

Anreicherung und Empfehlungen von Medieninhalten auf Basis von Linked Data



#DKT17

12. Oktober 2017 | Radoslaw Oldakowski

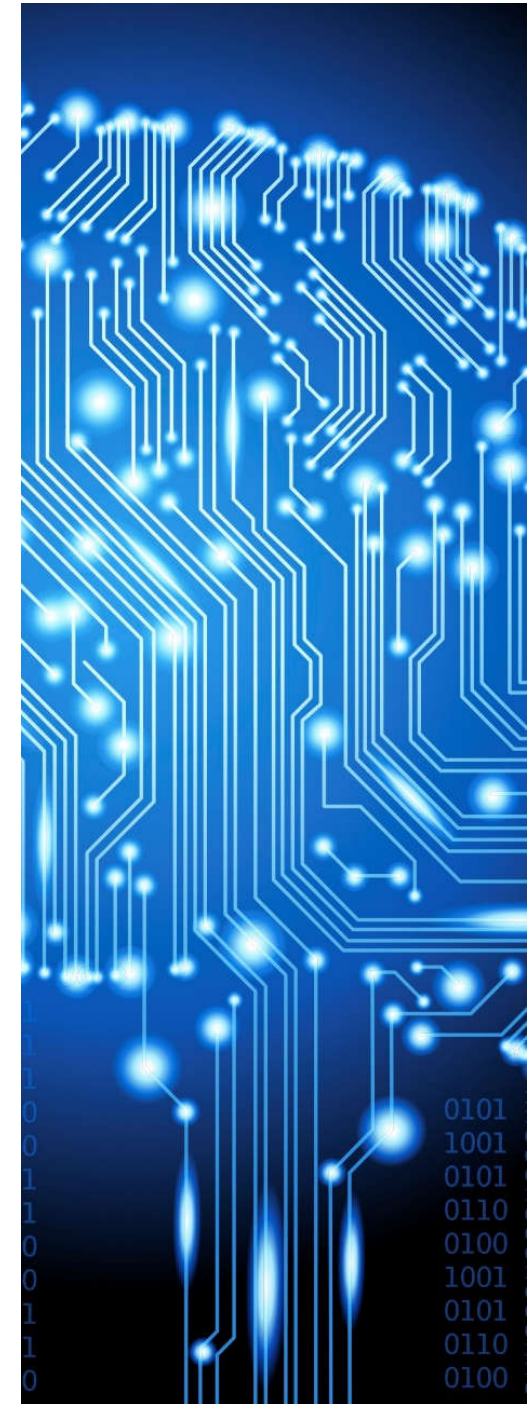


Agenda

- Kurze Vorstellung der Condat Smart Media Engine
- Ausgewählte Projektbeispiele:
 - zdf.de, ZDF-Mediathek
 - Öffentlich-rechtlicher Sender:
Programmempfehlungen
 - GVL: Datenbereinigung und Dublettensuche

Condat Smart Media Engine

- Analyse-, Such- und Recommendation-Server, basierend auf:
 - Semantischen Technologien
 - Frei verfügbaren Quellen (Linked Data)
- Allgemeine und domänenspezifische Ontologien
- „Semantische Fingerabdrücke“ von Kunden-Datenobjekten
- Cross-medial und cross-lingual
- REST-Schnittstellen für alle alle Funktionalitäten
- Diverse Tuning-Möglichkeiten

