

Dokumentation der Fingerübung

Name: EL Masiani

Vorname: Mimoun

Matrikelnummer: 6859378

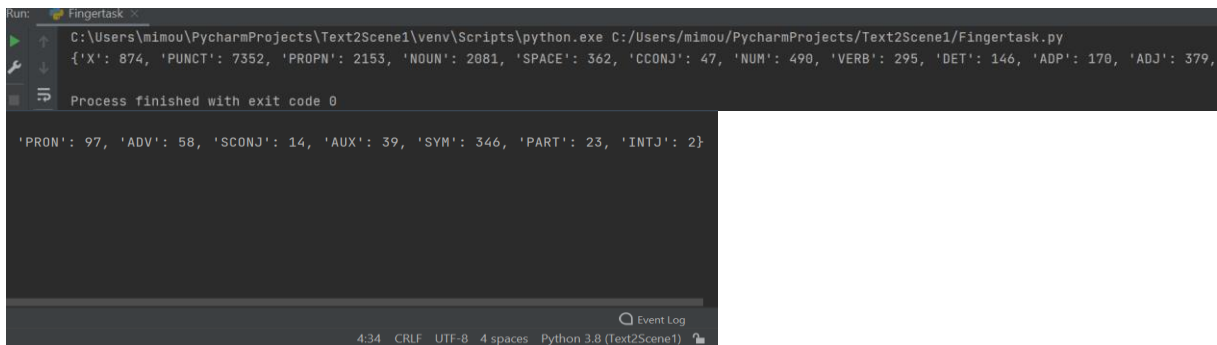
E-Mail: s8847553@stud.uni-frankfurt.de

Abgabetag: 03.11.2021

Im Folgenden werde ich die Dokumentation anhand von Screenshots der Konsole visuell deutlich machen.

Aufgabe 2.3

Wie oft kommen welche PoS-Tags vor?

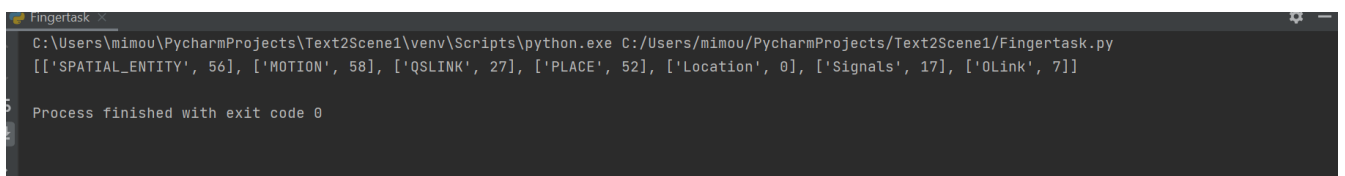


```
Run: Fingertask
C:\Users\mimou\PycharmProjects\Text2Scene1\venv\Scripts\python.exe C:/Users/mimou/PycharmProjects/Text2Scene1/Fingertask.py
{'X': 874, 'PUNCT': 7352, 'PROPN': 2153, 'NOUN': 2081, 'SPACE': 362, 'CCONJ': 47, 'NUM': 490, 'VERB': 295, 'DET': 146, 'ADP': 170, 'ADJ': 379, 'PRON': 97, 'ADV': 58, 'SCONJ': 14, 'AUX': 39, 'SYM': 346, 'PART': 23, 'INTJ': 2}
Process finished with exit code 0
```

Bei diesen zwei Screenshots erkennt man, die Häufigkeit eines POS.
Bsp:

'X' : 874 (POS-Bezeichnung mit dessen Häufigkeit)

Wie viele [SpatialEntities, Places, Motions, Locations, Signals, QsLinks, OLinks] gibt es?



```
Fingertask
C:\Users\mimou\PycharmProjects\Text2Scene1\venv\Scripts\python.exe C:/Users/mimou/PycharmProjects/Text2Scene1/Fingertask.py
[['SPATIAL_ENTITY', 56], ['MOTION', 58], ['QSLINK', 27], ['PLACE', 52], ['Location', 0], ['Signals', 17], ['OLink', 7]]
Process finished with exit code 0
```

