In [ ]:
```python
# if installations are not done
pip install nltk scikit-learn
import nltk
nltk.download('punkt')
nltk.download('punkt_tab')

nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')
nltk.download('all')

# full code

import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords, wordnet
from nltk.stem import PorterStemmer, WordNetLemmatizer
from nltk import pos_tag
from sklearn.feature_extraction.text import TfidfVectorizer


docs = [
    "The quick brown foxes are jumping over the lazy dog near the riverbank.",
    "Foxes and dogs are often seen together near riverbanks."
]

stemmer, lemmatizer = PorterStemmer(), WordNetLemmatizer()
tokens = word_tokenize(docs[0])
tokens_no_stop = [w for w in tokens if w.lower() not in stopwords.words('english')]
stemmed = [stemmer.stem(w) for w in tokens_no_stop]
lemmatized = [lemmatizer.lemmatize(w) for w in tokens_no_stop]

print("Tokens:", tokens)
print("POS Tags:", pos_tag(tokens))
print("No Stopwords:", tokens_no_stop)
print("Stemmed:", stemmed)
print("Lemmatized:", lemmatized)

tfidf = TfidfVectorizer()
```

```
matrix = tfidf.fit_transform(docs)
print("\nTF-IDF Matrix:\n", matrix.toarray())
print("Features:", tfidf.get_feature_names_out())
```

In [1]:
```
pip install nltk scikit-learn
```

Note: you may need to restart the kernel to use updated packages.Requirement already satisfied: nltk in c:\users\chatura karank
al\appdata\local\programs\python\python39\lib\site-packages (3.9.1)
Requirement already satisfied: scikit-learn in c:\users\chatura karankal\appdata\local\programs\python\python39\lib\site-packag
es (1.6.1)
Requirement already satisfied: click in c:\users\chatura karankal\appdata\local\programs\python\python39\lib\site-packages (fro
m nltk) (8.1.8)
Requirement already satisfied: regex>=2021.8.3 in c:\users\chatura karankal\appdata\local\programs\python\python39\lib\site-pac
kages (from nltk) (2024.11.6)
Requirement already satisfied: joblib in c:\users\chatura karankal\appdata\local\programs\python\python39\lib\site-packages (fr
om nltk) (1.4.2)
Requirement already satisfied: tqdm in c:\users\chatura karankal\appdata\local\programs\python\python39\lib\site-packages (from
nltk) (4.67.1)
Requirement already satisfied: numpy>=1.19.5 in c:\users\chatura karankal\appdata\local\programs\python\python39\lib\site-packa
ges (from scikit-learn) (2.0.2)
Requirement already satisfied: scipy>=1.6.0 in c:\users\chatura karankal\appdata\local\programs\python\python39\lib\site-packag
es (from scikit-learn) (1.13.1)
Requirement already satisfied: threadpoolctl>=3.1.0 in c:\users\chatura karankal\appdata\local\programs\python\python39\lib\sit
e-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: colorama; platform_system == "Windows" in c:\users\chatura karankal\appdata\local\programs\pytho
n\python39\lib\site-packages (from click->nltk) (0.4.6)

WARNING: You are using pip version 20.2.3; however, version 25.1.1 is available.
You should consider upgrading via the 'c:\users\chatura karankal\appdata\local\programs\python\python39\python.exe -m pip insta
ll --upgrade pip' command.

