Name : Yash Sarang	Class/Roll No. : D16AD / 47	Grade :

Title of Experiment :
Implementation of News summarizer using NLP

Problem Statement :

The problem is to develop a News Summarizer using Natural Language Processing (NLP) techniques. The goal is to create an automated system that can extract the most important information from a news article and generate a concise summary, making it easier for users to grasp the main points of the news quickly.

Description / Theory :

Description/Theory: News summarization is a process that involves condensing the essential information from a news article while retaining its core message and meaning. There are two primary approaches to news summarization: extractive and abstractive summarization. In extractive summarization, important sentences or phrases are selected directly from the article to form the summary. On the other hand, abstractive summarization involves paraphrasing and generating novel sentences to create a summary.

Flowchart :

1. Input a news article.
2. Preprocess the text (remove irrelevant information, tokenize into sentences, etc.).
3. Perform extractive summarization using TextRank or a similar algorithm.
4. Generate a summary by selecting the most important sentences based on scoring.
5. Display the summarized news article.

Program:

```
!pip install nltk newspaper3k
import nltk
from newspaper import Article
from nltk.tokenize import sent_tokenize
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.feature_extraction.text import CountVectorizer
nltk.download('punkt')
```

```
# Function to fetch and summarize news article
def summarize_article(url, num_sentences=3):
    # Download the article
    article = Article(url)
    article.download()
    article.parse()
    article.nlp()

    # Tokenize the article into sentences
    sentences = sent_tokenize(article.text)

    # Create a CountVectorizer to convert sentences to vectors
    vectorizer = CountVectorizer().fit_transform(sentences)
    vectors = vectorizer.toarray()

    # Calculate cosine similarity between sentence vectors
    similarity_matrix = cosine_similarity(vectors)

    # Sort sentences by similarity score
        sentence_scores = [sum(similarity_matrix[i]) for i in
range(len(sentences))]
      ranked_sentences = sorted(((sentence_scores[i], s) for i, s in
enumerate(sentences)), reverse=True)[:num_sentences]

    # Get the summarized text
    summarized_text = " ".join([s[1] for s in ranked_sentences])
    return summarized_text

# Sample news article URL
article_url =
"https://www.theguardian.com/sport/live/2023/oct/05/england-v-new-zealand
-cricket-world-cup-2023-opener-live-updates"

# Summarize the news article
summary = summarize_article(article_url)

# Display the summary
```

```
print("Article Summary:")
print(summary)
```

Output:

Article Summary:
Updated at 09.02 EDT 10h ago 08.56 EDT 4th over: New Zealand 19-1 (Conway 10, Ravindra 8) The reason it's a game of two ends is because Curran has got the length right – aiming for the top of off – and he's moving the ball both ways. Photograph: Andrew Boyers/Reuters Updated at 10.13 EDT 9h ago 09.28 EDT 11th over: New Zealand 92-1 (Conway 44, Ravindra 47) Buttler keeps the faith with Wood, but again the first ball of his over is struck for four – a classical straight drive from Conway. 8h ago 10.35 EDT 28th over: New Zealand 208-1 (Conway 108, Ravindra 97) Moeen returns and, as too often, the first ball of the over goes for four – a pull from Ravindra, who has been quietly motoring along at a run a ball.

Results and Discussions :

The results will include the generated summary for a given news article. The discussion will focus on the accuracy and conciseness of the summary, comparing it to the original article. The evaluation may include metrics such as Rouge scores to measure the quality of the summary.

Conclusion:

News summarization using NLP is a valuable tool for quickly obtaining the main points of a news article. Extractive summarization provides an effective way to generate summaries by selecting significant sentences. However, abstractive summarization, though more challenging, can potentially create more coherent and natural-sounding summaries. Implementing an automated news summarizer enhances information accessibility and can be used in various applications to improve news consumption. Further research and development can lead to more advanced summarization techniques and improved summary quality.

******