

**Vortrag von Ines Röhrer, Betreuer Max Hadersbeck**

**Thema: Musik und Ludwig Wittgenstein, semantische Suche in seinem Nachlass**

Im ersten Teil des Vortrags werden einige Hintergrundinformationen zum Thema erläutert. Der Hauptfokus der Arbeit liegt auf der eigens für den Nachlass von Wittgenstein konzipierten Suchmaschine WITTFind. Die Maschine bietet zwei Suchoptionen, eine regelbasierte- und eine semantische Suche. Die semantische Suche enthält bereits eine Kategorie für „Farbe“. Der Nachlass von L. Wittgenstein existiert in zwei Teilen, unveröffentlicht und „Open Source“. WITTFind arbeitet mit dem „Open Source“ Teil. Die Textstruktur im Nachlass besteht aus zwei Varianten: Manuskripte und Typoskripte. Die Texte sind wiederum unterteilt in einzelne Bemerkungen und haben eine individuelle Bezeichnung.

Motivation und Ziel der Arbeit ist es die semantische Suche von WITTFind zu erweitern und die Erwähnungen, welche sich auf Musik beziehen, in seinem Nachlass zu untersuchen. Konkret soll die WITTFind Webapplikation um eine Suchoption für Begriffe aus der Musik erweitert werden. Hierfür muss ein neues Modell auf der Webseite hinzugefügt und die Musikbegriffe strukturiert werden. Auch Untersuchungen über Ontologien der Musikbegriffe werden durchgeführt. Zur Vorgehensweise werden zunächst einzelne Komponenten genauer beschrieben. Für die Umsetzung ist die Erweiterung des Webfrontends ein wichtiger Teil. Die HTML Datei des Webfrontends wird erweitert, nach dem Vorbild der Farbsuche Kategorie. Die semantische Suche benötigt eine Erweiterung der Javascript Datei.

Die Musikbegriffe werden in einem Lexikon gespeichert. Die Grundlage für das Lexikon bildet eine Hausarbeit. Daraus werden nach Überarbeitung alle relevanten Begriffe extrahiert. Die Musikbegriffe werden für die semantische Suche in verschiedene Kategorien eingeteilt, z.B. Komponisten und Gattungen.

Um Wordclouds und Frequenzlisten darstellen zu können, sind Frequenzberechnungen der einzelnen Ausdrücke nötig. Die einzelnen Kategorien sollen als Wordclouds angezeigt werden. Wordclouds sind eine visuelle Darstellung von Wörtern, die je nach Wichtigkeit oder Häufigkeit, größer oder kleiner und in einer bestimmten Farbe dargestellt werden. Damit die Wörter korrekt gezählt werden, ist die Vollformenextraktion wichtig. Probleme die bei der Frequenzberechnung auftreten sind Leerzeichenfehler und multiple Satzvorkommen. Neben der Frequenzberechnung wird parallel noch der Kontext eines Begriffs extrahiert. Um die Kontexte zu extrahieren gibt es zwei Methoden: einen Ringbuffer oder Listenoperationen. Der Ringbuffer hat eine hohe Ausführungszeit, da für jeden Kontext ein Ringbuffer erstellt und der Text durchlaufen werden muss. Die Listenoperationen sind eine effizientere Methode und wurden in das Endsystem implementiert.

Aufgrund von Überlänge wurde der Vortrag an dieser Stelle abgebrochen.

**Verfasser: Thomas Ebert**