|  |
| --- |
| Christian Kraft  17.3.2019 |

## **Schriftliche Ausarbeitung zur Präsentationsprüfung als 5. PK des Abiturs**



**Inwieweit lassen sich Änderungen von Friedrich Schillers Schriftsprache durch statistische Methoden der Computerlinguistik qualifizieren und lassen sich damit parallelen vom Übergang von "Sturm und Drang" hin zur "Weimarer Klassik" nachweisen?**

Referenzfach: Deutsch

Begleitfach: Informatik

# Darstellung des Arbeitsprozesses

Themenfindung

Für mich war von Anfang an klar, dass ich meine 5. Prüfungskomponente dazu nutzen wollte, mich im Bereich der Informatik mit einem Thema zu beschäftigen, um meine Programmierkenntnisse weiter auszubauen. Diese sind bei mir in der Skriptsprache Python am meisten ausgebaut. Python erfreut sich gerade in den letzten Jahren enormer Beliebtheit, da sie relativ einfach zu erlernen ist und ein sehr breites Anwendungsspektrum bietet.

Gerade in den Bereichen „Data Science“, „machinelles lernen“, „künstliche Intelligenz“ und „IT-Sicherheit“ stößt man immer wieder auf Python.

Da ich mich bereits vor dieser Arbeit viel mit der Erstellung und im Allgemeinen mit Statistiken beschäftigte, wusste ich, dass es im englischsprachigen Raum einige Projekte gab, die sich beispielsweise mit Texten von Autoren beschäftigten, diese so kategorisierten und analysierten, sodass ein Computerprogramm danach in der Lage war mit einem selbst zu chatten, als würde der Autor dies tun.

Eine andere Arbeit verglich die Buchreihen von „Harry Potter“, „Hunger Games“, dem „Herr der Ringe“ und „Game of Thrones“ nach der Wortverwendung, die beliebtesten Wörter und der Schwierigkeit der Texte.

Da ich so etwas für deutsche Texte oder Autoren nicht fand und ich durch meine Kursfachwahl der Q – Phase für mein Referenzfach nur Deutsch und Politikwissenschaft zur Auswahl hatte, war die Entscheidung ziemlich schnell auf Deutsch gefallen.

Um den zeitlichen Rahmen der Arbeit nicht zu sprengen, entschied ich mich für Friedrich Schiller. Dass er mit der „Weimarer Klassik“ und dem „Sturm und Drang“ gleich in zwei Literaturepochen vertreten war, gab mir den Ansatz, ob ich Unterschiede nun auch in den Texten aus den beiden Epochen erkennen kann.

# Methoden und Medieneinsatz

Erstellung der Werkesammlung

Grundlage für diese Art der Textanalyse bilden die Werke als digitaler Text. Vorherige Texterkennung aus gescannten Büchern oder Aufzeichnungen wären auch möglich, der Aufwand und die Fehleranfälligkeit aber um einiges höher.

Von nun an habe ich versucht, so viele Werke wie möglich von Friedrich Schiller zu sammeln. Dabei versuchte ich auch gleich das Erscheinungsjahr und eine Eingruppierung in eine Werkkategorie zu dokumentieren.

Dies wurde beim Auffinden von einzelnen wenigen Werken händisch gemacht und bei gesammelten Werken durch selbstgeschriebene Python Skripte automatisiert aus E-Books und von Internetseiten heruntergeladen.

Textsäuberung und Anpassung („Cleaning“)

Da sich je nach Quelle verschiedenste Informationen wie Seitenzahlen oder Informationen des Verlages im Text befanden, mussten die Texte zu allererst bereinigt werden, sodass am Ende nur noch der eigentliche Text vorhanden war.

Statistikerstellung

Für die Statistiken wurden alle gesäuberten Werke in eine große Tabelle geladen, die Wortformen für jedes einzelne Wort bestimmt und die Anzahl der Substantive, Verben und Adjektive als Summe und als prozentualer Teilwert am Werk abgespeichert. Daraufhin wurden Algorithmen zur Bestimmung der Lesbarkeit, Komplexität und des Schwierigkeitsgrades angewendet und ebenfalls abgespeichert.

