

LAB#11

Artificial Intelligence
Abdul Haseeb

Lab Objectives

Working with
Regular
Expressions

Implementing
Bag of Words
Model

Regular Expressions

Strings with special syntax, used to match patterns from the given data

Applications:

- Password Criteria
- Matching email Id
- Finding all links in a web document

Regular Expressions

Example:

Regular
expression to find
digits pattern =

`r"\d"`

Bag of Words

One of the simplest and most commonly used techniques in natural language processing (NLP)

In the BoW model, a document is represented as a bag (set) of words, where:

The order of words does not matter.

The frequency of each word in the document is considered.

The model does not account for grammar or word order — it simply considers the presence and frequency of words.

Task#01

Working with Regular Expressions

Write a regular expression to match a string that starts with a capital letter and ends with a period.

Example:

- Input: "Hello world."
- Output: True

Write a regular expression that matches a string with exactly 5 digits (e.g., a zip code).

Task#02

Bag of Words

Look at the given Dialog:

•**Dialogue:**Character A: "The galaxy is in turmoil. The rebels are fighting to overthrow the empire, and we need to find a way to communicate with the leaders. We can't let the forces of darkness take over our world."Character B: "I agree. The resistance is our only hope. But how can we gain support from the people? We need a strategy to win their hearts and minds, to unite them against the empire."Character A: "We must focus on the key battles and disrupt their communication lines. We also need to create a powerful message that inspires hope. If we can sway public opinion, we can win this war."

•**Preprocess the Text:**

- Convert the dialogue to lowercase.
- Remove any punctuation and unnecessary characters.

•**Create a Bag of Words:**

- Convert the cleaned text into a list of words (tokens).

•**Count Word Frequencies**

- Count the frequency of each word in the dialogue.

•**Identify the Topic**