

Kapitel 1 – Überblick Content Management und Digitale Bibliotheken

Prof. Dr.-Ing. Stefan Deßloch
AG Heterogene Informationssysteme
Geb. 36, Raum 329
Tel. 0631/205 3275
dessloch@informatik.uni-kl.de

Digitale Bibliotheken und
Content Management

1

Überblick

- Was ist Content?
 - Daten, Dokumente, Multimedia-Objekte, ... zur Bereitstellung in Rechnernetzen
 - Content wird publiziert
 - Datenerfasser: Redakteure, Autoren
- Was ist Content einer Digitalen Bibliothek?
 - Dokumente, Multimedia-Objekte, ... von bleibendem Wert
 - Content wird archiviert
 - Nutzer: Leser, Kunden

Content

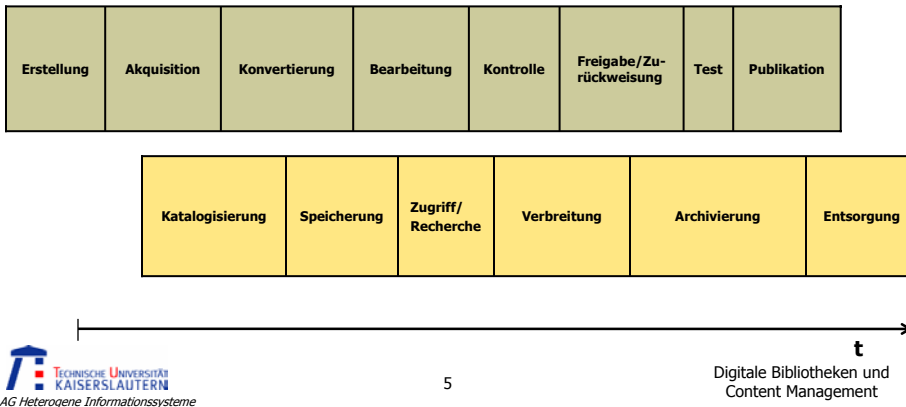
- Informationen, die in LANs oder WANs "veröffentlicht" (bereitgestellt) werden sollen
 - strukturiert, semi-strukturiert, unstrukturiert
 - > 85% des Content eines Unternehmens befindet sich ausserhalb von DB
 - Dateisysteme, spezialisierte Audio/Video – Systeme
- Beispiele
 - Rechnungen, Berichte, ...
 - gescannte Papier- oder Faxdokument
 - Strukturierte Daten in ERP, CRM
 - e-mail
 - Bürodokumente, Korrespondenz
 - Audio, Video, Photos
 - Web Content

Content Management

- Prozess der Verwaltung des bereitzustellenden Contents im LAN oder WAN
 - insbesondere Verwaltung multimedialer Dokumente
 - Speziell: Web Content Management: Zur Bereitstellung im Web
 - Noch spezieller: Verwaltung dynamischer Web-Seiten (gespeist aus Datenbanken)
 - Noch eingeschränkter: Autorenwerkzeug für dynamische Web-Seiten
 - Hier: Verwaltung multimedialer Dokumente (Content) zur Bereitstellung im LAN / WAN
- Was heisst Verwaltung?
 - Erstellen, Bearbeiten, Freigeben, Speichern, Suchen, Archiviere, ...
 - Wesentliche Aspekte
 - Extraktion/Aufbau von Metadaten zur weiteren Beschreibung von Dokumenten
 - Suche auf Metadaten und Inhalten

Content Management: Life Cycle

- Ebene 1
 - Zusammentragen, Gewinnen, Redaktionelles Arbeiten
- Ebene 2
 - Verwaltung, Suche, Archivierung