In [6]:
```python
import nltk
nltk.download('punkt')
nltk.download('punkt_tab')      # Internal resource (fixes your error)

nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')
nltk.download('all')
```

```
[nltk_data] Downloading package punkt to C:\Users\Chatura
[nltk_data]     Karankal\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping tokenizers\punkt.zip.
[nltk_data] Downloading package punkt_tab to C:\Users\Chatura
[nltk_data]     Karankal\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt_tab is already up-to-date!
[nltk_data] Downloading package stopwords to C:\Users\Chatura
[nltk_data]     Karankal\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to C:\Users\Chatura
[nltk_data]     Karankal\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     C:\Users\Chatura Karankal\AppData\Roaming\nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
[nltk_data] Downloading collection 'all'
[nltk_data]    |
[nltk_data]    | Downloading package abc to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\abc.zip.
[nltk_data]    | Downloading package alpino to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\alpino.zip.
[nltk_data]    | Downloading package averaged_perceptron_tagger to
[nltk_data]    |     C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Package averaged_perceptron_tagger is already up-
[nltk_data]    |       to-date!
[nltk_data]    | Downloading package averaged_perceptron_tagger_eng to
[nltk_data]    |     C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping
[nltk_data]    |       taggers\averaged_perceptron_tagger_eng.zip.
[nltk_data]    | Downloading package averaged_perceptron_tagger_ru to
[nltk_data]    |     C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping
[nltk_data]    |       taggers\averaged_perceptron_tagger_ru.zip.
[nltk_data]    | Downloading package averaged_perceptron_tagger_rus to
```

```
[nltk_data]    |        C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping
[nltk_data]    |        taggers\averaged_perceptron_tagger_rus.zip.
[nltk_data]    | Downloading package basque_grammars to
[nltk_data]    |        C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping grammars\basque_grammars.zip.
[nltk_data]    | Downloading package bcp47 to C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package biocreative_ppi to
[nltk_data]    |        C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\biocreative_ppi.zip.
[nltk_data]    | Downloading package bllip_wsj_no_aux to
[nltk_data]    |        C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping models\bllip_wsj_no_aux.zip.
[nltk_data]    | Downloading package book_grammars to C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping grammars\book_grammars.zip.
[nltk_data]    | Downloading package brown to C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\brown.zip.
[nltk_data]    | Downloading package brown_tei to C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\brown_tei.zip.
[nltk_data]    | Downloading package cess_cat to C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\cess_cat.zip.
[nltk_data]    | Downloading package cess_esp to C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\cess_esp.zip.
[nltk_data]    | Downloading package chat80 to C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\chat80.zip.
[nltk_data]    | Downloading package city_database to C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\city_database.zip.
[nltk_data]    | Downloading package cmudict to C:\Users\Chatura
[nltk_data]    |        Karankal\AppData\Roaming\nltk_data...
```

```
[nltk_data]    |   Package cmudict is already up-to-date!
[nltk_data]    | Downloading package comparative_sentences to
[nltk_data]    |     C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\comparative_sentences.zip.
[nltk_data]    | Downloading package comtrans to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package conll2000 to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\conll2000.zip.
[nltk_data]    | Downloading package conll2002 to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\conll2002.zip.
[nltk_data]    | Downloading package conll2007 to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package crubadan to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\crubadan.zip.
[nltk_data]    | Downloading package dependency_treebank to
[nltk_data]    |     C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\dependency_treebank.zip.
[nltk_data]    | Downloading package dolch to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\dolch.zip.
[nltk_data]    | Downloading package english_wordnet to
[nltk_data]    |     C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\english_wordnet.zip.
[nltk_data]    | Downloading package europarl_raw to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\europarl_raw.zip.
[nltk_data]    | Downloading package extended_omw to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package floresta to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\floresta.zip.
[nltk_data]    | Downloading package framenet_v15 to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\framenet_v15.zip.
[nltk_data]    | Downloading package framenet_v17 to C:\Users\Chatura
```

```
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\framenet_v17.zip.
[nltk_data]    | Downloading package gazetteers to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Package gazetteers is already up-to-date!
[nltk_data]    | Downloading package genesis to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Package genesis is already up-to-date!
[nltk_data]    | Downloading package gutenberg to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Package gutenberg is already up-to-date!
[nltk_data]    | Downloading package ieer to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\ieer.zip.
[nltk_data]    | Downloading package inaugural to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Package inaugural is already up-to-date!
[nltk_data]    | Downloading package indian to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\indian.zip.
[nltk_data]    | Downloading package jeita to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package kimmo to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\kimmo.zip.
[nltk_data]    | Downloading package knbc to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package large_grammars to
[nltk_data]    |       C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping grammars\large_grammars.zip.
[nltk_data]    | Downloading package lin_thesaurus to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\lin_thesaurus.zip.
[nltk_data]    | Downloading package mac_morpho to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\mac_morpho.zip.
[nltk_data]    | Downloading package machado to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package masc_tagged to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
```

```
[nltk_data]     | Downloading package maxent_ne_chunker to
[nltk_data]     |     C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Package maxent_ne_chunker is already up-to-date!
[nltk_data]     | Downloading package maxent_ne_chunker_tab to
[nltk_data]     |     C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping chunkers\maxent_ne_chunker_tab.zip.
[nltk_data]     | Downloading package maxent_treebank_pos_tagger to
[nltk_data]     |     C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping taggers\maxent_treebank_pos_tagger.zip.
[nltk_data]     | Downloading package maxent_treebank_pos_tagger_tab to
[nltk_data]     |     C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping
[nltk_data]     |       taggers\maxent_treebank_pos_tagger_tab.zip.
[nltk_data]     | Downloading package moses_sample to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping models\moses_sample.zip.
[nltk_data]     | Downloading package movie_reviews to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Package movie_reviews is already up-to-date!
