# Sitzungsprotokoll

Sitzung am: 04.08.2015 um: 10:00 Uhr

Ort: Büro Manuel Burghardt

Teilnehmer: Manuel Burghardt, Katia Buchhop, Thomas Schmidt, Florian Fuchs

Thema: Zwischenstandsbericht, Übergabe Analyse- und Metadatenreport, Besprechung Server

Fragen:

Verlauf:

Debatte über Datenbank:

Java-Server

Solr als IR-Instanz für die Indexierung und Suche

eXist als gute Lösung akzeptiert. (Kosten? )

Solr und eXist zusammen hätte doppelte Funktionalitäten…

Solr bietet weitere Suchfunktionen wie Rechtschreibfehlerberücksichtigung und Stemming; Felder für jedes Metadatum.

eXist ist ein Server, auf dem man die Daten hosten kann.

Im Grunde kann man alle Daten über Solr verwalten.

Datenformat: Alle bestehenden Daten (Metadaten, OCR/OMR-Outputs) in XML transformieren. Diese werden dann von Solr indexiert.

Durchführung: Konzeption lokal mittels Tomcat Server. Später Server über MI oder RZ.

Besprechung der Tool-Analyse:

Problem: Die Grounded Truth Dokumente wurden je nur von einer Person transkribiert und daher nicht auf Fehler oder Abweichungen geprüft. Frage: wie wurde mit Sonderfällen wie Bindestrichen umgeht.

Vorstellen der Analyse-Ergebnisse: Wichtigster Wert ist der Global Error Rate: Bei ABBYY müssen im Endergebnis 30% angepasst werden.

To Do:

Berichte in den nächsten zwei Tagen fertigstellen und an Burghardt senden.