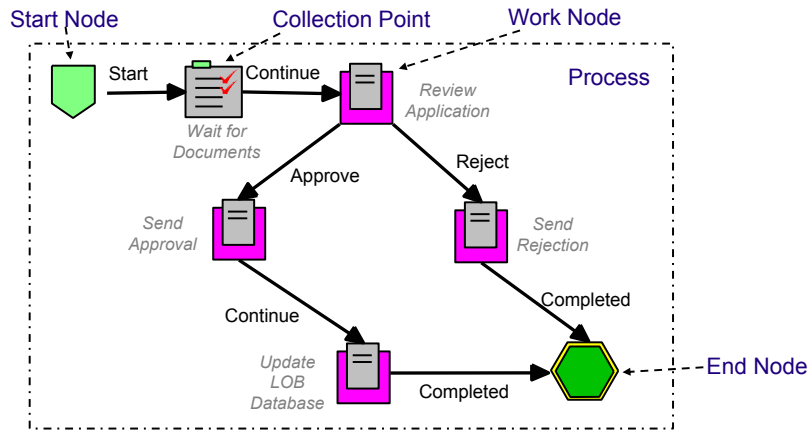


Document Management

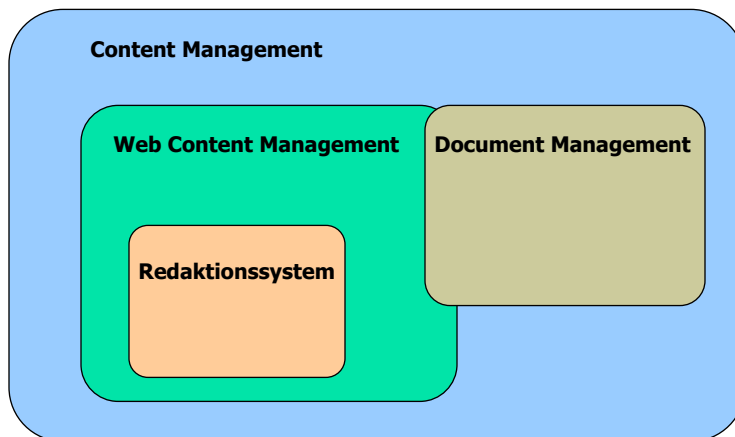
- Erstellung und Verwaltung von multimedialen Dokumenten
- Abbildung des Workflows zur Bearbeitung der Dokumente
 - document routing
- Normalerweise abgeschlossenes System in einem Unternehmen, kein E-Commerce / WAN-Szenario
- Vorteile gegenüber Web Content Management Systemen: Document Management Systeme verwalten oft auch den Content (die Dokumente)
- Dokument: oft Bürodokument mit wenigen, standardisierten Metadaten
- Dokument in Digitalen Bibliotheken: sehr stark strukturierte, dokumenttypabhängige Metadaten

Document Routing

- Workflow management als integraler Bestandteil



Content Management und Document Management



Enterprise Content Management

- Prämisse: alle Arten von (unstrukturiertem) Content sollen durch ein Repository verwaltet werden
 - unabhängig von einzelnen Applikationen, die den Inhalt nutzen, auf ihn zugreifen
 - siehe: Entwicklung von Datenbanksystemen
- Verwaltung und Zugriff
 - Speicherung, Suche
 - Zugriffskontrolle
 - Versionierung, Check-in/check-out
 - Warehousing vs. Föderationsansatz ("virtuelle" Integration)
- "Intelligence"
 - Analyse und Miningverfahren für unstrukturierte Inhalte
- Information Integration
 - Integration mit strukturierten Inhalten

Multimedia-Datenbanken

- Erweiterung der Funktionalität von Datenbanken zur Verwaltung multimediale Objekte
 - Suche nach Ähnlichkeit
 - spezielle Zugriffspfade
 - Speicherverwaltung
 - große Objekte
 - Auslieferung der Daten
- ... unter Beibehaltung der bekannten Vorteile von DBS
 - Datenmodellierung, Datenunabhängigkeit
 - Anfragesprachen, inhaltsorientierte Suche
 - Transaktionen
 - ...

XML-Datenbanken

- Verwaltung von Daten und Dokumenten, die in XML beschrieben sind
- Objektrelationale Datenbanksysteme mit
 - Datentyp XML
 - Import- / Export-Format XML (shreddern der Dokumente)
- XML-Datenbanksysteme
- XML-Suchmaschinen auf Dokumenten-Management-Systemen

Content / Document Management in XML-DB?

- Spezialisierte Content Management Systeme sind eher Redaktionssysteme
- Nur Verwaltung der Metadaten eines Dokuments in relationaler Datenbank
- Geeignet als Plattform: XML-Datenbanksystem zur Verwaltung der Daten und Dokumente, Suche nach Dokumenten, ...

Klassische Bibliothek

- Sammelt, stellt bereit und archiviert langfristig Dokumente von bleibendem Wert
- Erfasst Metadaten und macht diese für den Leser recherchierbar
- Ist Schnittstelle des Nutzers zu allen Publikationen, Leser muss nicht jeden einzelnen Verlag kontaktieren

Informationsflut

- Wissenschaftliche Zeitschriften
 - 1951: 10.000
 - 1993: 140.000
 - teilweise nur 100 Abonnenten, Kosten pro Jahr pro Abo teilweise 10.000 Euro

Probleme

- Wie soll man die relevante Literatur noch finden?
- Wie soll man die Literatur noch bezahlen? (Uni KL: Zeitschriftenetat der Informatik sinkt, Zeitschriftenpreise steigen weit über der Inflationsrate) also digital?

Digitale Bibliotheken

- Digitale Bibliothek ist
 - Sammlung von Dokumenten mit bleibendem Wert,
 - die erschlossen werden mit Metadaten,
 - langfristig zitierbar bleiben müssen,
 - einer Versionierung unterliegen können, aber selbst als Dokument unverändert bleiben,
 - als digitale Ware angesehen werden können.
- Digitale Bibliothek ist
 - ein Softwaresystem zur Erzeugung, Bereitstellung, Erschließung, Speicherung, Verteilung, Suche, Darstellung, Nutzung, Archivierung solcher Dokumente,
 - das weltweit verteilt sein kann, und aus Autoren, Anbietern (Verlagen), Vermittlern (Bibliotheken) und Nutzern besteht.
- Nicht nur wissenschaftliche Literatur: Bürodokumente, Videos, mp3-Files

Content Management vs. Digitale Bibliothek (1)

- Content Management - Schwerpunkte
 - Verwaltung der Inhalte von kurzlebigen Web-Präsentationen
 - Keine kundenseitige Content-Verwaltung vorgesehen
- Digitale Bibliotheken
 - Langfristige Verwaltung von Metadaten und Dokumenten
 - Anbieterseitige und kundenseitige Content-Verwaltung

Content Management vs. Digitale Bibliothek (2)

- Unterstützung der Phasen des CM Life Cycles mit verschiedener Gewichtung
- Content Management
 - Erstellen und Präsentieren von Content
 - Autor, Verlag
 - Im Content Management Life Cycle Betonung der oberen Ebene
- Digitale Bibliothek
 - Sammeln, Archivieren, Recherchieren und Nutzen von Content
 - Bibliothek, Leser
 - Im Content Management Life Cycle Fokussierung auf die untere Ebene
- Ein Content Management System kann auch Digitale Bibliotheken verwalten