

Protokolle zur Sitzung vom 19.06.2017 - Computerlinguistisches Arbeiten

22.06.2017

Drei Studenten stellten in der Sitzung am 19.06.2017 mit Hilfe einer Präsentation mit dem Beamer ihre Bachelorarbeitsthemen vor.

1. Präsentation:

Referentin: Ines Röhrer

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

Betreuer: Max Hadersbeck

In der Bachelorarbeit von Ines geht es um die Erweiterung der semantischen Suche von der WiTTFind Webapplikation um eine Suchoption für musikalische Begriffe. Hierzu wird ein neues Modul auf der Webseite hinzugefügt und die Musikbegriffe strukturiert.

Zum Beginn erklärte Ines was die WiTTFind Webapplikation ist und wie diese funktioniert.

Bei WiTTFind handelt es sich um eine Art Suchmaschine, mit der man gewisse Textstellen, Begriffe usw. im Nachlass Wittgensteins suchen kann.

Der Nachlass von Wittgenstein existiert in verschiedenen Teilen. WiTTFind arbeitet ausschließlich mit dem wesentlich kleineren Teil, der frei zugänglich und verfügbar ist. Es gibt jedoch auch einen unveröffentlichten Teil, der nicht frei verfügbar ist.

Generell gibt es zwei Varianten von Texten im Nachlass: Manuskripte und Typoskripte. Diese sind unterteilt in Bemerkungen, wobei jede Bemerkung eine individuelle Bezeichnung und eine einzigartige Identifikationsnummer hat, z.B. „MS-104_92“.

Da die Musik eine sehr große Rolle im Leben Wittgensteins gespielt hat, ist es interessant, die musikalischen Erwähnungen in seinem Nachlass zu untersuchen.

Dabei stellen sich Ines drei grundlegende Fragen:

1. Inwiefern kann man Musikbegriffe als Ontologien modellieren?
2. Welche Relationen existieren zwischen den Ausdrücken?
3. Welche vorhandenen Tools kann man nutzen?

Wie ging sie bei der Arbeit vor? Ein wichtiger Teil der Umsetzung ist die Erweiterung des Webfrontends. Hierbei hat sie die HTML-Datei nach Vorbild des Farbenmoduls und die Javascript Datei für die semantische Suche erweitert. Dies veranschaulicht Ines an einer Folie ihrer Präsentation.

Für die Einteilung des Musiklexikons arbeitet Ines mit einem Musikwissenschaftler zusammen, der ihr bei der Einteilung und bei wichtigen Begriffen hilft. Hierbei ist die Überarbeitung und die Auswahl der Begriffe ein wichtiger Schritt. Einige Begriffe werden jedoch auch nicht übernommen, wie z.B. der Ausdruck „C“. „C“ kann im Nachlass als Tonart auftreten, jedoch auch in mathematischen Formeln. Um die Ergebnisse nicht zu verfälschen und unnötige Begriffe auszugeben, lässt Ines solche Arten von Begriffen weg.

Im Anschluss zeigt Ines die Einteilung der semantischen Kategorien. Hierzu zählen bspw.

„Komponisten“, „Gattungen“ und „Instrumente“. Insgesamt hat Ines sechs Kategorien festgelegt. Für diese Beispiele wurden 15, 6 und 15 unterschiedliche Begriffe gefunden. Die Häufigkeit zeigt Ines mithilfe von Wordclouds auf. Um diese darzustellen und Frequenzlisten anbieten zu können, sind Frequenzberechnungen der einzelnen Ausdrücke nötig.

Für die technische Umsetzung werden die Wörter in Dictionarys gespeichert mit der Herkunftsdatei als Value. Die Frequenzen werden beim Textdurchlauf hochgezählt. Um auch Wörter wie „Wagner'sche“ etc. mit aufnehmen zu können, verwendet Ines eine Endungsliste und ein Vollformenlexikon. Bei der Frequenzberechnung treten jedoch auch Probleme auf, wie Leerzeichenfehler und multiple Satzvorkommen im Nachlass. Des weiteren spielt auch der Kontext

eine wichtige Rolle. Dieser kann mit zwei unterschiedlichen Methoden berücksichtigt werden. Zum einen mit Hilfe eines Ringbuffers und zum anderen mit einer Liste, welche mittels Abfrage überprüft ob ein Wort relevant ist. Ist dies der Fall, extrahiert eine Funktion der Liste den Kontext. Da der Ringbuffer eine zu hohe Ausführungsdauer hat, wird die zweite Variante in der finalen Fassung verwendet.

Als Abschluss erklärte Ines noch, dass es viele Relationen zwischen Musikbegriffen gibt und diese auf der Webseite zugänglich gemacht werden sollen. Dies wird am besten durch eine Ontologie modelliert, welche die Verbindungen der Begriffe darstellen kann.