

Referat Faridis Alberteris (Montag, den 22.05.2017)

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

Das erste Referat wurde von der Frau Alberteris vorgestellt. Im Rahmen des Digital – Humanities Projekts „Wittgenstein in Co-Text“ in der Zusammenarbeit mit dem Wittgenstein Archivs der Universität Bergen in Norwegen wurde das ganze analysiert und bearbeitet.

Das Ziel dieser Arbeit ist die Optimierung der linguistischen Suche beim XML-annotierten Nachlass von Ludwig Wittgenstein durch die optimale Nutzung der XML-Annotation und die Verbesserung der XML Edition mit den Schwerpunkt auf Personennamen und die Erweiterung des Lexikons.

Es wurde uns erläutert, dass es drei verschiedene XSLT Dateien Editionen gibt. Die erste ist die ORG Datei. Diese beinhaltet alle Optionen. Die zweite ist die NORM Datei, hier aufgelistet, was sein sollte und zuletzt die DIPLO Datei. Diese Datei zeigt wie Ludwig Wittgenstein eigentlich den Text verfasst hat.

Auf der nächsten Folie wurde erklärt, dass ein Tagger und zwar der TreeTagger auf den Texten von Ludwig Wittgenstein genutzt wird. Der Tagger taggt die NORM-Dateien und generiert die NORM-tagged.xml Dateien. Dies wird anbei Folien mit Beispielen erklärt.

Durch verschiedene Schritte versuchen wir hier die Semantische Suche in WittFind zu verbessern bzw Eigennamen zu finden. Hier wird die Form erläutert, welche genutzt wird.

Die Resultate zeigen, dass aus 20 Dateien das WittFind beim jetzigen nutzen in 13 Dateien insgesamt 168 Treffer hat. Jedoch das Empfohlene System findet insgesamt 833 Treffer.

Zur Schlussfolgerung wurde uns gesagt, dass wir hier versuchen die linguistische Suche in WittFind zu verbessern und das Lexikon zu erweitern.

Das Referat wurde mit einigen interessanten Fragen abgeschlossen, wieso Beispielsweise durch ein „Filter“ mehr Treffer gibt, in diesem Fall 6x so viele.