S. B. JAIN INSTITUTE OF TECHNOLOGY, MANAGEMENT & RESEARCH, NAGPUR.

Practical No. 8

Aim: Declare a list of words and perform stemming on each word using PorterStemmer() and LancasterStemmer(),perform lemmatization on each word of the sentence using WordNetLemmetizer().

Name of Student: Shrutika Pradeep Bagdi

Roll No.: CS22130

Semester/Year: IV/VII

Academic Session: 2025-2026

Date of Performance: _____

Date of Submission: _____

AIM: Declare a list of words and perform stemming on each word using PorterStemmer() and LancasterStemmer(), perform lemmatization on each word of the sentence using WordNetLemmetizer().

OBJECTIVE/EXPECTED LEARNING OUTCOME:

- Understanding Stemming.
- Understanding Lemmatization.

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement:

Software Requirement:

THEORY:

Lemmatization is a text pre-processing technique used in natural language processing (NLP) models to break a word down to its root meaning to identify similarities. For example, a lemmatization algorithm would reduce the word *better* to its root word, or lemme, *good*.

In stemming, a part of the word is just chopped off at the tail end to arrive at the stem of the word. There are different algorithms used to find out how many characters have to be chopped off, but the algorithms don't actually know the meaning of the word in the language it belongs to. In lemmatization, the algorithms do have this knowledge. In fact, you can even say that these algorithms refer to a dictionary to understand the meaning of the word before reducing it to its root word, or lemma.

So, a lemmatization algorithm would know that the word *better* is derived from the word *good*, and hence, the lemme is *good*. But a stemming algorithm wouldn't be able to do the same. There could be over-stemming or under-stemming, and the word *better* could be reduced to either *bet*, or *bett*, or just retained as *better*. But there is no way in stemming that can reduce better to its root word *good*. This is the difference between stemming and lemmatization.

Lemmatization has applications in:

1. Biomedicine: Using lemmatization to parse biomedicine literature may increase the efficiency of data retrieval tasks.
2. Search engines
3. Compact indexing: Lemmatization is an efficient method for storing data in the form of index values.

CODE:

```
import nltk
from nltk.stem import PorterStemmer, LancasterStemmer
from nltk.stem import WordNetLemmatizer
from nltk.corpus import wordnet
nltk.download('punkt')
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')
nltk.download('averaged_perceptron_tagger_eng')
nltk.download('punkt_tab')
porter = PorterStemmer()
lancaster = LancasterStemmer()
lemmatizer = WordNetLemmatizer()
def get_wordnet_pos(tag):
    if tag.startswith('J'):
        return wordnet.ADJ
    elif tag.startswith('V'):
        return wordnet.VERB
    elif tag.startswith('N'):
        return wordnet.NOUN
    elif tag.startswith('R'):
        return wordnet.ADV
    else:
        return None
user_input = input("Enter a sentence or list of words: ")
```

```
words = nltk.word_tokenize(user_input)
pos_tags = nltk.pos_tag(words)
print(f"\n{'Word':<15}{'PorterStemmer':<20}{'LancasterStemmer':<20}{'Lemmatizer (with POS)':<25}")
print("-" * 85)
for word, tag in pos_tags:
    porter_stem = porter.stem(word)
    lancaster_stem = lancaster.stem(word)
    wntag = get_wordnet_pos(tag)
    if wntag is None:
        lemma = lemmatizer.lemmatize(word) # Default to noun if POS not found
    else:
        lemma = lemmatizer.lemmatize(word, pos=wntag)
    print(f"{word:<15}{porter_stem:<20}{lancaster_stem:<20}{lemma:<25}")
```

OUTPUT (SCREENSHOT):

```
Enter a sentence or list of words: The children are playing happily with studies they better understand.

Word           PorterStemmer       LancasterStemmer    Lemmatizer (with POS)
-------------------------------------------------------------------------------------
The            the                 the                 The
children       children            childr              child
are            are                 ar                  be
playing        play                play                play
happily        happili             happy               happily
with           with                with                with
studies        studi               study               study
they           they                they                they
better         better              bet                 good
understand     understand          understand          understand
.              .                   .                   .
```

CONCLUSION:

DISCUSSION AND VIVA VOCE:

- What is Stemming?
- What is Lemmatization?
- Explain porter stemmer.
- Explain Lancaster Stemmer.