcheln</aktivitäten>
<saison>Ganzjährig</saison>
</sehenswürdigkeit>
</region>
```

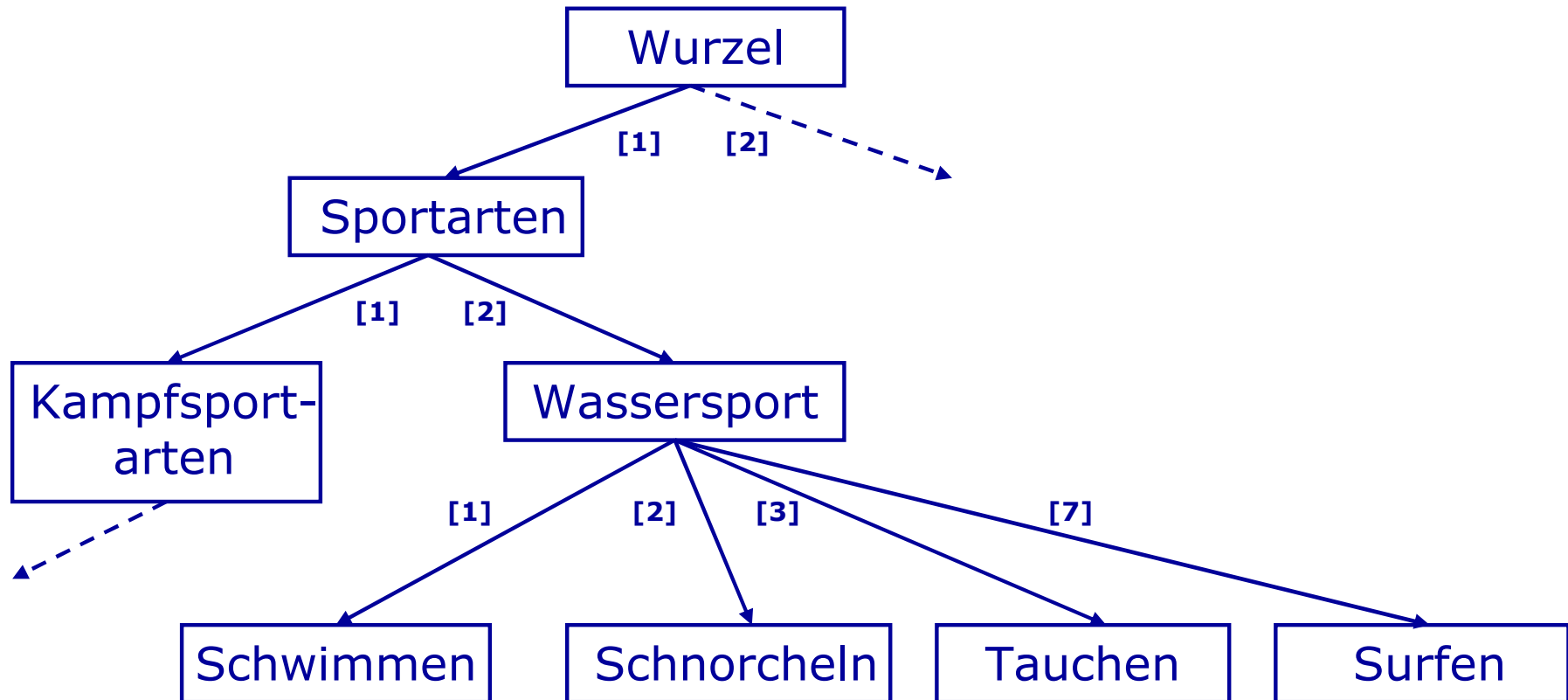
Beispiel-Anfrage 1:

```
Select P
From http://www.wehwehweh.de/reiseziele.xml
Where region.(ort|sehenswürdigkeit) As P
And P.# LIKE "%Schwimmen%"
And P.# LIKE "%Sommer%"
```

Beispiel-Anfrage 2:

```
Select P
From http://www.wehwehweh.de/reiseziele.xml
Where region.~ort As P
And P.# ~ "Schwimmen"
And P.# ~ "Sommer"
```

XXL - Beispiel (Begriffshierarchie)



XQL

- **X**ML **Q**uery **L**anguage
- 1998 von J. Robie, J. Lapp und D. Schach entwickelt
- baut auf XPath auf

- einfache Syntax
- kompakte XQL-Ausdrücke (leicht zu lesen und zu schreiben)
- XQL-Ausdrücke als Strings zur Integration in URLs

- keine Variablen
- kein Construct-Block!
- Format:

`pfad-ausdruck[filter]: /wurzel/eintrag[wert='abc']`

XQL - Beispiel

