

Protokoll zur Sitzung vom 19.06.2017

Kurzvorstellung von der Bachelorarbeit von Ines Röhrer

BA-Betreuer: Hr. Dr. Maximilian Hadersbeck

Thema: Musik und Ludwig Wittgenstein: Semantische Suche in seinem Nachlass

Ines hat in der heutigen Sitzung ihre Bachelorarbeit vorgestellt, in der es um semantische Suche nach musikalischen Erwähnungen in Wittgensteins Nachlass geht. Der Hauptfokus der Arbeit liegt auf dem Nachlass und WITTFind, eine eigens für den Nachlass konzipierte Suchmaschine, bei der es 2 Suchoptionen existieren: regelbasiertes und semantisches Suchen. WITTFind arbeitet mit dem freizugänglichen Teil von Wittgensteins Nachlass, der wesentlich kleiner ist als der unveröffentlichte Teil. Es gibt 2 Varianten von Texten im Nachlass: Manuskripte und Typoskripte, die in einzelne Bemerkungen unterteilt sind. Jede Bemerkung hat eine individuelle Bezeichnung, bestehend aus „Ms“/„Ts“ und Id-Nummer.

Ziel der Arbeit war, die semantische Suche von der WITTFind Webapplikation (WebFrontend) um eine Suchoption für musikalische Begriffe zu erweitern. Dafür hat Ines die Erweiterungen in der HTML und Javascript Dateien für semantische Suche durchgeführt. Die ganzen Musikbegriffe wurden aus einer Hausarbeit eines anderen Studenten übernommen. Es wurden aber nicht alle Musikbegriffe berücksichtigt. Ines hat ein Beispiel angeführt, in dem der Ausdruck „c“ in einer Bemerkung als ein musikalischer Ton und in einer anderen als eine mathematische Konstante erwähnt wurde. Da der Nachlass keine geordnete Struktur oder Kapitel hat, war es manchmal schwer festzustellen, zu welchem Thema ein Begriff genau gehört. Aus diesem Grund wurden viele potenziell richtige Begriffe weggelassen. Alle Wörter müssen für die Suche in Kategorien eingeteilt werden (z.B.: Komponisten, Instrumente u.a.). Alle Wörter aus einer Kategorie werden in WITTFind als Wordclouds dargestellt, wo das häufigste Wort am größten und das seltenste Wort am kleinsten geschrieben wird. Ines hat ein Beispiel für Wordclouds aus der Kategorie „Komponisten“ angeführt, wo die am seltensten erwähnten Komponisten sehr klein geschrieben wurden, sodass man die Namen nicht lesen konnte (was ihrer Meinung nach zu den Nachteilen von der Darstellungsmethode gehört). Um Wordclouds darzustellen und Frequenzlisten anzubieten, sind Frequenzberechnungen des einzelnen Ausdrucks nötig. Um diese zu bekommen, hat Ines die Wörter in Dictionary gespeichert mit Herkunftsdatei als Value damit das Wort in seiner Kategorie bleibt. Frequenzen werden bei Textdurchlauf hochgezählt. Für eine korrekte Frequenz ist Vollformenextraktion wichtig, für die Ines eine Endungsliste und ein Vollformenlexikon verwendet hat. Obwohl Endungsliste eine „unprofessionelle“ Methode ist, hat sie bessere Ergebnisse geliefert. Ines Meinung nach liegt es daran, dass sie kein gutes Lexikon zur Verfügung hatte. Zu den Problemen bei den Frequenzberechnungen gehören Leerzeichenfehler und multiple Satzvorkommen, weil Wittgenstein sowohl auf Deutsch als auch auf Englisch geschrieben hat und öfters seine Gedanken übersetzt oder wiederholt hat. Im Laufe der Arbeit verschiebt sich der Fokus auf den interessanten Kontext der Musikbegriffe. Für Kontextextraktion wurden 2 Methoden umgesetzt: Ringbuffer und Listenoperation. Die beiden Methoden sind gut, aber Ines hat sich für Listenoperation entschieden, weil Ringbuffer zeitintensiver war. Ines hat beide Varianten von Kontext ausgegeben: mit und ohne Stopwörter. Eine vernünftige Idee war, Kontextextraktion gleichzeitig mit Frequenzberechnung durchzuführen, sodass nur ein Durchlauf benötigt war.

Ines Meinung nach war die Arbeit im Großen und Ganzen erfolgreich, auch wenn einige Ergebnisse nicht perfekt sind. An der Stelle sollte Herr Schulz die Studentin unterbrechen, weil die Präsentation etwa länger schon gedauert hat.