

Zusammenfassung zum Vortrag über die Bachelorarbeit zum Thema „Musik und Wittgenstein: Semantische Suche in seinem Nachlass“ von Ines Röhrer, zusammengefasst von Korbinian Schmidhuber

Zu Beginn des Vortrages wird die am CIS entwickelte Suchmaschine „WiTTFind“ vorgestellt, mit der man den Nachlass von Ludwig Wittgenstein durchsuchen kann. In dieser sind derzeit zwei verschiedene Arten der Suche implementiert: eine Regelbasierte und eine Semantische Suche. In der Semantischen Suche kann man nach Kategorien und Unterkategorien in dem Nachlass suchen. Z.B. gibt es eine Suchkategorie „Farbe“, bei der dann verschiedene Klassen, wie beispielsweise Grundfarben, Zwischenfarbe, und weitere angeklickt werden können. Klickt man eine dieser Klassen an, wird eine Frequenzliste und eine Word Cloud zu Begriffen in dieser Klasse angezeigt. Nach diesem Vorbild sollte eine neue Kategorie zum Thema „Musik“ implementiert werden.

Bei dem Nachlass von Wittgenstein gibt es öffentlich zugängliche Dokumente, aber auch der Öffentlichkeit noch nicht zugängliche Dokumente, welche den Großteil der Dokumente des Nachlasses darstellen. WiTTFind arbeitet mit den öffentlich zugänglichen Dokumenten. Dies hat zur Folge, dass im späteren Verlauf der Arbeit bestimmte Begriffe nicht aufgenommen wurden in die Frequenzlisten, da sie in diesem Teil der Dokumente nur sehr wenig bis gar nicht vorkommen, sondern nur in den unveröffentlichten Dokumenten frequent genug auftreten, als dass man sie als relevant betrachten kann.

Es gibt 2 Varianten von Texten in diesem Nachlass: Manuskripte und Typoskripte. Alle Bemerkungen haben eine Bezeichnung („Ms“ für Manuskript, „Ts“ für Typoskript) und eine eindeutige Identifikationsnummer (z.b. „Ms -104_92“)

Die Motivation für die Arbeit bestand darin, dass die semantische Suche von WiTTFind um eine Kategorie erweitert wird, die für Wittgenstein ein sehr wichtigen Teil seines Lebens war: Musik. Es sollte also ein neues Modul für musikalische Begriffe erstellt werden. Ein zweites Ziel war, Ontologien für diese Begriffe zu untersuchen und zu sehen ob man diese Begriffe als Ontologien modellieren kann und welche Relationen zwischen den Begriffen bestehen.

Um dies zu realisieren musste Ines zunächst einen lokalen Webserver erstellen, um die Änderungen zunächst lokal zu implementieren und zu testen, so dass diese nicht auf dem Server von WiTTFind getestet werden mussten. Anschließend mussten die HTML Dateien angepasst werden, um auf der Webseite von WiTTFind das neue Modul anzuzeigen, und anschließend mussten die Javascript Dateien angepasst werden, in denen die Berechnungen implementiert sind.

Die Basis für die Bestimmung der Musikbegriffe stellte eine Hausarbeit dar. Ein wichtiger Schritt war zunächst mal die Überarbeitung und Auswahl der Begriffe, da nicht alle Begriffe übernommen werden (z.B. aufgrund von fehlender Frequenz in den Texten, oder Ambiguität, die nicht ohne weiteres aufgelöst werden konnte, wie z.b. in dem Begriff „c“, wie in der Note „c“ gegenüber „c“ in mathematischen Formeln).

Herr Schulz nannte an dieser Stelle einen Möglichkeit für ein Vorgehen, um Ontologien aufzustellen, und zwar indem man sich die Kategorien der Named Entities (Personen, Location, usw.) zu Hilfe zieht und anhand dessen Ontologien aufstellt.

Zum Schluss wurden die Kategorien „Komponist“, „Instrument“, „Gattung“, „Intervalle“, „Bezug zu Komposition“ und „Sonstige Begriffe“ verwendet.

Um Word Clouds erstellen zu können sind zuvor Frequenzberechnungen notwendig. Für diese werden die Wörter in einem Dictionary gespeichert mit der Herkunftsdatei als Value. Anschließend

werden für jedes Vorkommen in einem Dokument die Frequenz hochgezählt. Damit diese Frequenzberechnung korrekt ist, ist die Extraktion der Vollformen wichtig. Dies wurde zunächst mit der Verwendung von Endungslisten, anschließend mit der Verwendung eines Vollformenlexikons realisiert.

Probleme, die bei der Berechnung der Frequenzen vorkamen waren einerseits Leerzeichenprobleme, z.B. „awaltz“ anstatt „a waltz“, die durch immer bessere Korpora, ohne Leerzeichenfehler behoben wurden, und andererseits die Problematik von multiplen Satzvorkommen. Es musste geklärt werden, ob es ein Problem darstellt, wenn ein Begriff mehrfach vorkommt, weil der Satz in dem es vorkommt mehrfach vorkommt (z.B. durch Zitieren).

Im Laufe der Arbeit wurde der Fokus der Arbeit immer mehr auf den Kontext der Musikbegriffe verschoben. Es wurden verschiedene Methoden benutzt, um Kontexte zu extrahieren. Im Laufe der Untersuchungen wurde der betrachtete Kontextumfang immer kleiner. Es wurden zwei verschiedene Datenstrukturen für diese Kontexte verwendet und miteinander verglichen: Ringbuffer und Listen.

In Bezug auf eine Ontologie stellte sich heraus, dass eine solche zu erstellen komplexer ist als anfangs erwartet. Ines stoß gegen Ende der Arbeit auf eine Musikontologie, die es bereits gibt und entwickelte mit dieser eine prototypenhafte Modellierung einiger Komponisten.

Als Fazit führte sie auf, dass die Arbeit sich als viel interdisziplinärer herausstellt, als erwartet und dass die Arbeit zu einem erfolgreichem Ergebnis führte, auch wenn die Ergebnisse nicht perfekt seien.