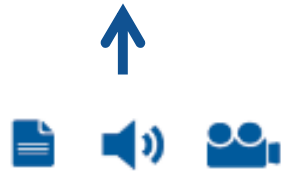
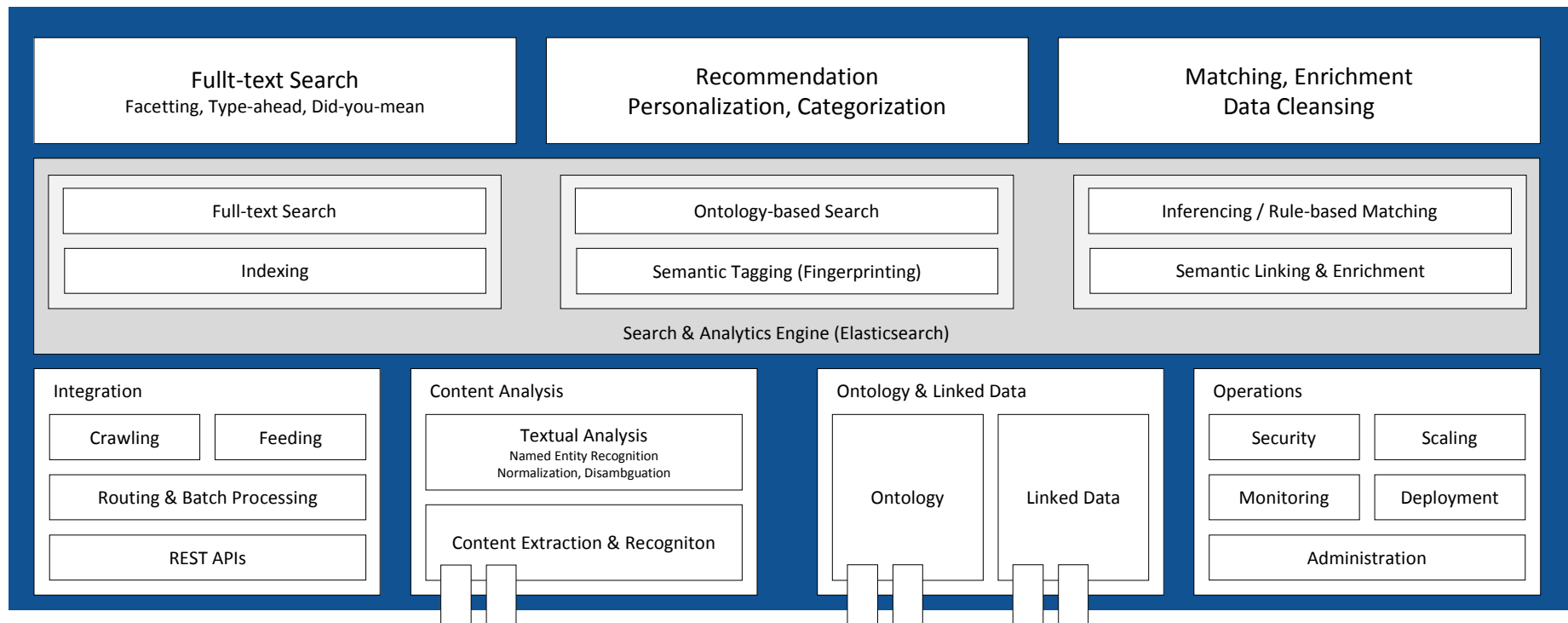


CSME - Architektur

Search

Recommend

Discover



CSME Hauptanwendungsfälle

- Semantische Suche nach Content
 - Volltextsuche (mit Suchraumerweiterung)
 - Objekttyp-übergreifend (Metadaten, Text, Audio, Video)
 - Sprachen-übergreifend
- Ähnlichkeitssuche
 - Objekttyp-übergreifend
 - Sprachen-übergreifend
 - Spezialfall Dublettensuche
- Kategorisierung von Content („Themen-Kanäle“)
- Kategorie-basierte Suche
- Nutzerprofil-basierte Varianten („Nutzer-Kanäle“)



zdf.de, ZDF-Mediathek

- Ohne CSME, aber viele Lösungsansätze übernommen
- Suche (Volltext, Typeahead, Did-you-mean, facettiert, etc.)
- Item-basierte und personalisierte Empfehlungen
- Weitere Recommendation-Verfahren auf Basis von Elasticsearch
 - Collaborative Filtering
 - Cluster-basierte Recommendations
- Services sehr stark parametrisierbar:
 - Attribut- bzw. Item-Boosting (z.B. Aktualitätsfunktion)
 - Umfangreiche Filter
 - Fachliches Ergebnis-Tuning



Öffentlich-rechtlicher Sender: Programmempfehlungen

- Vollständig auf CSME-basis entwickelt
- Programmempfehlungen für Zuschauer
 - item-to-item, personalisiert
 - DbPedia als Wissenmodell (automatische Updates)
 - 5 Sprachen (fr, de, en, es, pl)
 - Cross-medial, z.B. Text → Video
 - Kanalempfehlungen / Mood-basierte Empfehlungen
 - Manuelle Kanaldefinitionen
 - Semi-automatische Kanaldefinition



GVL: Datenbereinigung und Dublettensuche

- Große Anzahl an Dubletten, stark heterogene und zum Teil falsche Schreibweisen von Titeln, Künstlernamen, etc.
- Textbereinigung nach speziellen Regeln
 - Umwandlung von Großschrift
 - Normierung von Trennzeichen, Werkverzeichnissen, Angaben aus dem Klassikbereich (Dur-, Mol-, op., etc.)
 - Künstlersuche, Namenskorrektur und Vereinheitlichung unter Anbindung domänenspezifischer Ontologien: **discogs**, **MusicBrainz**
- Anreicherung von Produktinformationen (z.B. Mitwirkende)
- Maßgeschneiderte Dublettensuchverfahren



Danke!

Fragen?

Kontakt:

www.condat.de

rol@condat.de

www.condat.de

