NLP CHEAT SHEET

NLP is a part of Artificial Intelligence which aims to the manipulation of the human/natural language. It is used for extracting meaningful insights from textual datasets. NLP is mainly used for Text Analysis, Text Mining, Sentiment Analysis, Speech Recognition, Machine Translation, etc.

Natural Language Toolkit (NLTK)

It is a toolkit that is a collection of libraries and programs which is used for Statistical Natural Language Processing. NLTK has been successfully used as a platform for prototyping and building research systems

Importing:

import nltk

Operations Using NLTK:

1. Lemmatization

```
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()
lemmatizer.lemmatize('thoughts')
```

Output:
'thought'

2. Stemming

```
from nltk.stem import PorterStemmer
ps = PorterStemmer()
ps.stem('working')
```

Output:
'work'

3. Tokenization

```
from nltk.tokenize import word_tokenize
word_tokenize("Hello I am Himanshu
                Sharma")
```

Output:
['Hello', 'I', 'am', 'Himanshu', 'Sharma']

Similarly, NLTK has the essential functionalities required for almost all kinds of natural language processing tasks with Python.

Pattern

Pattern is an open-source python library and performs different NLP tasks. It is mostly used for text processing due to various functionalities it provides. Other than text processing Pattern is used for Data Mining i.e we can extract data from various sources such as Twitter, Google, etc. using the data mining functions provided by Pattern.

Importing:

import pattern.en

Operations using Pattern:

1. Sentiment Analysis

```
from pattern.en import sentiment
print(sentiment("The worst is yet to come"))
```

Output:
(-1.0, 1.0)

2. Spellcheck

```
from pattern.en import suggest
print(suggest("amog"))
```

Output:
[('among', 0.99339), ('amoy', 0.00220), ('amos', 0.00220)]

3. Data Mining

Pattern can be used for data mining using different platforms like google, facebook, wkipedia, etc. The mining function is defined under pattern.web

```
from pattern.web import Google, Wikipedia
```

Pattern is a highly valuable learning environment for due to its easy syntax, it serves as a rapid development framework for web developers.

Textblob

Textblob is built on top of NLTK and Pattern also it is very easy to use and can process the text in a few lines of code. Textblob can help you start with the NLP tasks. It performs different operations on textual data such as noun phrase extraction, sentiment analysis, classification, translation, etc.

Importing:

from textblob import TextBlob

Operations using TextBlob:

1. Noun Phrases

```
from textblob import TextBlob
text = 'Hello I am Himanshu Sharma'
blob = TextBlob(text)
blob.noun_phrases
```

Output:
WordList(['hello', 'himanshu sharma'])

2. Parsing

```
blob.parse()
```

Output:
'Hello/UH/O/O I/PRP/B-NP/O am/VBP/B-VP/O Himanshu/NNP/B-NP/O Sharma/NNP/I-NP/O'

3. N-Grams

```
blob.ngrams(4)
```

Output:
[WordList(['Hello', 'I', 'am', 'Himanshu']), WordList(['I', 'am', 'Himanshu', 'Sharma'])]

TextBlob objects can be treated as Python strings that are trained in Natural Language Processing. It focuses on providing access to common text-processing operations

TextHero

Texthero is a library that is used to analyze and process the textual datasets and make them zero to hero. It is a python package that is used to work with textual data efficiently and quickly. It supports text visualization: vector space visualization, place localization on maps

Importing:

import texthero as hero

Operations using TextHero:

1. Visualizing Wordcloud

```
hero.visualization.word_cloud(df)
```


2. Tokenize

```
hero.tokenize(df)
```

```
0       [tiffanylue, know, listenin, bad, habit, earli...
1       [layin, n, bed, headache, ughhhh, waitin, call]
2       [funeral, ceremony, gloomy, friday]
3       [wants, hang, friends, soon]
4       [dannycastillo, want, trade, someone, houston,...
               ...
39995                          [johnlloydtaylor]
39996              [happy, mothers, day, love]
39997   [happy, mother, day, mommies, woman, man, long...
39998   [niariley, wassup, beautiful, follow, peep, ne...
39999   [mopedronin, bullet, train, tokyo, gf, visitin...
Name: clean_content, Length: 40000, dtype: object
```

Texthero is helpful in saving time and efforts due to its ease of use. It is used by developers because it is easy to use and runs blazingly fast.