Auswertung der Statistiken

Die Auswertung der Statistiken wird über die Erscheinungsjahre hin verglichen um dann in der Präsentation eine Antwort auf das Thema zu finden. Da Friedrich Schiller als Verleger, Dramen

und Gedichtsautor, Universitätsprofessor,... sehr viele unterschiedliche Textarten verfasst hat, wird dies noch einmal hinsichtlich der unterschiedlichen Textarten aufgegliedert.

Medieneinsatz

Aufgrund der großen Datenmenge (664 einbezogene Werke, 13 688 620 Zeilen Text) kam nur ein Computer in Frage. Auf diesem waren sowohl normale Büroprogramme wie ein Textverarbeitungsprogramm, als auch die Programmierumgebung, welche um zahlreiche Programmierbibliotheken, welche der Standardsprache, weitere Funktionen hinzufügt, installiert. Die Ergebnisse werden in einer Powerpoint Präsentation vorgestellt. Für etwaige Fragen, ist die Programmierumgebung jederzeit griffbereit. Zur Klärung von Fragen könnten auch kleine Zeichnungen an der Tafel oder auf einem Flip Chart helfen.

# Gründe für eine Einzelprüfung

Da ich durch Schule, Vollzeitarbeit und einem kleinen Kind meine Zeit nicht wirklich planen kann, wäre ich kein verlässlicher Partner für so eine Präsentationsprüfung geworden. Meine Hauptarbeitszeit war nachts oder für kleine Arbeiten auch die Fahrt zur Arbeit morgens im Zug.

# Reflexion

# Quellenangabe

Literatur

|  |  |  |
| --- | --- | --- |
| **Titel** | **Autor / Verlag** | **Erscheinungsjahr** |
| Textanalytics with Python | Dipanjan Sarkar / Apress | 2016 |
| Fluent Python | Luciano Ramalho / O‘Reilly | 2016 |
| Natural Language Processing with Python | Steven Bird, Ewan Klein & Edward Loper / O‘Reilly | 2009 |
| Texte, Themen und Strukturen | Margret Fingerhut, Bernd Schurf / Cornelsen, Volk und Wissen | 2011 |
| Friedrich Schiller | Claudia Pilling, Diana Schilling, Mirjam Springer / Rowohlt | 2014 |
| Friedrich Schiller – sämtliche Dramen | Friedrich Schiller / e-artnow | 2015 |

Internet - die wichtigsten Quellen für Werke von Schiller

|  |  |
| --- | --- |
| Link | Kommentar |
| www.gutenberg.org | Aktuell für Nutzer aus Deutschland gesperrt, viele Gedichte hatte ich von da |
| https://gutenberg.spiegel.de/autor/friedrich-schiller-518 | Viele Standardwerke – Gedichte wurden automatisiert heruntergeladen |
| http://www.wissen-im-netz.info/literatur/schiller | Die Briefe von Schiller habe ich von hier. Diese wurden bei gleichem Empfänger im jeweiligen Jahr zu einem zusammengefasst |
| www.stackoverflow.com | Zahlreiche „Programmier – Fragen“ und Probleme konnten über das Forum beantwortet werden. |

# Zeitlicher Ablauf der erfolgten Arbeitsschritte

|  |  |  |
| --- | --- | --- |
| **Datum** | **Kommentar** | **Arbeitsschritt** |
| 10.09.2017 | Erste Recherche im Internet | Themenfindung |
| 13.09.2017 | Einrichtung der Programmierumgebung | Vorbereitung |
| 16.09.2017 | Erste deutschsprachige Bücher werden gesammelt, Festlegung auf Schiller noch nicht erfolgt. | Themenfindung |
| 20.09.2017 | Erste Programmiertests | Textsäuberung |
| 11.11.2017 | Aufbau einer Datenbank für die Texte | Erstellung der Werksammlung |
|  |  |  |
|  |  |  |

# Selbstständigkeitserklärung