

# SW03 - Gruppe 1

## Einführung

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

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 - Tokenisieren

Tokenisieren 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', 'from', 'Chuck', 'Norris', '.']
```

#### Code

```
def exercise_01():  
    exercise_01_text = "When Alexander Graham Bell invented the telephone he had three missed  
    ↪ calls from Chuck Norris."  
    token_wort = tokenize.word_tokenize(exercise_01_text)  
    print(token_wort)
```

### Aufgabe 2 - Stop Words

Entfernen sie die Stop Words aus dem tokenisierten Text.

#### Screenshot

```
['When', 'Alexander', 'Graham', 'Bell', 'invented', 'the', 'telephone', 'he', 'had', 'three',  
↪ 'missed', 'calls', 'from', 'Chuck', 'Norris', '.']  
['When', 'Alexander', 'Graham', 'Bell', 'invented', 'telephone', 'three', 'missed', 'calls',  
↪ 'Chuck', 'Norris', '.']
```

## Code

```
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 Problematik 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 graham bell invented the telephone he had three missed calls from  
↪ chuck norris.

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

## Code

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