Stemming ¶

In [1]:
```
import nltk
from nltk.stem import PorterStemmer
```

In [2]:
```
# Create a Porter Stemmer object
porter_stemmer = PorterStemmer()
```

In [7]:
```
words = ["running", "ran", "jumps", "jumped", "happiness", "happy"]

stemmed_words = [porter_stemmer.stem(word) for word in words]

for original, stemmed in zip(words, stemmed_words):
    print(f"{original} -> {stemmed}")

# The above for is same as the one below
# for i in range(len(words)):
#     print(f"{words[i]} -> {stemmed_words[i]}")
```

```
running -> run
ran -> ran
jumps -> jump
jumped -> jump
happiness -> happi
happy -> happi
```

Lemmatization

In [10]:
```
from nltk.stem import WordNetLemmatizer
from nltk.tokenize import word_tokenize
nltk.download('punkt')
nltk.download('wordnet')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\ASUS\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\ASUS\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

Out[10]:
```
True
```

In [11]:
```
lemmatizer = WordNetLemmatizer()
```

In [12]:
```
paragraph = """
Natural Language Processing (NLP) is a subfield of artificial intelligence tha
between computers and humans using natural language. It involves the developme
that enable machines to understand, interpret, and generate human-like text. N
various applications, including sentiment analysis, machine translation,
chatbots, and information extraction
"""

tokens = word_tokenize(paragraph)
lemmatized_tokens = [lemmatizer.lemmatize(token) for token in tokens]

for original, lemmatized in zip(tokens, lemmatized_tokens):
    print(f"{original} -> {lemmatized}")
```

```
Natural -> Natural
Language -> Language
Processing -> Processing
( -> (
NLP -> NLP
) -> )
is -> is
a -> a
subfield -> subfield
of -> of
artificial -> artificial
intelligence -> intelligence
that -> that
focuses -> focus
on -> on
the -> the
interaction -> interaction
between -> between
computers -> computer
and -> and
humans -> human
using -> using
natural -> natural
language -> language
. -> .
It -> It
involves -> involves
the -> the
development -> development
of -> of
algorithms -> algorithm
and -> and
models -> model
that -> that
enable -> enable
machines -> machine
to -> to
understand -> understand
, -> ,
interpret -> interpret
, -> ,
and -> and
generate -> generate
human-like -> human-like
text -> text
. -> .
NLP -> NLP
plays -> play
a -> a
crucial -> crucial
role -> role
in -> in
various -> various
applications -> application
, -> ,
including -> including
sentiment -> sentiment
```

```
analysis -> analysis
, -> ,
machine -> machine
translation -> translation
, -> ,
chatbots -> chatbots
, -> ,
and -> and
information -> information
extraction -> extraction
```

In [ ]: