

Vortrag von Matthias Baumgart, M. Sc. Professur Medieninformatik

Zusammenfassung:

Die Komplexität der Prozesse und Aufgaben im Forschungsmanagement an Hochschulen ist seit einigen Jahren durch die Zunahme an Aufwänden, sowohl auf Seite der Wissenschaftler als auch auf administrativer Ebene gekennzeichnet. Diese liegen dabei insbesondere im Bereich der Texterstellung und -verarbeitung im Bereich der Forschungsanträge, wissenschaftlichen Publikationen sowie weiteren Dokumenten im Umfeld der Wissenschaftskommunikation. Gleichzeitig existieren Methoden und Technologien in den Bereichen des *Information Retrievals*, des *Maschinellen Lernens* und der *Semantischen Technologien*, die grundsätzlich für die qualitative und quantitative Analyse und Bewertung dieser Texte, ihrer Merkmale sowie dazugehöriger Metadaten geeignet sind.

An dieser Stelle setzt diese Arbeit an, in dem sie zunächst mit Fokus auf die Optimierung von digitalen, kognitiven und kommunikativen Aufwänden im Lebenszyklus von öffentlich geförderten Forschungsprojekten die aktuellen Entwicklungen und Technologien eruiert. Diese Betrachtung stellt die Grundlage für die Ableitungen von Kriterien dar, die auf eine Gesamtarchitektur zielen, welche wissenschaftliche Texte mit der Erzielung einer hohen Textqualität annotiert, trainiert und evaluiert. Die Kriterien wurden in Anforderungen überführt, welche die Grundlage für die Konzeptentwicklung im Bereich Datenhandhabung (K_{Daten}), Textanalyse ($K_{\text{Textqualität}}$) und Einschätzung der Textqualität ($K_{\text{Textqualität}}$) darstellten. Die Fusionierung der drei technischen Konzepte zur Gesamtsystemarchitektur und der Ableitung der Systemkomponenten bildet die Grundlage für das Framework zum Management und zur Evaluierung wissenschaftlicher Texte (FELIX).

FELIX dient als prototypisches System für die durchgeführte Studie zur computergestützten Assistenz zur Evaluation wissenschaftlicher Texte. Die hierfür erforderlichen Datenkorpora wurden aus Forschungsanträgen der SMWK Forschungsförderung 2020/2021/2023, Forschungsanträgen im Rahmen der Ausschreibungen des Arbeitskreises E-Learning und aus Publikationen der 17. Nachwuchswissenschaftlerkonferenz aufgebaut. Zur Gewinnung von Erkenntnissen wurde explorativ auf verschiedenste Verfahren aus dem Bereich *Maschinelles Lernen* gesetzt, die Vorhersagewahrscheinlichkeiten für die Bewilligung / Nicht-Bewilligung bzw. Annahme / Ablehnung von Forschungsanträgen und Publikationen ermöglichen.

Die durchgeführten Experimente zielen dabei auf Basis der Extraktion von Textmerkmalen und Meta-Daten auf die formale und statistische Analyse sowie regelbasierte Klassifikation, auf *Low-Level-Klassifikationen* für Textprüfungen (z. B. Lesbarkeitsindizes) und *High-Level-Klassifikationen* für Textprüfungen (z. B. *Sentiment- und Ähnlichkeitsanalyse*, *Deep Learning*) ab.

Zwei Experimente im Kontext *Transfer-Learning* dienen weiterführend der Evaluation der Ergebnisse unter Einsatz unbekannter Daten und der Fusion zu einer additiven Ergebnisformel, deren Grundlage mehr als 68 Tsd. Einzelbeobachtungen in fünf Experimenten waren. Im Ergebnis wurde das vorgestellte System FELIX positiv evaluiert. Es ermöglicht die Analyse wissenschaftlicher Texte und ihrer Metadaten und erzielte regelmäßig Vorhersagewahrscheinlichkeiten von 75 % - 89 %, bei einzelnen Subexperimenten sogar noch deutlich genauere Ergebnisse.

Mit der Konzeption und Evaluierung des Systems FELIX ergeben sich wissenschaftliche und praktische Implikationen, die zur Optimierung von Aufwänden im Forschungsmanagement führen. Dazu zählt die automatische Analyse von Texten zur Identifikation von Mustern, Texten und Korrelationen, die Klassifizierung nach definierten Kriterien (z. B. Fachdisziplin des Antragstellers, Projektvolumen, Lesbarkeit einzelner Kapitel) oder die Vorhersage der Bewilligung auf Basis gelernter Informationen.

Mit dem prototypischem System FELIX lassen sich forschungsbezogene Aufwände durch Methoden des *Maschinellen Lernens*, des *Information Retrievals* und der *Semantischen Technologien* senken, indem die Effizienz des Peer-Review-Prozesses verbessert und somit ein effektiveres Forschungsmanagement erreicht werden kann. Durch den Einsatz einzelner und kombinierter Verfahren aus dem Bereich *Maschinelles Lernen* lassen sich eine Reihe von Kriterien automatisch erfassen, klassifizieren und bewerten. Der Rechenaufwand für den Einsatz des Systems FELIX lässt sich basierend auf einer additiven Ergebnisformel an spezifische Anforderungen anpassen.

Es ist jedoch zu beachten, dass die automatisierte Bewertung von Forschungsanträgen die menschliche Tätigkeit nicht ersetzen kann und sollte. Vielmehr stellt das System ein Assistenzsystem dar, welches Zeit und Kosten sparen und von subjektiven Einflüssen freie Empfehlungen liefern oder als prototypisches System zur Weiterentwicklung in anderen Domänen oder im Rahmen weiterführender Forschungsarbeiten Verwendung finden kann.

Öffentliche Verteidigung im Rahmen des Promotionsverfahrens

“Exploration maschineller Verfahren zur Entwicklung eines methodischen Frameworks zur Evaluierung wissenschaftlicher Texte im Forschungsmanagement”

7. Dezember 2023, 13.00 Uhr

Technische Universität Chemnitz, Straße der Nationen 62, Raum: 1/336 (neu: A12.336)

Alle interessierten Personen sind eingeladen.