[nltk_data]     | Downloading package mte_teip5 to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\mte_teip5.zip.
[nltk_data]     | Downloading package mwa_ppdb to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping misc\mwa_ppdb.zip.
[nltk_data]     | Downloading package names to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Package names is already up-to-date!
[nltk_data]     | Downloading package nombank.1.0 to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     | Downloading package nonbreaking_prefixes to
[nltk_data]     |     C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\nonbreaking_prefixes.zip.
[nltk_data]     | Downloading package nps_chat to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\nps_chat.zip.
```

```
[nltk_data]     | Downloading package omw to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Package omw is already up-to-date!
[nltk_data]     | Downloading package omw-1.4 to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Package omw-1.4 is already up-to-date!
[nltk_data]     | Downloading package opinion_lexicon to
[nltk_data]     |     C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\opinion_lexicon.zip.
[nltk_data]     | Downloading package panlex_swadesh to
[nltk_data]     |     C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     | Downloading package paradigms to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\paradigms.zip.
[nltk_data]     | Downloading package pe08 to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\pe08.zip.
[nltk_data]     | Downloading package perluniprops to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping misc\perluniprops.zip.
[nltk_data]     | Downloading package pil to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\pil.zip.
[nltk_data]     | Downloading package pl196x to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\pl196x.zip.
[nltk_data]     | Downloading package porter_test to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping stemmers\porter_test.zip.
[nltk_data]     | Downloading package ppattach to C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\ppattach.zip.
[nltk_data]     | Downloading package problem_reports to
[nltk_data]     |     C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]     |   Unzipping corpora\problem_reports.zip.
[nltk_data]     | Downloading package product_reviews_1 to
[nltk_data]     |     C:\Users\Chatura
[nltk_data]     |     Karankal\AppData\Roaming\nltk_data...
```

```
[nltk_data]    |    Unzipping corpora\product_reviews_1.zip.
[nltk_data]    | Downloading package product_reviews_2 to
[nltk_data]    |     C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Unzipping corpora\product_reviews_2.zip.
[nltk_data]    | Downloading package propbank to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package pros_cons to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Unzipping corpora\pros_cons.zip.
[nltk_data]    | Downloading package ptb to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Unzipping corpora\ptb.zip.
[nltk_data]    | Downloading package punkt to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Unzipping tokenizers\punkt.zip.
[nltk_data]    | Downloading package punkt_tab to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Package punkt_tab is already up-to-date!
[nltk_data]    | Downloading package qc to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Unzipping corpora\qc.zip.
[nltk_data]    | Downloading package reuters to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package rslp to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Unzipping stemmers\rslp.zip.
[nltk_data]    | Downloading package rte to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Unzipping corpora\rte.zip.
[nltk_data]    | Downloading package sample_grammars to
[nltk_data]    |     C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Unzipping grammars\sample_grammars.zip.
[nltk_data]    | Downloading package semcor to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    | Downloading package senseval to C:\Users\Chatura
[nltk_data]    |     Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |    Unzipping corpora\senseval.zip.
[nltk_data]    | Downloading package sentence_polarity to
[nltk_data]    |     C:\Users\Chatura
```

```
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\sentence_polarity.zip.
[nltk_data]    |   Downloading package sentiwordnet to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\sentiwordnet.zip.
[nltk_data]    |   Downloading package shakespeare to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Package shakespeare is already up-to-date!
[nltk_data]    |   Downloading package sinica_treebank to
[nltk_data]    |       C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\sinica_treebank.zip.
[nltk_data]    |   Downloading package smultron to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\smultron.zip.
[nltk_data]    |   Downloading package snowball_data to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Package snowball_data is already up-to-date!
[nltk_data]    |   Downloading package spanish_grammars to
[nltk_data]    |       C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping grammars\spanish_grammars.zip.
[nltk_data]    |   Downloading package state_union to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\state_union.zip.
[nltk_data]    |   Downloading package stopwords to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Package stopwords is already up-to-date!
[nltk_data]    |   Downloading package subjectivity to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\subjectivity.zip.
[nltk_data]    |   Downloading package swadesh to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\swadesh.zip.
[nltk_data]    |   Downloading package switchboard to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping corpora\switchboard.zip.
[nltk_data]    |   Downloading package tagsets to C:\Users\Chatura
[nltk_data]    |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]    |     Unzipping help\tagsets.zip.
[nltk_data]    |   Downloading package tagsets_json to C:\Users\Chatura
```

```
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Unzipping help\tagsets_json.zip.
[nltk_data]      | Downloading package timit to C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Unzipping corpora\timit.zip.
[nltk_data]      | Downloading package toolbox to C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Unzipping corpora\toolbox.zip.
[nltk_data]      | Downloading package treebank to C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Package treebank is already up-to-date!
[nltk_data]      | Downloading package twitter_samples to
[nltk_data]      |       C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Package twitter_samples is already up-to-date!
[nltk_data]      | Downloading package udhr to C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Unzipping corpora\udhr.zip.
[nltk_data]      | Downloading package udhr2 to C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Unzipping corpora\udhr2.zip.
[nltk_data]      | Downloading package unicode_samples to
[nltk_data]      |       C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Unzipping corpora\unicode_samples.zip.
[nltk_data]      | Downloading package universal_tagset to
[nltk_data]      |       C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Unzipping taggers\universal_tagset.zip.
[nltk_data]      | Downloading package universal_treebanks_v20 to
[nltk_data]      |       C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      | Downloading package vader_lexicon to C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Package vader_lexicon is already up-to-date!
[nltk_data]      | Downloading package verbnet to C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Unzipping corpora\verbnet.zip.
[nltk_data]      | Downloading package verbnet3 to C:\Users\Chatura
[nltk_data]      |       Karankal\AppData\Roaming\nltk_data...
[nltk_data]      |     Unzipping corpora\verbnet3.zip.
```

