

**DEPARTMENT OF COMPUTER ENGINEERING**

## CSL804 Computational Lab II

**Eighth Semester, 2021-2022 (Even Semester)**

**Name of Student :** Saurav Kumar

## Roll No. 23

**Division :** BE – CMPN

**Day/ Session :** Monday/Afternoon

**Venue :** SLRTCE Lab 305

## Experiment No. 3

**Title of Experiment :** Generate N-grams from sentences for English and any Indian language.

## Date of Conduction :

**Date of Submission :**

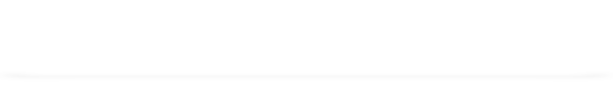
|  |  |  |
| --- | --- | --- |
| **Particulars Max. Marks Marks Obtained** | | |
| Preparedness and Efforts(PE) | **3** |  |
| Knowledge of tools(KT) | **3** |  |
| Debugging and results(DR) | **3** |  |
| Documentation(DN) | **3** |  |
| Punctuality & Lab Ethics(PL) | **3** |  |
| **Total** | **15** |  |

**Grades – Meet Expectations (3 Marks), Moderate Expectations (2 Marks), Below Expectations (1 Mark)**

**Checked and Verified by Name of Faculty :** Prof. Neelam Kulkarni

## Signature :

**Date :**

EXPERIMENT NO: 3

N-GRAMS

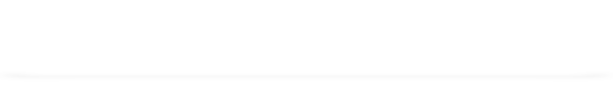
**AIM:** Generate N-grams from sentences for English and any Indian language.

**SOFTWARE:** Python, NLTK, CLTK

# THEORY:

An N-gram language model predicts the probability of a given N-gram within any sequence of words in the language. A good N-gram model can predict the next word in the sentence i.e the value of p(w|h)

Example of N-gram such as unigram (“This”, “article”, “is”, “on”, “NLP”) or bi-gram (‘This article’, ‘article is’, ‘is on’,’on NLP’).



# IMPLEMENTATION:

N-GRAM CODE

import nltk

from nltk.util import ngrams

# Function to generate n-grams from sentences. def extract\_ngrams(data, num):

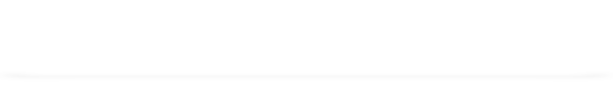
n\_grams = ngrams(nltk.word\_tokenize(data), num) return [ ' '.join(grams) for grams in n\_grams]

data = 'A class is a blueprint for the object.' print("1-gram: ", extract\_ngrams(data, 1)) print("2-gram: ", extract\_ngrams(data, 2)) print("3-gram: ", extract\_ngrams(data, 3)) print("4-gram: ", extract\_ngrams(data, 4))

OUTPUT

Text

Description automatically generated

N-GRAMS USING TEXTBLOB

CODE

# Function to generate n-grams from sentences. from textblob import TextBlob

def extract\_ngrams(data, num):

n\_grams = TextBlob(data).ngrams(num)

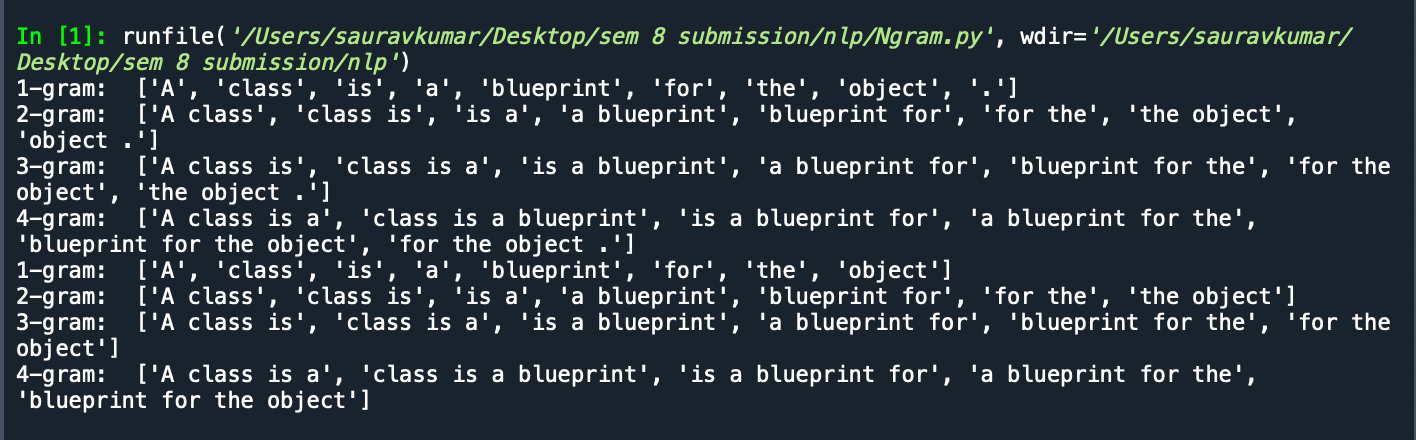
return [ ' '.join(grams) for grams in n\_grams]

data = "N"# hTNी (e˛ +T"

print("N-gram using Textblob :")

print("1-gram: ", extract\_ngrams(data, 1)) print("2-gram: ", extract\_ngrams(data, 2)) print("3-gram: ", extract\_ngrams(data, 3)) print("4-gram: ", extract\_ngrams(data, 4))

OUTPUT



# CONCLUSION:

Thus we have generated N grams from sentences for English and any Indian language.