

## Experiment 12: Text Summarization using TF-IDF

### #Objective:

To implement a text summarization technique using TF-IDF to identify the most important sentences in a given text.

### #Code

```
import nltk
from sklearn.feature_extraction.text import TfidfVectorizer
from nltk.tokenize import sent_tokenize
import numpy as np

def summarize_text(text, num_sentences=3):
    sentences = sent_tokenize(text)
    vectorizer = TfidfVectorizer()
    tfidf_matrix = vectorizer.fit_transform(sentences)
    sentence_scores = tfidf_matrix.sum(axis=1).flatten()
    ranked_sentence_indices = np.argsort(sentence_scores)[::-1]
    selected_sentences = [sentences[i] for i in ranked_sentence_indices[:min(num_sentences,
len(sentences))]]
    summary = ''.join(selected_sentences)
    return summary
```

```
text = """
```

Natural language processing (NLP) is a branch of artificial intelligence (AI) that enables computers to comprehend, generate, and manipulate human language. Natural language processing has the ability to interrogate the data with natural language text or voice. This is also called “language in.” Most consumers have probably interacted with NLP without realizing it. For instance, NLP is the core technology behind virtual assistants, such as the Oracle Digital Assistant (ODA), Siri, Cortana, or Alexa. When we ask questions of these virtual assistants, NLP is what enables them to not only understand the user’s request, but to also respond in natural language. NLP applies both to written text and speech, and can be applied to all human languages. Other examples of tools powered by NLP include web search, email spam filtering, automatic translation of text or speech, document summarization, sentiment analysis, and grammar/spell checking. For example, some email programs can automatically suggest an appropriate reply to a message based on its content—these programs use NLP to read, analyze, and respond to your message.

```
"""
```

```
summary = summarize_text(text, 5)
print("Summary:")
print(summary)
```

### #Output:

**Summary:**

**Summary:**

*When we ask questions of these virtual assistants, NLP is what enables them to not only understand the user's request, but to also respond in natural language. Other examples of tools powered by NLP include web search, email spam filtering, automatic translation of text or speech, document summarization, sentiment analysis, and grammar/spell checking. For example, some email programs can automatically suggest an appropriate reply to a message based on its content—these programs use NLP to read, analyze, and respond to your message. For instance, NLP is the core technology behind virtual assistants, such as the Oracle Digital Assistant (ODA), Siri, Cortana, or Alexa. Natural language processing (NLP) is a branch of artificial intelligence (AI) that enables computers to comprehend, generate, and manipulate human language.*

### **#Conclusion:**

This experiment highlights the foundational importance of tokenization in natural language processing (NLP). Through various methods—including NLTK-based sentence and word tokenization, naive splitting, and the use of the TextBlob library—we gained insights into breaking down text into meaningful components. By identifying individual words, phrases, and sentence structures, tokenization facilitates subsequent tasks in the NLP pipeline. Additionally, tools like TextBlob allow for deeper analysis, such as part-of-speech tagging and sentiment analysis, underscoring how tokenization enables detailed text processing and understanding.