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

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

|  |  |
| --- | --- |
| **Erfolge**, behobene Risiken oder Probleme (Farbcodierung = wie wichtig war Erledigung für Gesamterfolg) | * Erfolg 1 (wichtiger Milestone): Konvertierung von BibTeX-Code in Literaturstrings in Python (Bibliothek Pybtex) für die Stile Plain, APA und MLA: Die im Web gefundene Sammlung von BibTeX-Code „anthology.bib“ enthält Einträge des Typs Article, Proceedings und Inproceedings, die mithilfe von Pybtex konvertiert werden konnten. Daraus wurden Trainingsdaten mit den Attributen „Literaturstring“, „Literaturtyp“, „Style“ und „Bibtex“ erstellt. Pybtex kann später noch angepasst werden, um auch Referenzen in weiteren Stilen zu erzeugen. * Erfolg 2 (mittelmäßig wichtiger Milestone): <Beschreibung> * Erfolg 3 (unwichtiger Milestone): <Beschreibung> |
| **Risiken und Probleme** (Farbcodierung = wie groß ist das Risiko für Gesamterfolg) | * Risiko 1 (hohes Risiko): <Beschreibung> * Risiko 2 (mittleres Risiko): Bias in den Trainingsdaten: Durch die einheitliche Konvertierung von Bibtex-Code zu Literaturstrings mithilfe einer Bibliothek wird das “Rauschen” in Realbeispielen vernachlässigt. Auch wenn jemand den Stil Apa benutzt, benutzt er vielleicht hier ein Komma statt einem Punkt, der andere klammert das Jahr und der Dritte nicht. Manche schreiben nur die Seitenzahlen, andere setzen davor pages, p., oder pp. Die von uns erzeugten Traningsdaten werden dieses "Rauschen" nicht abbilden. Andererseits sind diese Trainingsdaten mit dem jeweiligen Stil gelabelt, d.h. so kann das Modell eine grobes Muster für den jeweiligen Stil lernen. Das ist der Vorteil zu gescrapten Referenzen, deren Stil ohne weiteres unklar ist.Es wird sich herausstellen, inwieweit sich hieraus Probleme ergeben. * Risiko 3 (geringes Risiko): <Beschreibung> |
| **Geplante nächste Aktivitäten** (Farbcodierung = wie wichtig ist Erledigung für Gesamterfolg) | * Aktivität 1 (wichtige Aktivität): Überprüfung der Qualität der Trainingsdaten (Sind die mit Pybtex erzeugten Literaturstring nach den jeweiligen Richtlinien des Styles formatiert?) * Aktivität 1 (wichtige Aktivität): Recherche von Bibtex-Code für books, inbooks, incollections, phdthesis * Aktivität 2 (mittelmäßig wichtige Aktivität): * Aktivität 3 (wichtiger Aktivität): Verwendung der generierten Trainingsdaten für das ML-Modell: Liefert das Modell anhand der selbst erzeugten Trainingsdaten zufriedenstellende Ergebnisse? |
| Offene Fragen, generelle Anmerkungen | <optional> |

<DATUM>