OCR Renderfarmen und TEI

Christian Mahnke
SUB Göttingen

Inhalt

- Grundlagen
  - OCR Server
  - Volltextkodierung
  - Integration in Workflowsystem
- Umsetzung
  - Server
  - Formate
  - Präsentation
- Ausblick

Grundlagen

OCR für die Massendigitalisierung

Bisherige Vorgehensweise

- Rohdaten in proprietären Formaten („Stapel“)
- Ergebnisse schwer integrierbar
- Manuell via Desktopsoftware (wenig Automatismen)
- Oder: Integration als Programmbibliothek (Prozessmanagement nicht integriert)

OCR für die Massendigitalisierung

Wirkliche Anforderungen

- Skalierbarkeit
- Management des Prozesses in größeren Einheiten
- Integrierbarkeit in bestehende Infrastruktur
- Hoher Grad an Automatisierung
- Massenverarbeitung

Volltextkodierung

- XML basiert
- Offenes Format
- Hohe Flexibilität für verschiedenen Anwendungsszenarien
- Bestehende Vokabularien nutzen
- Breite Community – hohes Potential für Nachnutzung

Integration in den Digitalisierungsworkflow

- Keine manuelle Interaktion im Regelfall
- Kein Trainingsaufwand für Personal
- Priorisierung (Echtzeit OCR für Metadateneditor vs. OCR als Workflowschritt)
- Kapselung der Details der Orchestrierung (Servicekonzept)
- Steuerbar durch Software (API / Webservices)

Umsetzung

Server (Soft- und Hardware)

- Software
  - Abbyy Recognition Server 2.0
  - Teilung zwischen Management- und Processingknoten
  - Keine seitenbasierte Lizenzierung
- Hardware
  - Cluster aus Bladeservern
  - Derzeit 16 CPU Kerne

Server (Kommunikation)

- WebDAV Schnittstelle
  - HTTP basiert (keine Probleme mit Firewalls)
  - Viele Implementierungen des Protokolls
  - Nutzbar als Netzlaufwerk
- Steuerung
  - XML Tickets (erzeugt durch Programmbibliothek)

Formate

- TEI basiert
- Indexformat für einfache Indexierung
- Zukünftig: Volltextformat für Nachnutzung
  - Geeignet für Erweiterungen wie Annotationen
  - Nutzbar für elektronische Editionen

Produktion

- Derzeit: Rekursives abarbeiten von Verzeichnisbäumen
- Zukünftig: Prozesssteuerung als Teil von Goobi
- Indexierung beim Import in das DMS

Präsentation

- Volltexte für die Suche und Wortkoordinaten für die Darstellung
- Index (Lucene) wird durch Typo3 (CMS) abgefragt
- Zukünftig: Darstellung in der Oberfläche
- Beispiele

Demo 1


Loginstatus:
SUB Göttingen
logout
Suchen Sie hier im Archiv nach
Direktsuche
DigiZeitschriften
DAS DEUTSCHE DIGITALE ZEITSCHRIFTENARCHIV
START
SUCHE
ZEITSCHRIFTEN
OPEN ACCESS
INFORMATION
FAQ
KONTAKT
english
Sie sind hier: Suche > Suche
Suche | Hilfe zur Suche
Suche
hoffnung
6127 Treffer
1|2|3|4|5|6|7|8|9|10|11
1 Die Aufgabe christlicher Eschatologie
Autor: Pannenberg, Wolfhart
...ein Leben ohne Hoffnung , und ein... ...ein Leben ohne Hoffnung leben wir eigentlich... ...Diesseits«2 . Die Hoffnung auf ein Leben... ...eschato - logischen Hoffnung der Christen ,... [mehr]
Zeitschrift für Theologie und Kirche : ZThK / Zeitschriftenband 92 / Artikel
2 Glaube auf Hoffnung - Hoffnung für Japan? Missionarische Verkündigung im Dienst am Menschen heute
Autor: Rosenkranz, Gerhard
Suchen in
Volltext und Metadaten
Volltext
Metadaten
Sortieren nach
Index
Autor
Erscheinungsjahr
Importdatum
Titel
Treffer in Sammlungen
Anglistik (28)
Buch- / Bibliothekswesen (358)
Erziehungswissenschaften (622)
Geowissenschaften (329)

Demo 2


Loginstatus:
SUB Göttingen
logout
Suchen Sie hier im Archiv nach
Direktsuche
DigiZeitschriften
DAS DEUTSCHE DIGITALE ZEITSCHRIFTENARCHIV
START
SUCHE
ZEITSCHRIFTEN
OPEN ACCESS
INFORMATION
FAQ
KONTAKT
english
Sie sind hier: Seitenansicht
Inhaltsverzeichnis | Bibliografische Info | Seitenansicht | DFG Viewer
Die Aufgabe christlicher Eschatologie
Zeitschrift für Theologie und Kirche : ZThK / Zeitschriftenband / Artikel
http://resolver.sub.uni-goettingen.de/purl?PPN507831411_1995_0092
78 : 72
72
Wolfhart Pannenberg
sensiblen Menschen nur dann möglich, wenn er oder sie an eine Lebenserfüllung über dieses irdische Dasein hinaus glauben können. Sonst bleiben nur Betäubung durch die Betriebsamkeit der Arbeit oder die Ablenkungen der Freizeit oder aber ein Leben ohne Hoffnung, und ein Leben ohne Hoffnung leben wir eigentlich auch da, wo wir es in der einen oder andern Form der Betäubung unserer tieferen Ängste oder unserer inneren Leere verbringen oder auch in unrea-

Ausblick

- Goobi
  - Anbindung des GBV OCR Clusters
- IMPACT
  - EU Projekt zur Verbesserung von Fraktur OCR
- TextGrid
  - Integration von OCRopus

Fragen?

mahnke@sub.uni-goettingen.de