

Der Dokumentenserver DOL

<http://dol.uni-leipzig.de>

D. Sosna

Inst. für Informatik, Abt. Datenbanken (Prof. Rahm)

Inhalt:

- Ziele und Konzept
- Workflow zum Einbringen eines Dokuments
- Datenausgabe, Metadatenexport
- Akzeptanz
- Weiterentwicklungen

Entwicklungsziele:

- Dauerhaftes Vorhalten der bibliographischen Angaben und der Dokumente in hoher Qualität
Bereitstellen einer stabilen Adresse
- Komfortabler Zugriff über das Internet
- Dezentrale Verwaltung der Dokumente
Flexibilität bezüglich der Anpassung an Verwaltungsstrukturen
- Weitgehende Nutzung existierender Software

Grundkonzept:

- Dokumente in mehreren Formaten im Dateisystem
- DBVS zur Metadatenverwaltung und Anfragebeantwortung
- Weitgehende Automatisierung der Abläufe mit Bearbeiterkontrolle
- Nutzung erprobter Software für Teilprozesse
- Konfigurierbarkeit mittels XML-Dateien

Workflow zum Einbringen eines Dokuments

Nutzer: Aufruf WEB-Seite, Ausfüllen Formular mit bibl. Angaben
Absenden der Einverständniserklärung

System: Konvertierungen, Eintragungen, Protokollierung,
Nachricht an Bearbeiter

Bearbeiter: Plausibilitätskontrolle → Status *plausibel*.

Nach Eingang der Einverständniserklärung:

Freischalten des Dokuments → Status *gültig*.

Hintergrundaktivitäten: dokumentbezogene Aufzeichnung aller Aktionen
zeitgesteuerte Mahnungen

Suche:

- nach bibliographischen Angaben (Metadaten)
- Volltextsuche (mit Ranking)
- Navigation (Organisationsstruktur, Dokumenttyp, Zeit, ...)

Ausgabemöglichkeiten:

- Metadaten (als HTML-Texte)
- Seitenvorschau (als Bilder)
- Dokumente als PS- und PDF-Datei (und als Textdatei)
- Anfrageschnittstelle zur Einbindung in WWW-Seiten
- Metadaten und Volltexte an Suchmaschinen

Anschluß an Datenverbände (Protokolle):

- Dienst-Protokoll → NCSTRL
- DublinCore, eingebettet in WEB-Seiten
- Crawler-Interface → Google, AltaVista
- Weitere Interfaces bei Bedarf

Statistik:

Leipzig	Dokumente	432 + 10 + 8
	Volumen	2,5 GByte
	Zugriffe	220 000 / Jahr
	Probleme	6,4%, autom. Konv. PS → PDF
Stanford	Dokumente	598
	Volumen	620 MByte
	Zugriffe	300 000 / Jahr
	Probleme	4,3%, autom. Konv. PS → PDF

Software:

- htDig (Volltext-Indexierung)
- pscript (Konvertierung, Seitenvorschau)
- Relationales DBVS mit SQL und JDBC
- WEB-Server, JAVA
- eigene JAVA-Programme (14 000 Zeilen)
- Betriebssystem UNIX

Rollen:

- Entwickler - Admin - Bearbeiter

Möglichkeiten der Weiterentwicklung

S.Melnik: Übergabe in ein Open-Source-Projekt

- + Multimediale Dokumente
- + Dokumente aus mehreren Bestandteilen
- + Verbesserte Systempflege
- ? Authentizität der Dokumente, Signaturen
- ? Multilinguale Dokumente
- ? Verteiltes System, Datenreplikation
- ? OCR Texterkennung

Weitere Informationen:

- <http://dol.uni-leipzig.de>
- Rahm, Schwipper, Sosna: Dienst für Online-Dokumente gestartet
<http://dol.uni-leipzig.de/pub/1998-86>
- Melnik, Rahm, Sosna: DOL - An Interoperable Document Server
<http://dol.uni-leipzig.de/pub/2001-27>
- Kontakt:

rahm@informatik.uni-leipzig.de

<http://dbs.uni-leipzig.de>