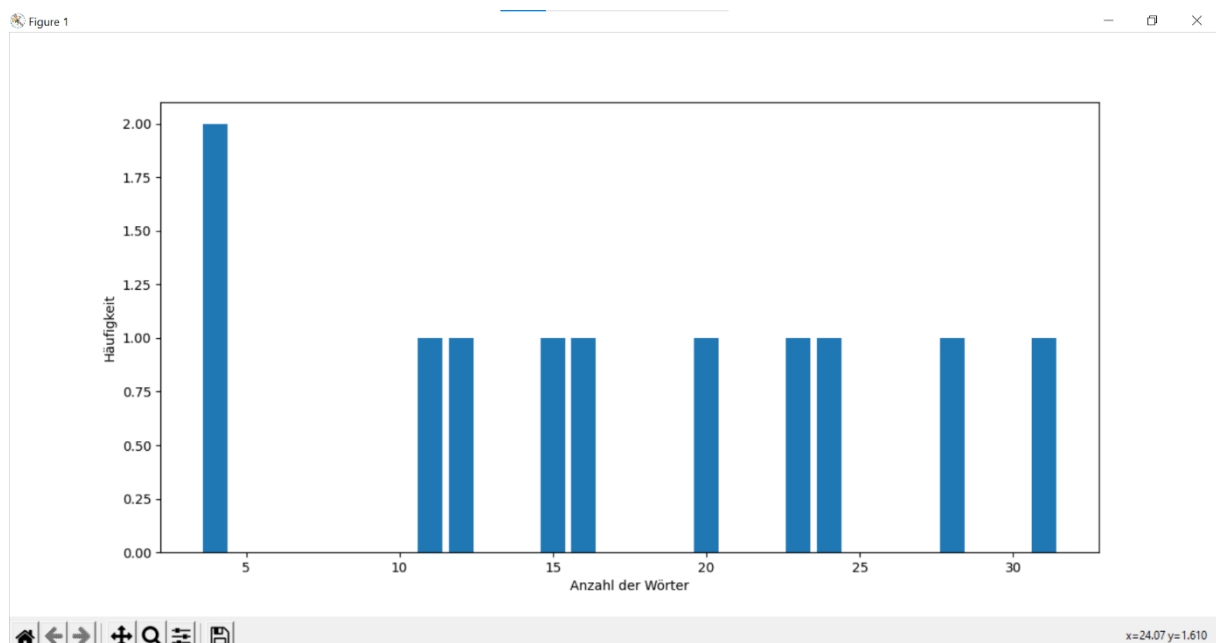
Man erkennt hierbei, dass die große Liste in viele kleine Listen unterteilt ist. In jeder "Unterliste" stehen jeweils die Entities mit ihrer jeweiligen Häufigkeit.

Wie oft kommen welche QsLink Typen vor? (DC,EC, ...)?

```
Run: Fingertask
C:\Users\mimou\PycharmProjects\Text2Scene1\venv\Scripts\python.exe C:/Users/mimou/PycharmProjects/Text2Scene1/Fingertask.py
Counter({'IN': 18, 'EC': 7, 'TPP': 1, 'DC': 1})
Process finished with exit code 0
```

Aus diesem Screenshot geht hervor, dass ich mir der Counter-Methode die Häufigkeiten der QS-Link-Typen abzählen kann. Jedoch musste vorher aus den QS-Link-Zeilen dies gefiltert werden, damit sie finalisierend gezählt werden konnten. (siehe Pythoncode unter der Funktion **qsLinks()**).

