**Aufgabenpaket 1**

**Geöffnet:** Montag, 25. April 2022, 00:00

**Fällig:** Donnerstag, 28. April 2022, 00:00

**1.1 Rechercheaufgabe: Autoren**

In der Statistik geht es um Beobachtungen, und darum, wie wir in der Wissenschaft mit Beobachtungen umgehen (Dazu mehr in der nächsten Unterrichtseinheit). Für unsere Übungen brauchen wir also Beobachtungen, und um die Sache für alle spannend zu gestalten, wird jede/jeder von Ihnen eigene Beobachtungen anstellen, ergo Daten sammeln, die wir dann nach und nach auswerten.

Zunächst wollen wir untersuchen, wie alt Autorinnen/Autoren werden. Wählen Sie eine Gruppe von Autoren, z.B. aus einer bestimmten Epoche (100 Jahre), einem Genre und einer Sprache, und recherchieren Sie, vorzugsweise in einer Bibliographie, für 20 Autoren, welches Alter diese jeweils erreicht haben (Geburts- und Todesjahr). Dafür sollte die gewählte Epoche zumindest so weit zurück liegen, daß keiner der betreffenden Autoren mehr lebt. Ansonsten wählen Sie Epoche/Genre/Sprache ganz nach Ihren eigenen Interessen oder Vorlieben.

Sammeln Sie die Daten in einer Tabelle mit drei Spalten: "Autor", "Geburtsjahr" und "Todesjahr". Speichern Sie die Daten gleich in einem geeigneten Tabellenformat, am besten .csv, aber auch MS Excel oder Open Office - Dateiformate sind hier noch möglich. Schicken Sie mir ihre Tabelle.

1.2 Übungsaufgabe: Alter

Wie alt werden denn nun die Autorinnen/Autoren in ihrer Tabelle, wenn Sie das jemand fragt? Können Sie darauf eine Antwort (oder mehrere) geben, die sinnvoll/präzise/nützlich/umfassend ist, die weniger als 20 Zahlen enthält? Formulieren Sie eine Antwort in Ihrer Email.

1.3 Rechercheaufgabe: Drama

Suchen Sie ein Theaterstück (gerne digital in XML-TEI) aus. Irgendeines, ganz nach Ihrem Geschmack.

Beispiele für Quellen:

- Deutsch: https://textgridrep.org/

- Englisch (Shakespeare): http://www.folgerdigitaltexts.org

- Französich: https://github.com/cligs/theatreclassique

- Spanisch: https://github.com/GHEDI/BETTE/tree/master/corpus

- Latein/Altgriechisch: http://www.perseus.tufts.edu/

Erzeugen Sie eine Tabelle, in der jede Zeile einer Person entspricht, mit folgenden Spalten:

- id

- label (Name der Person)

- gender

- role (protagonist, lover, antagonist, other)

- importance (primary, secondary, minor)

- per\_mes\_sps (Anzahl von Sprechakten dieser Person; dafür berechnen bzw. zählen wir wie viele sp (oder speaker) Elemente diese Person spricht

Sie können selber definieren, wie Sie die Werte für role (protagonist, lover, antagonist, other) und importance (primary, secondary, minor) den Personen zuteilen.

Speichern Sie die Daten gleich in einem geeigneten Tabellenformat, am besten .csv, aber auch MS Excel oder Open Office - Dateiformate sind hier noch möglich. Schicken Sie mir ihre Tabelle.

**1.4 Technische Vorraussetzungen und Installation: Python in Jupyter Notebooks**

Sie werden in diesem Kurs immer wieder Berechnungen anstellen müssen. Der bequemste Weg, statistische Berechnungen durchzuführen, bei moderneren Verfahren auch teilweise der einzig praktikable, ist die Nutzung einer Programmier- oder Skriptsprache mit entsprechenden mathematischen Funktionen. Die Programmiersprache Python, die Sie ohnehin in Ihrem Studium lernen müssen, bietet Ihnen auch dafür eine ganze Reihe komfortabler Lösungen, und die sollen Sie auch ihm Rahmen dieses Seminars kennen lernen. Keine Sorge, Sie brauchen dafür keine Vorkenntnisse und was wir hier im Kurs machen ist aus programmiertechnischer Sicht alles sehr basal. Ordentliches Programmieren lernen bei Herrn Vitt.

