

CAP7: Suche in strukturierten (XML-)Dokumenten

Slide 1

Norbert Fuhr

Kai Großjohann

Stefan Kokkelink

19. September 2000

Motivation

- Kombination von inhaltsorientierter und strukturorientierter Suche
- Einfacher Ansatz: Dokument ist Menge von Feldern
(Wird bspw. verwendet von Harvest)
- Einfacher Ansatz reicht nicht aus
Dokument mit mehreren Autoren, speichere Email-Adresse zu jedem Autor: Zuordnung zwischen Autornamen und Email-Adressen geht verloren.

Slide 2

Lösung

- Verwende XML als Dokumentenformat
- Entwickle Anfragesprache (XIRQL) für XML-Dokumente mit

Unterstützung für

- inhaltsorientierte Suche
- strukturorientierte Suche

(XIRQL Zwischensprache analog zu SQL bei RDBMS.)

- Implementiere System (DeslRe) für Indexierung und Retrieval

Slide 3 **Anwendungsbeispiel MathNet**

- Migration von SOIF nach RDF/XML.
- Umfasst u.a. MPRESS: 40.000 Metadaten-Records in SOIF; davon ca. 50% gemäß Dublin Core
- Konvertierung dieser Daten nach RDF
- Suche auf XML-Repräsentation der RDF-Daten

Slide 4 • Partner-Projekte zu CAP7 zum Erstellen angereicherter Metadaten

Systemarchitektur

Slide 5