Verteilung der Satzlänge graphisch darstellen (x: Satzlänge, y: Wie häufig)?



```
Run: Fingertask
C:\Users\mimou\PycharmProjects\Text2Scene1\venv\Scripts\python.exe C:/Users/mimou/PycharmProjects/Text2Scene1/Fingertask.py
[(4, 2), (31, 1), (28, 1), (23, 1), (11, 1), (16, 1), (15, 1), (24, 1), (12, 1), (20, 1)]
Process finished with exit code 0
```

Bei diesen Screenshots erkennt man, dass die Häufigkeiten der Anzahl der Wörter in einem Satz dargestellt wurden. Auf der Konsole erkennt man, dass in einer Liste von Tupeln an erster Stelle jeweils die Anzahl der Wörter pro Satz bzw. an zweiter Stelle die jeweilige Häufigkeit steht.

Bsp: (4, 2)

→ Anzahl der Wörter pro Satz = 4

→ Häufigkeit von Sätzen bestehend aus vier Wörtern = 2

Diese auf der Konsole dargestellten Ergebnisse wurden dann mithilfe eines Diagrammes visualisiert.

→ X-Achse = Anzahl der Wörter pro Satz

→ Y-Achse = Häufigkeit der der Wörter (x)

Welche Links (QSLinks, OLinks) werden von welchen Präpositionen (markiert durch SPATIAL SIGNAL) getriggert (z.B. wie oft werden QSLinks durch die Präposition „on“ getriggert)?

```
C:\Users\mimou\PycharmProjects\Text2Scene1\venv\Scripts\python.exe C:/Users/mimou/PycharmProjects/Text2Scene1/Fingertask.py
[('s4', 'on'), ('s5', 'between'), ('s0', 'where'), ('s7', 'where'), ('s1', 'in'), ('s8', 'in'), ('s10', 'on'), ('s9', 'in'), ('s11', 'in'),
Process finished with exit code 0
```

```
ngertask.py
's9', 'in'), ('s11', 'in'), ('s12', 'where'), ('s13', 'At'), ('s15', 'in'), ('s16', 'on'), ('s19', 'in'), ('s3', 'in'), ('s2', 'in'), ('s6', 'at')]
```

Hierbei war wichtig zu beachten, dass man die IDs der QS-Links und der O-Links mit den IDs der SPATIAL_SIGNAL vergleicht und bei übereinstimmenden IDs die jeweilige Präposition mit dessen IDs herausschreibt. Hierbei habe ich zwei Screenshots erstellt, da die Anzahl der übereinstimmenden ID ziemlich hoch ist.

Welches sind die fünf häufigsten „MOTION“ Verben (und wie oft kommen diese vor)?

```
Run: Fingertask
C:\Users\mimou\PycharmProjects\Text2Scene1\venv\Scripts\python.exe C:/Users/mimou/PycharmProjects/Text2Scene1/Fingertask.py
[('visited', 5), ('ride', 3), ('biking', 3), ('biked', 2), ('rode', 2)]
Process finished with exit code 0
```

Bei diesem Screenshot erkennt man deutlich, welche Verben im Text am häufigsten genannt werden. In einer Liste stehen jeweils in Tupeln, wobei das erste Element des Tupels das Verb bzw. das zweite Element die jeweilige Anzahl veranschaulicht.

Aufgabe 2.4

Lieber Herr Henlein,

ich habe mich in den letzten knapp 14 Tagen sehr oft damit beschäftigt, diese Aufgabe zu lösen. Jedoch habe ich es ehrlicherweise immer noch nicht verstanden. Ich habe zahlreiche Kommilitoninnen und Kommilitonen gefragt. Diese haben es versucht mir zu erklären, jedoch konnte ich aus den ganzen Erklärungen kein Verständnis für die Aufgabe entwickeln. Die Aufgabenstellung ist klar, aber jedoch konnte ich keine eigene Lösung diesbezüglich programmieren.

Ich bitte hierbei um Verständnis.

Verwendete Quellen:

- <https://spacy.io/usage/linguistic-features>
- <https://www.kite.com/python/answers/how-to-find-the-most-common-elements-of-a-list-in-python>
- https://www.w3schools.com/python/ref_string_splitlines.asp
- <https://www.bigdata-insider.de/was-ist-spacy-a-980360/>
- <https://spacy.io/usage/linguistic-features>