|  |  |
| --- | --- |
| Projektstatus Report | Status gesamt:  **grün** |

# Projekt: BibTexKonverter, Gruppe: David Konieczny, Lars Lafleur, Constantin Schulz

10.07.2024

|  |  |
| --- | --- |
| **Erfolge**, behobene Risiken oder Probleme (Farbcodierung = wie wichtig war Erledigung für Gesamterfolg) | * Erfolg 1 (wichtiger Milestone): Mehrere String-Referenz können verarbeitet werden. OCR-Erkennung implementiert, Sprachauswahl für OCR durch Dropdown-Menü * Erfolg 2 ( wichtiger Milestone ): Logik für Evaluationsmetrik fertiggestellt. Anwendung der Metriken auf Testfälle steht noch aus. * Erfolg 3 (wichtiger Milestone): Folgendes Felder können bereits erfolgreich extrahiert werden: authors, editors, year, volume, number, edition, pages, url, doi, month, address, publisher, school. Zusätzlich noch titel, booktitle und journal * Erfolg 4 ( wichtiger Milestone): Debugging erfolgreich durchgeführt und funktionstüchtiger Prototyp erstellt |
| **Risiken und Probleme** (Farbcodierung = wie groß ist das Risiko für Gesamterfolg) | * Risiko 1 (mittleres Risiko ): Klassifizierer für die Referenzformate (article, book, etc.) benötigen noch weiteres Training für höhere Genauigkeit. * Probleme 1: Pybliometrics, Scholarly, Arxi und Pybliographer können nicht im Rahmen des Praktikums umgesetzt werden und bedarf weiterer Untersuchungen. Abfrage der Informationen eher langsam und nicht komfortabel. * Problem 2: POS-Tagging, um title, booktitle, journal und series zu extrahieren, hat sich als problematisch herausgestellt und kann im Rahmen des Praktikums nicht mehr umgesetzt werden. Die Extraktion dieser Felder erfolgte nun ohne Zuhilfenahme der Tagger. |
| **Geplante nächste Aktivitäten** (Farbcodierung = wie wichtig ist Erledigung für Gesamterfolg) | * Aktivität 1 (wichtiger Aktivität): Verwendung der generierten Trainingsdaten für das ML-Modell: Liefert das Modell anhand der selbst erzeugten Trainingsdaten zufriedenstellende Ergebnisse? * Aktivität 2 (wichtiger Aktivität): Optimierung der Klassifizierer der Referenztypen * Aktivität 3 ( mittelmäßig wichtige Aktivität ): Optimierung der virtuellen Entwicklungsumgebung zur Installation des Projektes * Aktivität 4 ( mittelmäßig wichtige Aktivität ): Finale Evaluierung erstellen und Visualisierung der Ergebnisse beenden * Aktivität 5 ( mittelmäßig wichtige Aktivität ): Fortführung der Dokumentation der bisherigen Methoden und Prozesse * Aktivität 6 ( mittelmäßig wichtige Aktivität ): Genauere Analyse der Testdaten |
| Offene Fragen, generelle Anmerkungen |  |