**TEXT AND SPEECH ANALYSIS-CCS369**

**Observation**

**Experiment 1- Create Regular Expressions in python for detecting word patterns, and tokenizing texts:**

1. **Email Address Detection**

import re

pattern = r'[a-zA-Z0-9\_,+-]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-.]+'

text = "Contact us at support@example.com or info@domain.org"

emails = re.findall(pattern,text)

print(emails)

**Output:**

['support@example.com', 'info@domain.org']

1. **URL Detection**

import re

pattern = r'https?://(?:www\.)?[^\s/$.?#].[^\s]\*'

text = "Visit https://www.example.com or https://blog.example.org for info"

urls = re.findall(pattern,text)

print(urls)

**Output:**

['[https://www.example.com](https://www.example.com/)', '[https://blog.example.org](https://blog.example.org/)']

1. **Date Detection: dd-mm-yy, dd/mm/yyyy Format**

import re

text = "Today's date is 04/08/2025 or 04-08-25"

date\_pattern = r'\d{2}[-/]\d{2}[-/]\d{2,4}'

dates = re.findall(date\_pattern, text)

simple\_dates = [f"{d[:2]}/{d[3:5]}/{d[-2:]}" for d in dates]

print(simple\_dates)

**Output:**

['04/08/25', '04/08/25']

1. **Tokenize Sentence into Words**

import re

sentence = "Hello,World! Let's tokenize this sentence."

words = re.findall(r'\b\w+\b',sentence)

print("words:",words)

**Output:**

words: ['Hello', 'World', 'Let', 's', 'tokenize', 'this', 'sentence']

1. **Tokenize Paragraph into Sentences**

import re

paragraph = "Hello world! This is a test. Let's see how it works? Great."

sentences = re.findall(r'[^.!?]+[.!?]',paragraph)

print("sentences:",[s.strip() for s in sentences])

**Output:**

sentences: ['Hello world!', 'This is a test.', "Let's see how it works?", 'Great.']