

Matthias Melcher

KI-basierte Texterkennung für tschechische Quellen (20. Jahrhundert, handgeschrieben)

Call for Expressions of Interest

Je größer die zu verarbeitenden Datenmengen für ein Forschungsprojekt sind, desto wichtiger wird es, handgeschriebene Quellen in maschinenlesbare Texte zu verwandeln. Für zahlreiche Sprachen gibt es für diesen Vorgang bereits ausgefeilte Tools (bspw. als öffentlich zugängliche Texterkennungsmodelle auf *Transkribus*). Auch das Collegium Carolinum hat mit der Erstellung eines KI-Modells für Tschechisch und Slowakisch hier einen großen Fortschritt erzielt – dieses Modell ist jedoch vor allem auf gedruckte Texte ausgelegt. Für handgeschriebene tschechische Texte fehlt nach meinem derzeitigen Wissensstand ein vergleichbares Werkzeug.

Da ich im Rahmen meines Dissertationsprojekts ein kleines Korpus tschechischer handschriftlicher Quellen transkribiere, würde ich mich gerne mit Kolleg:innen in der bohemistischen Community vernetzen, die möglicherweise vor einer ähnlichen Herausforderung stehen.

Alle, die Interesse an einer gemeinschaftlichen Arbeit zur Erstellung eines KI-Modells für handgeschriebene tschechische Quellen des 20. Jahrhunderts haben, sind herzlich eingeladen, sich bei mir zu melden.

Kontakt: Matthias Melcher, matthias.melcher@lmu.de