Sie brauchen in diesem Seminar drei Dinge:

1. Eine funktionierende Python-Installation

2. Eine Reihe von Paketen, die Funktionen für die wissenschaftliche Datenverarbeitung bereit stellen

3. Ein Interface, in dem Sie nicht nur programmieren, sondern vor allem auch sinnvoll ihre Lerninhalte protokollieren können.

All das liefert die Python-Distribution Anaconda (https://www.anaconda.com/) in einer Installtion, und darum wird Anaconda die Referenzinstallation für alle Übungen in diesem Kurs sein. Neben all den Paketen, die wir aus inhaltlichen Gründen brauchen, enthält Anaconda das sogenannte "Jupyter Notebook" (https://jupyter.org/), ein browserbasiertes, interaktives Interface, in dem man interaktiv ausführbaren Code mit umfassender Textdokumentation in der sehr einfachen Markupsprache "Markdown" (https://de.wikipedia.org/wiki/Markdown) kombinieren kann. Sie werden sehen, was ich meine, sobald Sie mal so ein Notebook vor sich haben. Sie können in diesen Notebooks Funktionen und Code testen, es als Notizformat für sich selbst verwenden, und ggf. ihre Übungs- und Testaufgaben darin erledigen und sie als PDF exportieren.

Jupyter Notebooks (und der Rest von Anaconda) werden Sie in diesem Studium vermutlich noch bis in die Masterarbeit begleiten. Investieren Sie also die Zeit, auch wenn es am Anfang aufwändig erscheint, Jupyter und den Rest von Anaconda auf ihrem eigenen Computer zum Laufen zu bringen. Es lohnt sich, und das Internet ist reich an Anleitungen und Problemlösungen. Eine sehr gute Anleitung extra für "Digital Humanists" bietet "The Programming Historian":

https://programminghistorian.org/en/lessons/jupyter-notebooks

Die Seite ist generell sehr zu empfehlen, die Tutorials sind sehr umfangreich, auf die Bedürfnisse der Geisteswissenschaften zugeschnitten, und sie unterliegen durch Peer-Review einer gewissen Qualitätskontrolle.

Im Notfall, falls sie sich wirklich an der Installtion die Zähne ausbeißen, bietet Google mit Google Colab (https://colab.research.google.com/notebooks/intro.ipynb) inzwischen eine Jupyter-Oberfläche incl. aller wichtigen Python-Pakete als Webservice an. Kurz: Sie können statt der Installation auf ihrem Computer auch auf den Google-Service zurückgreifen, wenn Sie einen Google-Account haben. Der Funktionsumfang sollte für den Bedarf dieses Seminars auf jeden Fall reichen. Aber: Ich erwähne das ausdrücklich aus Ausweich-/Notlösung, denn:

1. Sie werden im Studium noch so viel mit Anaconda und Jupyter zu tun haben, daß Sie irgendwann eh nicht mehr um die Installation herum kommen.

2. Auch wenn realistisch gesehen die meisten wohl einen Google-Account haben, fände ich es sehr problematisch, das hier im Seminar zur Voraussetzung zu machen.

**Daher ihre Aufgabe:**

Installieren sie Python Anaconda auf ihrem Rechner. Anaconda enthält u.a. auch das Paket "Jupyter". Suchen Sie sich ein Tutorial (z.B. https://www.dataquest.io/blog/jupyter-notebook-tutorial/ oder das aus "The Programming Historian" s.o.) um sich mit den Grundfunktionen von Jupyter vertraut zu machen. Erstellen Sie ein Notebbook, in dem (kurz) erklärt wird, wie man in Python Zahlen addiert, subtrahiert, multipliziert und dividiert. Verwenden Sie Überschriften für die Textformatierung. Speichern Sie dieses Notebook und senden Sie es mit als .pdf oder .ipynb Datei zu.