```
<filme>
  <film datum="01.07.1998">
    <titel>Das fünfte Element</titel>
    <regie>Luc Besson</regie>
    <produzent>...</produzent>
    <genre>Science Fiction</genre>
    <beschreibung>...</beschreibung>
  </film>
  <film datum="01.05.2001">
    <titel>Planet der Affen</titel>
    <regie>Tim Burton</regie>
    <produzent>...</produzent>
    <genre>Science Fiction</genre>
    <beschreibung>...</beschreibung>
  </film>
</filme>
```

Beispiel-Anfrage

```
/filme[//filme/film/regie='Luc Besson']
```

```
/filme/film/titel[//regie='Luc Besson']
```

```
//film[genre=//film[titel='Planet der Affen']/genre]
```

```
//film/@datum[//film/titel='Planet der Affen']
```

Ergebnis

Alle Filme ab Knoten „filme“, bei denen Luc Besson Regie geführt hat

Liefert alle Titel der Filme, bei denen Luc Besson Regie geführt hat

Liefert alle Filme, bei denen das Genre dasselbe ist, wie bei „Planet der Affen“

Liefert das Datum des Films mit dem Titel „Planet der Affen“

XIRQL

- **X**ML **I**nformation **R**etrieval **Q**uery **L**anguage
- Ziel: Information Retrieval auf XML-Dokumenten ermöglichen
- Erweiterung der XML-Anfragesprache XQL:
- bisher nur ansatzweise als Prototyp implementiert
(HyRex, Hypermedia Retrieval Engine for XML)

1. Gewichtung

