School of Computer Science and Artificial Intelligence

Course Code: 21CS121

Course Name: Natural Language Processing

Course Type: Specialization Elective

Date: 0/08/2024

Name: U. Anurag

Hlt .No: 2203A54050

Assignment-02

1.Take your own text or take text as “Hello there! How are you doing today? NLP is fascinating." Implement Tokenization in the text. [CO1]

2. Take your own words or take words = ["running", "ran", "runs", "easily", "fairly"]. Implement Stemming in the text. [CO1]

3. Implement representation of word on any text or take text as “NLP is fun and interesting.", "NLP involves linguistics and computer science." [CO1]

4. Implement Representation of Sentences on following or take any other sentence

"NLP is an interesting field.", "It involves processing natural language." [CO1]

**MY Answer**

1.Take your own text or take text as “Hello there! How are you doing today? NLP is fascinating." Implement Tokenization in the text. [CO1]

Step 1: I have took the text as “Hello there! How are you doing today? NLP is fascinating."

Step 2: Use nltk lib in python and import it

Import nltk

Step 3: download punkt if required

nltk.download('punkt')

step 4: import word tokenizer from nltk.tokenizer

from nltk.tokenize import word\_tokenize

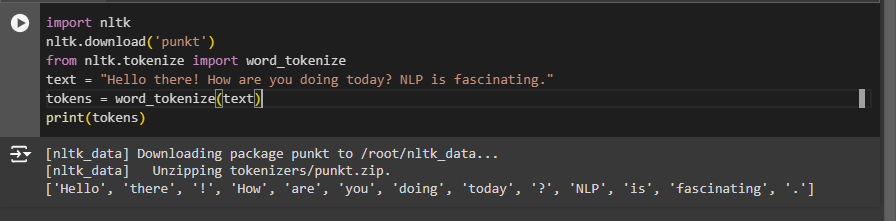
step 5: Take the Text for tokenization “Hello there! How are you doing today? NLP is fascinating."

text = "Hello there! How are you doing today? NLP is fascinating."

Step 6: Tokenizing the text using function word\_tokenize(given text) and print

tokens = word\_tokenize(text)

print(tokens)



2. Take your own words or take words = ["running", "ran", "runs", "easily", "fairly"]. Implement Stemming in the text. [CO1]

Step 1: from library nltk.stem import PorterStemmer function

from nltk.stem import PorterStemmer

step 2: Give Words for stemming

words = ["running", "ran", "runs", "easily", "fairly"]

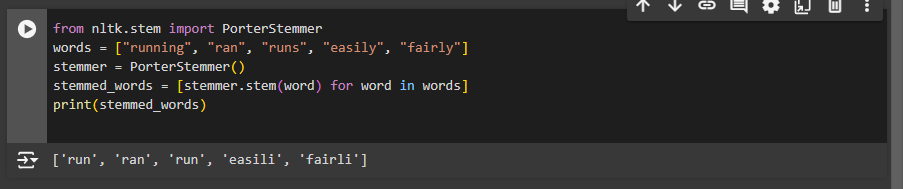
step 3: Create a stemmer by using inbuild function PorterStemmer().

stemmer = PorterStemmer()

step 4: Stem each word using for loop and print the stemmed\_words.

stemmed\_words = [stemmer.stem(word) for word in words]

print(stemmed\_words)



3. Implement representation of word on any text or take text as “NLP is fun and interesting.", "NLP involves linguistics and computer science." [CO1]

Step 1: import CountVectorizer from sklearn .feature library

from sklearn.feature\_extraction.text import CountVectorizer

step 2 : Give Sentences

sentences = ["NLP is fun and interesting.", "NLP involves linguistics and computer science."]

step 3: Initialize the CountVectorizer

vectorizer = CountVectorizer()

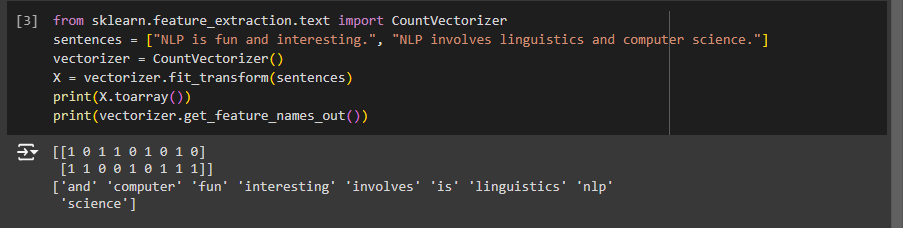
Step 5: Fit the model and transform the sentences into word vectors

X = vectorizer.fit\_transform(sentences)

Step 6 :Convert to an array to visualize and print it

print(X.toarray())

print(vectorizer.get\_feature\_names\_out())



4. Implement Representation of Sentences on following or take any other sentence

"NLP is an interesting field.", "It involves processing natural language."

Step 1: Import CountVectorizer from scikit-learn

from sklearn.feature\_extraction.text import CountVectorizer

step 2: Give Sentences in sentences variable.

sentences = ["NLP is an interesting field.", "It involves processing natural language."]

step 3: Initialize the CountVectorizer

vectorizer = CountVectorizer()

Step 4: Fit the model and transform the sentences into vectors

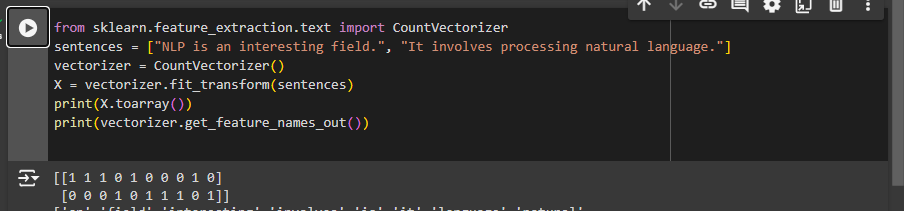
X = vectorizer.fit\_transform(sentences)

Step 5: Convert to an array to visualize the matrix

print(X.toarray())

Step 6: Get the feature names (words) from the vectorizer

print(vectorizer.get\_feature\_names\_out())



Link of the assignment of collab: [click here](https://colab.research.google.com/drive/1HDPjMa5tQ_GDTqOOIahFTrqXD0kbzU0g?usp=sharing)(use collage mail to access the file ).