

EXPT. NO.	NAME	M T W T F S S	Page No.: 7	YOUVA
		Date: 5/2/25		

EXPERIMENT NO. 4

AIM:

Introduction to open source NLP tools like NLTK, etc.

THEORY:

Natural Language Processing (NLP) is a field of Artificial Intelligence (AI) that enables computers to understand, interpret and generate human language. It combines linguistics, machine learning and computational techniques to process textual and spoken data.

Importance of NLP:

- ① Sentiment Analysis (e.g. analyzing customer feedback)
- ② Chatbots and Virtual Assistants.
- ③ Machine Translation (e.g. Google Translate)
- ④ Speech Recognition (e.g. Siri, Alexa)
- ⑤ Text summarization and Keyword Extraction

Open Source NLP Libraries:

Several open-source NLP tools simplify text processing and analysis. Some of the most widely used ones are:

shreyas kasture

Teacher's Signature: _____

EXPT. NO.	NAME	M T W T F S S
		Page No: <u>3</u> Date: <u>5/2/25</u>
		YOUVA

1. NLTK (Natural Language Toolkit) :

- It is one of the oldest and most widely used NLP libraries for Python. It provides tools for tokenization, stemming, lemmatization, parsing, that is good for educational and research purposes.

Key features :

- ① Tokenization (splitting text into words or sentences)
- ② Stemming & Lemmatization (reducing words to their root form)
- ③ Part-of-Speech (POS) Tagging
- ④ Named Entity Recognition (NER)
- ⑤ Stopword Removal

code :

```
import nltk
from nltk.tokenize import word_tokenize
nltk.download('punkt')
text = "Natural language processing"
tokens = word_tokenize(text)
print(tokens)
```

shreyas kasture

EXPT. NO.	NAME	M T W T F S S						
		Page No. <u>4</u> Date <u>5/2/25</u>						

2. spacy :

- It is a high-performance NLP library optimized for speed and large-scale text processing. It supports deep learning-based NLP models and are more efficient than NLTK for production-level applications.

Key features :

- ① Fast tokenization and Named Entity Recognition (NER)
- ② Dependency Parsing
- ③ POS Tagging
- ④ Pre-trained models for multiple languages

Code :

```
import spacy
nlp = spacy.load("en-core-web-sm")
text = "Apple products are dephering asset."
doc = nlp(text)
for token in doc:
    print(token.text, token.pos_, token.dep_)
```

shreyas kasture

EXPT. NO.	NAME	M T W T F S S	Page No: 10	YOUVA
		Date: 5/2/25		

3. Text Blob :

- It is built on NLTK, but provides a simpler API for NLP tasks. It is used for quick prototyping and sentiment analysis.

Key Features :

- ① Sentiment Analysis.
- ② Text Classification.
- ③ Spelling Correction.

Code:

```
from textblob import TextBlob
text = "Textblob makes NLP easy to use."
blob = TextBlob(text)
print(blob.sentiment)
```

CONCLUSION:

Open-source NLP tools like NLTK, spacy and TextBlob provide essential functionalities for processing and analyzing text. NLTK is suitable for academic purposes, spacy is ideal for production applications and TextBlob simplifies common NLP tasks.

Teacher's Signature: _____