Out[6]: True

In [9]:
```python
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords, wordnet
from nltk.stem import PorterStemmer, WordNetLemmatizer
from nltk import pos_tag
```

```python
from sklearn.feature_extraction.text import TfidfVectorizer


docs = [
    "The quick brown foxes are jumping over the lazy dog near the riverbank.",
    "Foxes and dogs are often seen together near riverbanks."
]

stemmer, lemmatizer = PorterStemmer(), WordNetLemmatizer()
tokens = word_tokenize(docs[0])
tokens_no_stop = [w for w in tokens if w.lower() not in stopwords.words('english')]
stemmed = [stemmer.stem(w) for w in tokens_no_stop]
lemmatized = [lemmatizer.lemmatize(w) for w in tokens_no_stop]

print("Tokens:", tokens)
print("POS Tags:", pos_tag(tokens))
print("No Stopwords:", tokens_no_stop)
print("Stemmed:", stemmed)
print("Lemmatized:", lemmatized)

tfidf = TfidfVectorizer()
matrix = tfidf.fit_transform(docs)
print("\nTF-IDF Matrix:\n", matrix.toarray())
print("Features:", tfidf.get_feature_names_out())
```

```
Tokens: ['The', 'quick', 'brown', 'foxes', 'are', 'jumping', 'over', 'the', 'lazy', 'dog', 'near', 'the', 'riverbank', '.']
POS Tags: [('The', 'DT'), ('quick', 'JJ'), ('brown', 'NN'), ('foxes', 'NNS'), ('are', 'VBP'), ('jumping', 'VBG'), ('over', 'I
N'), ('the', 'DT'), ('lazy', 'JJ'), ('dog', 'NN'), ('near', 'IN'), ('the', 'DT'), ('riverbank', 'NN'), ('.', '.')]
No Stopwords: ['quick', 'brown', 'foxes', 'jumping', 'lazy', 'dog', 'near', 'riverbank', '.']
Stemmed: ['quick', 'brown', 'fox', 'jump', 'lazi', 'dog', 'near', 'riverbank', '.']
Lemmatized: ['quick', 'brown', 'fox', 'jumping', 'lazy', 'dog', 'near', 'riverbank', '.']

TF-IDF Matrix:
 [[0.         0.16999204 0.23891789 0.23891789 0.         0.16999204
  0.23891789 0.23891789 0.16999204 0.         0.23891789 0.23891789
  0.23891789 0.         0.         0.71675366 0.        ]
 [0.36469323 0.25948224 0.         0.         0.36469323 0.25948224
  0.         0.         0.25948224 0.36469323 0.         0.
  0.         0.36469323 0.36469323 0.         0.36469323]]
Features: ['and' 'are' 'brown' 'dog' 'dogs' 'foxes' 'jumping' 'lazy' 'near' 'often'
 'over' 'quick' 'riverbank' 'riverbanks' 'seen' 'the' 'together']
```

In [ ]: