

Kolloquium, 22. Mai 2017

Dayyan Smith

Optimierung der linguistischen Suche beim XML-annotierten Nachlass von Ludwig Wittgenstein

Vortragende: Faridis Alberteris Azar

Betreuer: Dr. Maximilian Hadersbeck

Faridis schreibt ihre Bachelorarbeit im Rahmen des Digital Humanity Projekts “Wittgenstein in Co-Text”, ein Projekt in Zusammenarbeit mit dem Wittgenstein Archiv der Universität Bergen (WAB) in Norwegen. Das Ziel ist die Optimierung der linguistischen Suche beim XML-annotierten Nachlass von Ludwig Wittgenstein. Dafür soll die XML-Annotation optimal ausgenutzt werden und die XML-Edition verbessert werden. Die XML Dateien liegen in drei Versionen vor: original, normalisiert und diplomatisch. Diese verschiedenen Versionen ermöglichen es Wittgensteins nachträglichen Verbesserungen, nachzuvollziehen und stellen einen normalisierten Text zur Verfügung. Der TreeTagger von Helmut Schmid wird verwendet um die normalisierten Dateien zu taggen. Ursprünglich war in den XML annotierten Texten kein Element für Eigennamen vorhanden. Im März 2017 hat das Wittgenstein Archiv mit Eigennamen annotierte Texte bereitgestellt. In einem ersten Schritt mussten die Eigennamen in der normalisierten Datei lokalisiert werden. Im zweiten Schritt ging es dann darum, die semantische Suche in WittFind zu verbessern, um Eigennamen zu finden und neue Kategorien in WittFind und im CIS Lexikon einzutragen. Getestet wurden die Verbesserungen auf 20 Dateien. Vor den Verbesserungen konnten in dreizehn dieser Dateien insgesamt 168 Personennamen gefunden werden. Mit Verbesserungen konnten in insgesamt 833 Personennamen verteilt über alle Dateien gefunden werden.