SW03 - Gruppe 1

Einführung

M. Nebroj, S. Hauri, S. Ineichen, R. Schwarzentruber

2019-10-04

Testatübung SW03

Aufgabe 0 - Installations Hinweis

export NLTK_DATA=~/Documents/Studium.Local/KBDS/nltk_data

Packages:

```
import nltk
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('maxent_ne_chunker')
nltk.download('words')
nltk.download('stopwords')
nltk.download('wordnet')
```

Aufgabe 1 - Tokenizieren

Tokenizieren sie folgenden Text:

When Alexander Graham Bell invented the telephone he had three missed calls from Chuck Norris.

Output

['When', 'Alexander', 'Graham', 'Bell', 'invented', 'the', 'telephone', 'he', 'had', 'three', 'missed', 'calls', 'f

Code

```
def exercice_01():
    exercice_01_text = "When Alexander Graham Bell invented the telephone he had three missed calls from Chuck
    token_wort = tokenize.word_tokenize(exercice_01_text)
    print(token_wort)
```

Aufgabe 2 - Stop Words

Entfernen sie die Stop Words aus dem tokenizierten Text.

Screenshot

['When', 'Alexander', 'Graham', 'Bell', 'invented', 'the', 'telephone', 'he', 'had', 'three', 'missed', 'calls', 'f
['When', 'Alexander', 'Graham', 'Bell', 'invented', 'telephone', 'three', 'missed', 'calls', 'Chuck', 'Norris', '.

Code

```python
def exercice_02():
    stop_words = set(stopwords.words('english'))
    filtered_sentence = []

    for w in exercice_01():
        if w not in stop_words:
            filtered_sentence.append(w)
    print(filtered_sentence)
```

Aufgabe 3 - Stemming / Lemmatization

Führen sie auf dem gleichen Text Stemming und Lemmatization aus. (Welche Probelmatik entdecken sie in den Resultaten?)

- Problem A
- Problem B

Screenshot

Stem When Alexander Graham Bell invented the telephone he had three missed calls from Chuck Norris.: when alexander g

Lemmatise When Alexander Graham Bell invented the telephone he had three missed calls from Chuck Norris.: When Alexan

Code

```python
def exercice_03():
    stemmer = PorterStemmer()
    lemmatiser = WordNetLemmatizer()
    exercice_03_text = "When Alexander Graham Bell invented the telephone he had three missed calls from Chuck
    print("Stem %s: %s" % ((exercice_03_text), stemmer.stem((exercice_03_text))))
    print("Lemmatise %s: %s" % ((exercice_03_text), lemmatiser.lemmatize((exercice_03_text))))
```