# Sitzungsprotokoll

Sitzung am: 22.05.2015 um: 9.00Uhr

Ort: Informationszentrum

Teilnehmer: Miriam Nickl, Katia Buchhop, Thomas Schmidt, Florian Fuchs

Thema: Austauschen der Zwischenergebnisse; Vorbereiten der Präsentation

Agenda:

* Präsentation der Rechercheergebnisse OCR
* Präsentation der Rechercheergebnisse ONR
* Präsentation der Analyse der Excel-Tabellen
* Erstellen der Zwischenpräsentation

Verlauf:

* Besprechung Metadaten:
  + Untersuchen der verschiedenen Ausprägungen
  + Diskussion von Ungereimtheiten im Datensatz
  + 🡪 ToDo: Liste der möglichen Ausprägungen und Häufigkeit
* Besprechung: OCR
  + Bisherige Ergebnisse: Paper über Fehler und Fehlerarten/ Maße
  + Tabelle für die Bewertung der Tool
* Besprechung ONR
  + Alle Marktführer kostenpflichtig (Erkennungsrate ca 90% (gedruckte Noten))
  + SharpEye Music Reader kann Musik und Text lesen; gute Performance; AUsgeb MusicXML
* Besprechung Workflow und Probleme
  + Frage: Namespaces? Wie führt man die Datensätze zusammen?
  + Idealfall: User muss nur das Blatt in einem Programm öffnen und ONR/ OCR werden von dort aus gestartet…
* Planung Präsentation
  + Vorstellen der Verschiedenen Baustellen des Problems
  + Erklären auf welche Probleme wir uns fokussieren (und warum)
* Idee:
  + WorkflowErgebinss Upload
  + 🡪Tool zum Editieren der OCR-Texte
* ToDo:
  + 🡪WorkFlow und Datenbank
  + 🡪Dann Frontend
  + 🡪Dann Visualisierung und Analyse
* Evaluation der Tool:
  + Tool zur automatischen Evaluation verwenden.
  + Ausgewählte Dokumente händisch auszeichnen und OCR vergleichen
* Idee:
  + Automatisiertes Zählen von Wörtern, Texten, Noten, Tonart, Takt, und Strophen
  + Auswerten unter bestimmten Kriterien wie z.B bestimmter Sangesort o.ä.
  + 🡪 Möglich sobald die Texte maschinenlesbar erfasst sind
  + 🡪 MusicXML bietet viele Werte
  + 🡪 Kulturwissenschaftler fragen was noch interessant wäre
  + 🡪 Suche nach Musikläufen???
* Gesprächsbedarf:
  + Suche?
  + Searchengine oder direkt auf Datenbank?
* Frage: Wie gehen wir mit fehlerhaften Datensätzen um?
* Frage: Wie normalisieren wir die Daten für das Info Retrival?
* Frage: BitBucket nur 5 User möglich? Wie mit Stakeholder und Projektbetreuer?

Zusammenfassung:

Frontend:

3 Komponenten

* Import
* Metadaten Auszeichnungstool
* Visualisierung

Nächste Schritte:

* Scanns erstellen
* Testkorpus händisch annotieren
* Test mit verschiedenen Tools + automatische Auswertung
* Frage: Wann ist die nächste Sitzung und wann welche Präsentation?

Aufgabenverteilung (zu Montag, 23.05.2015):

* Miriam: Gültigkeit der Metadatenwerte prüfen (Gibt es die Städte noch?)
* Katia: Ausprägungen und Häufigkeiten der Metadaten analysieren; Dokumente für Doku suchen und in Dropbox speichern
* Thomas: OCR-Übersichtstabelle (für Mac geeignet? ); (ggf. Kurztest des Preprosessing )
* Florian: OMR ausprobieren und Übersichtstabelle erstellen (Verfügbarkeit von Testversionen beachten, für Mac geeignet?)
* Miriam u. Katia Vorbereitung der Projektmanagement-Dokumente am Do, 28.5.2015