

Projektantrag im Programm "e-Research-Technologien" des DFG-Bereichs
"Wissenschaftliche Literaturversorgungs- und Informationssysteme" (LIS)
Interdisziplinäres Infrastrukturvorhaben an der Humboldt-Universität zu Berlin:
Malte Dreyer (Computer- und Medienservice, CMS),
Prof. Dr. Anke Lüdeling (Institut für deutsche Sprache und Linguistik, IdSL)

Daidalos-Projekt

Entwicklung einer AI-gestützten Software zur semantischen Analyse lateinischer und griechischer Texte

Das Daidalos-Projekt bietet eine Anlaufstelle für semantisch orientierte philologische Forschungsfragen, z. B. wie mit Faktualität (Gegensatzpaar: *fabula - probata*, "bloße Erzählung – überprüfte Fakten") in verschiedenen Gattungen umgegangen wird. Es ermöglicht, relevante Texte einheitlich bis auf die Tokenebene zu referenzieren und mit generischen Metadaten (z. B. Entstehungszeit, Autorschaft) zu versehen. Außerdem macht es anfallende Forschungsdaten für die eigene Arbeit (*Identity and Access Management*) wie auch für andere Forschende nachvollziehbar sowie nachnutzbar (Forschungsdatenbank). Im Sinne einer agilen Softwareentwicklung setzen wir hierbei vor allem auf eine enge Zusammenarbeit mit den zukünftigen Nutzenden, die wir zu einer Interessengemeinschaft zusammenführen wollen. Computerlinguistisch gesehen folgen wir der jüngsten Entwicklung der *Distributional Semantic Models* auf der Basis von *Machine Learning* und wollen deren traditionelle Schwächen, z. B. Dekontextualisierung und Nivellierung semantischer Relationen, überwinden. Auf der Grundlage korpusbasierter semantischer Analysen durch den interaktiv visualisierten Output maschineller Lernprozesse können Forschende außerdem bestimmte Themenfelder wie z. B. 'gerechter Krieg' oder 'Weiblichkeit' untersuchen, indem sie verwandte ähnliche Wörter (Synonyme) und typische Verwendungskontexte (Kollokationen) finden. Für den Projektverlauf sind zwei Phasen geplant: In den ersten zwei Jahren soll einerseits der Prototyp für *Machine-Learning*-basierte semantische Textanalysen fertiggestellt, andererseits der Aufbau einer Fachcommunity mithilfe von Workshops, Foren etc. vorangetrieben werden. In der sich anschließenden zweiten Phase soll ein dezidiertes Schulungskonzept zur Verbesserung der Digital Literacy unter Forschenden und Lehrenden entwickelt und umgesetzt werden, während die Software aktualisiert (Modellierung, Optimierung) bzw. nutzungsorientiert fortentwickelt wird (Personalisierung, Kollaborationsoption, Workflow-Verfeinerung, intelligente Suchalgorithmen).