```
/document[.//heading cw 'XQL' and .//section/* cw 'XQL']
```

```
/document[0.6 * .//heading cw 'XQL'+ 0.4 * .//section/* cw 'XQL']
```


XIRQL (2)

2. Vage Prädikate und Datentypen

- „Finde den Politiker, der so ähnlich wie Manela heißt und um 1950 Mitglied des African National Congress war“.
- Unterstützung von Prädikaten wie „ähnlich wie“, „um“, „ungefähr“ (%soundlike%)
- neue Datentypen (Domäne, vage Prädikate)

3. Relevanz-orientierte Suche

- nur relevante Teile eines Dokumentes zurückliefern („kleinsten Context“)
- „Root-Nodes“

4. structural relativism

- kein Unterschied zwischen Attributen und Elementen

5. Hyperlinks

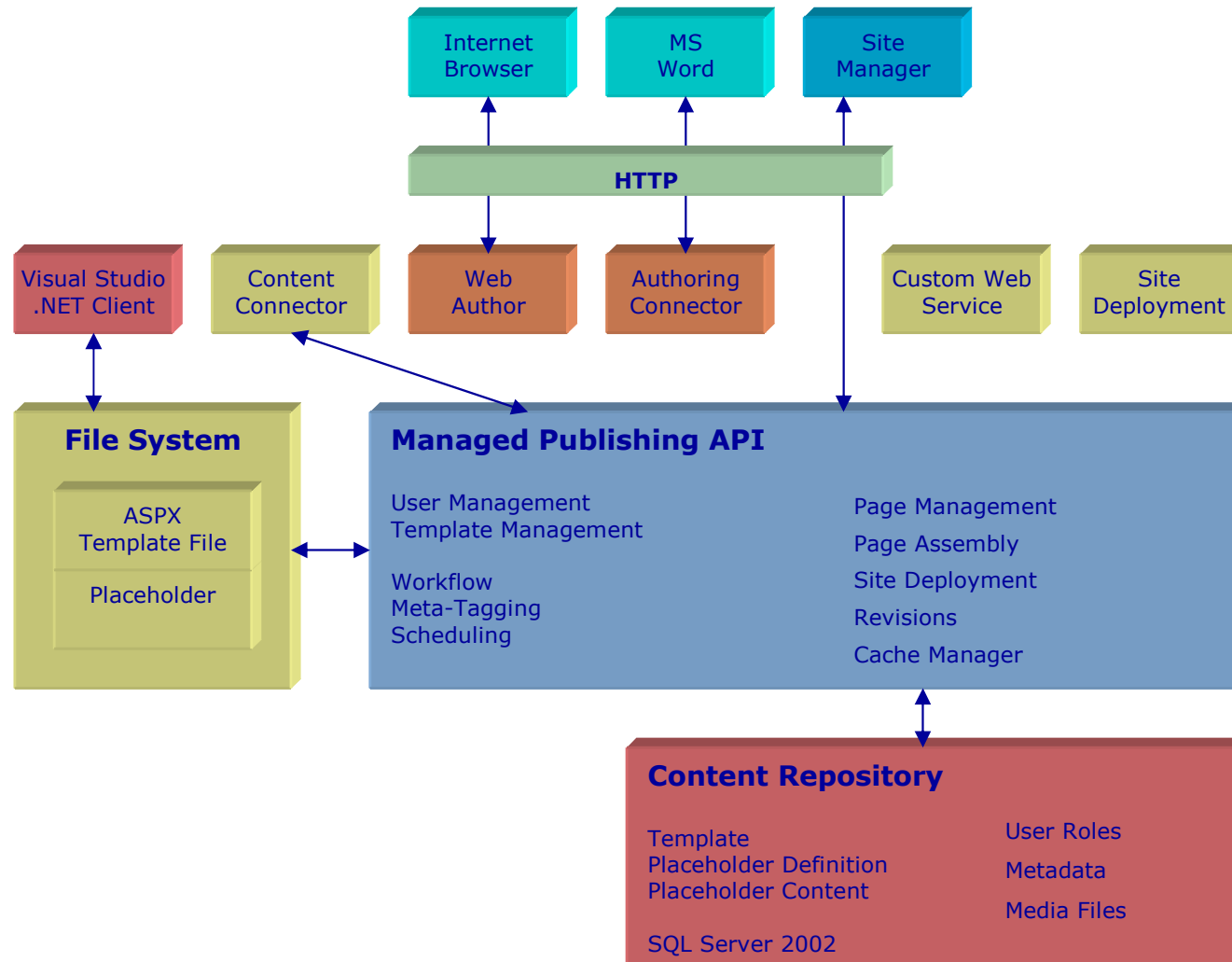
`/artikel/zitate/zitat ==> /artikel[autor='Prof. Jones']`

Content-Management-Systeme: Produkte

Microsoft Content Management Server

- Web-Content Management System
- Unterstützung von Workflows
- Vier Benutzergruppen
 - (Seiten-Entwickler, -Administratoren, -Manager, - Autoren)
- mittels Templates „frei“ definierbare Benutzerinterfaces
- Einbindung von Web-Services
- dynamisch erzeugbare *Sitemaps*

MS CM-Server - Architektur



Zope

- Web-Content Management System
- von Digital Creations entwickelt
- seit 1998 OpenSource-Projekt (Lizenz- und gebührenfrei)
- große Entwicklergemeinschaft („Zope Products“)

- mit Python / C programmiert
- eigene Objektdatenbank (ZODB)
- eigener Web-Server
- Workflow-Komponente (Add-On)
- *kein* WYSIWYG-Editor (DTML/HTML)

- Root Folder
- Control_Panel
- Examples
- QuickStart
- Test
- WiFiSNet
- acl_users
- temp_folder
- Zope Corporation
- Refresh

Folder at / Help!

Select type to add... Add

Type	Name	Size	Last Modified
<input type="checkbox"/>	Control_Panel (Control Panel)		2003-01-16 13:00
<input type="checkbox"/>	Examples (Example Applications)		2003-01-16 12:59
<input type="checkbox"/>	QuickStart (Zope QuickStart Outline)		2001-01-20 22:39
<input type="checkbox"/>	Test (Test)		2003-01-16 15:38
<input type="checkbox"/>	WiFiSNet (Homepage des Wireless FileSharingNet)		2003-01-16 15:55
<input type="checkbox"/>	acl_users (User Folder)		1998-12-30 03:23
<input type="checkbox"/>	browser_id_manager (Browser Id Manager)		2003-01-16 12:59
<input type="checkbox"/>	index_html	1 Kb	2001-01-20 22:39
<input type="checkbox"/>	session_data_manager (Session Data Manager)		2003-01-16 12:59
<input type="checkbox"/>	standard_error_message (Standard Error Message)	1 Kb	2001-01-20 22:39
<input type="checkbox"/>	standard_html_footer (Standard Html Footer)	1 Kb	2001-01-20 22:39
<input type="checkbox"/>	standard_html_header (Standard Html Header)	1 Kb	2001-01-20 22:39
<input type="checkbox"/>	standard_template.pt	1 Kb	2003-01-16 13:00
<input type="checkbox"/>	temp_folder (Temporary Folder)		2003-01-16 15:54

Rename Cut Copy Delete Import/Export Select All

Zope (3)

DTML-Datei

```
<dtml-var standard_html_header>
<p><dtml-var logo></p>
<p>Willkommen!</p>
<dtml-var standard_html_footer>
```

Generierte HTML-Datei

```
<html>
<body bgcolor="#FFFFFF" text="Gray">
<p></p>
<p>Willkommen!</p>
<hr>
<p>Letzte Änderung 10.03.2003</p>
</body>
</html>
```

Zusammenfassung

- Content Management Systeme (Kennzeichnen, Architektur)
- XML-Anfragesprachen (XML-QL, XXL, XQL, XIRQL)
- Content Management Systeme - Produkte

Ausblick

- (W-) CMS versprechen erhebliche Kosteneinsparungen
- Internetpräsenz heutzutage fast ein MUSS
- unüberschaubare Anzahl an Produkten
- wachsende Bedeutung von XML bei (W-) CMS
(ICE-Protokoll als Standard bei W3C vorgeschlagen)

Fragen?

