

## Natural Language Processing



## Natural Language Processing (NLP)

- A branch of artificial intelligence
- How to deal with text data
  - Gather, clean, decipher, analyze

## Natural Language Processing (NLP)

---

```
: import nltk  
nltk.download("all")
```

## Natural Language Processing (NLP)

```
import nltk
```

```
nltk.download("all")
```

```
[nltk_data] Downloading collection 'all'
[nltk_data] |
[nltk_data] | Downloading package abc to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\abc.zip.
[nltk_data] | Downloading package alpino to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\alpino.zip.
[nltk_data] | Downloading package biocreative_ppi to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\biocreative_ppi.zip.
[nltk_data] | Downloading package brown to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\brown.zip.
[nltk_data] | Downloading package brown_tei to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\brown_tei.zip.
[nltk_data] | Downloading package cess_cat to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\cess_cat.zip.
```

## Natural Language Processing (NLP)

```
[nltk_data] | Unzipping misc\perluniprops.zip.
[nltk_data] | Downloading package nonbreaking_prefixes to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\nonbreaking_prefixes.zip.
[nltk_data] | Downloading package vader_lexicon to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package porter_test to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping stemmers\porter_test.zip.
[nltk_data] | Downloading package wmt15_eval to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping models\wmt15_eval.zip.
[nltk_data] | Downloading package mwa_ppdb to
[nltk_data] | C:\Users\bhuang\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping misc\mwa_ppdb.zip.
[nltk_data] |
[nltk_data] Done downloading collection all
```

- <https://www.nltk.org/install.html>

## Natural Language Processing



## Natural Language Processing

```
: from nltk.tokenize import word_tokenize  
from nltk.corpus import stopwords  
from nltk.stem import PorterStemmer  
from nltk.stem import WordNetLemmatizer  
from nltk import FreqDist
```



## Lion King Review on Amazon Video

The animators deserve five stars as the scenery and animals are lovely. I had to keep reminding myself that this was animated and not a live action film. Only the very first opening scene was filmed. Otherwise, the entire movie is the work of animators. Amazing. However, as a story it really falls flat. I was very disappointed. Several songs are changed, and many of the iconic scenes are scaled way back and lose their charm completely. Can You Feel The Love Tonight, during the day??? Late afternoon at best. Weird. I have four kids, they were all disappointed except the two year old who thought it was pretty great. At other points I thought it would build up to a good pun or exciting scene and it would just fall flat instead. Missed opportunities all over. The hyenas weren't funny at all. I hoped that it would pick up with Timon and Pumbaa's arrival, but it didn't. They were definitely cuter than the rest, but Timon was a real disappointment, not very funny at all. Pumbaa was actually really cute, but they got rid of, They call me Mr. Pig! So sad. I love that part. Otherwise, the animals' faces are all pretty blank for the entire movie, so the emotive acting is really unbelievable, and at times feels super overworked. The sad parts are awkward because the voice acting doesn't match the lions' relatively blank faces. Overall we all kinda hated the movie, and definitely won't watch it again. Maybe the animation should have been narrated by David Attenborough. The story would have probably been more interesting than talking realistic animals with blank stares.

## Lion King Review on Amazon Video

```
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.stem import WordNetLemmatizer
from nltk import FreqDist
```

```
review="The animators deserve five stars as the scenery and animals are lovely. I had to keep reminding
```

```
#Tokenize words and sentences from review
```

```
word_token = word_tokenize(review.lower())
```

```
print(word_token)
```

## Lion King Review on Amazon Video

```
# filter out stopwords

mystopwords = set(stopwords.words("english"))

filtered_words = []
for word in word_token:
    if word not in mystopwords and word.isalpha():
        filtered_words.append(word)

print(filtered_words)
```

```
# stemming - eliminating affixes (suffixes, prefixes, infixes, circumfixes) to obtain a word stem

ps = PorterStemmer()

stemmed_words=[]
for word in filtered_words:
    stemmed_words.append(ps.stem(word))

print("stemmed words: ",stemmed_words)
```

## Lion King Review on Amazon Video

```
# Calculating frequency of words
freq = FreqDist(stemmed_words)

for word, frequency in freq.most_common(5):
    print("{}:{}".format(word, frequency))

#plot frequency for top 10
freq.plot(10,title="Top 10 from Review", linewidth=10, color="g")
```

## Natural Language Processing



## NLP Lemmatizer

```
from nltk.stem import WordNetLemmatizer
```

```
lemmatizer = WordNetLemmatizer()
```

```
Word = "skies"
```

```
print(lemmatizer.lemmatize(Word))
```