

Protokoll 1 für die Sitzung vom 19.06.2015

Ivana Daskalovska 11139620

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

Student: Ines Röhler

Betreuer: Maximilian Hadersbeck

Im Hauptfokus der Bachelorarbeit von Frau Röhler steht WiTTFind und der Nachlass Ludwig Wittgensteins. WiTTFind ist eine am CIS konzipierte Suchmaschine speziell für den Nachlass von Wittgenstein. Innerhalb von WiTTFind sind zwei Suchoptionen möglich: Regelbasiertes und Semantisches Suchen. Der größte Teil von Wittgensteins Nachlass wurde leider bisher nicht veröffentlicht. WiTTFind arbeitet daher mit dem kleineren veröffentlichten Teil des Nachlasses. Dieser Teil enthält zwei Arten von Texten: Manuskripte und Typoskripte, die in einzelne Bemerkungen unterteilt sind. Jede Bemerkung hat eine individuelle Bezeichnung, die aus 'Ms' oder 'Ts' und einer einzigartigen Identifikationsnummer besteht. z.B.: 'Ms-104_92'. Frau Röhler arbeitet im Rahmen ihrer Bachelorarbeit an der Erweiterung der semantischen Suche von WiTTFind. Sie untersucht konkret die musikalischen Erwähnungen in seinem Nachlass. Das ist besonders wichtig, weil Musik eine sehr große Rolle in Ludwig Wittgensteins Leben gespielt hat.

Die Hauptziele ihrer Arbeit sind:

1. Erweitern der semantischen Suche der WiTTFind Webapplikation um ein neues Tool für die Suche nach musikalischen Begriffen.
2. Untersuchungen über Ontologien für diese Musikbegriffe. Sie ist daran interessiert, inwiefern man diese Musikbegriffe als Ontologie modellieren kann, welche Relationen zwischen den Ausdrücken bestehen und welche vorhandenen Tools man hierzu nutzen kann.

Vorgehensweise:

Wichtiger Teil der Umsetzung ist die direkte Erweiterung des Webfrontends. Dafür hat Frau Röher einen lokalen Webserver auf dem eigenen Laptop aufgesetzt, damit sie nicht mit der Online Version arbeiten muss, was ihre Arbeit erschweren würde. Dann hat sie ein neues Modul in der HTML Datei nach Vorbild der bereits existierenden Farbensuche implementiert und ebenso die JavaScript Datei für die semantische Suche erweitert.

Als Basis für die ganze Arbeit mit den Musikbegriffen hat sie eine Hausarbeit von einem Musikwissenschaftler verwendet. Das war ein wichtiger Schritt für die Überarbeitung und Auswahl der Begriffe und erste Einteilung in Kategorien, weil sie selber nicht so viel Wissen in diesem Bereich hat und eine richtige Einteilung in Kategorien sehr wichtig ist. Sie hat nur einen Teil von den Begriffen aus der Hausarbeit übernommen, weil nicht alles für die digitale Verarbeitung relevant war.

Die Einteilung der Begriffe in Kategorien wurde dadurch erschwert, dass einige Begriffe im Nachlass in mehreren Kontexten verwendet werden, die nicht unbedingt mit Musik zu tun haben. Z.B. wird „c“ einmal als Musikton verwendet und einmal in einer mathematischen Formel.

Wörter müssen für die Suche in Kategorien eingeteilt werden, wie z.B. Komponisten, Gattungen, Instrumente, Intervalle, Bezug zu Komposition oder sonstige Begriffe.

Um Wordclouds und Frequenzlisten darstellen und anbieten zu können, werden die Vorkommen der einzelnen Ausdrücke berechnet. Damit werden Wordclouds der einzelnen Kategorien angezeigt. Je häufiger ein Wort vorkommt, desto größer wird die Schriftgröße mit der es in der Wordcloud dargestellt wird. Manchmal werden die häufigsten oder seltensten Wörter nicht dargestellt, weil sie zu groß bzw. zu klein sind, um dargestellt werden zu können.

Wörter werden für die Frequenzberechnung in einem Wörterbuch abgespeichert mit Herkunftsdatei als Wert. Die Frequenzen werden bei jedem Textdurchlauf hochgezählt. Die

Vollformenextraktion ist wichtig für eine korrekte Frequenz. Dafür verwendet Frau Röhrer eine selbsterstellte Endungsliste und dann ein Vollformenlexikon. Sie erzielt bessere Ergebnisse mit der eigenen Endungsliste. Probleme, die bei der Frequenzberechnung vorkommen sind Leerzeichenfehler, welche im Laufe der Zeit mit manueller Überprüfung fast alle entfernt wurden und mehrfache Satzvorkommen.

Der Fokus verschiebt sich im Laufe der Arbeit auf die interessanten Kontexte der Musikbegriffe. Die Kontexte werden extrahiert und der Umfang des Kontextes wird berechnet. Es werden zwei Varianten von Kontexten betrachtet: mit und ohne Stoppwörter. Frau Röhrer benutzt dabei zwei Extraktionsmethoden:

1. Ringbuffer: Datenstruktur "Warteschlange" mit festgesetzter Größe, wobei Anfang und Ende verbunden sind. Wenn der Ringbuffer gefüllt ist, überschreiben neue Elemente die Ältesten ('FIFO'). Der Ringbuffer wird als eigene Objektklasse implementiert.
2. Listenoperationen: Mit einer Abfrage wird überprüft, ob ein Wort relevant ist. Bei relevanten Wörtern, extrahiert eine Funktion aus einer Liste von Bemerkungen den Kontext. Anhand einer übergebenen Ausschlussliste wird eine Kontextvariante festgelegt. Der Umfang ist somit dynamisch verschiebbar.

Die Methode mit Listenoperationen wurde als effizientere Methode in der finalen Fassung der Implementation verwendet.

Als nächstes ging Frau Röhrer auf Ontologien ein. Sie stellte fest, dass es viele Relationen zwischen den Musikbegriffen gibt. Sie möchte diese auf der Webseite zugänglich machen. Außerdem wollte sie die Relationen durch eine Ontologie modellieren, welche die Verbindungen darstellen kann. Doch dies stellt sich als komplexer heraus als gedacht. Im Laufe der Arbeit an den Ontologien wurde zufällig eine schon bestehende Musikontologie entdeckt 'The Music Ontology'. Mithilfe dieser Ontologie gelang eine erste, prototypenhafte Modellierung einiger Komponisten. Auf dieser Basis kann sicher die Entwicklung von Musikontologien bei L. Wittgenstein noch ausgebaut werden.

Fazit ihrer Arbeit: Einige Ziele sind im Laufe der Arbeit weggefallen oder dazugekommen. Größtenteils war die Arbeit erfolgreich, auch wenn einige Ergebnisse noch verbesserungswürdig sind. Überraschend war, wie stark interdisziplinär das Thema ihrer Arbeit gewesen ist.