Experiment: 11

AIM: Perform text summarization using WordNet

OBJECTIVES: To implement a basic **text summarization** technique by identifying important concepts using **WordNet**.

REQUIREMENTS: Python (version 3.x or above), NLTK library, WordNet.

THEORY:

Text Summarization is a key technique in **Natural Language Processing (NLP)** that uses algorithms to reduce large texts while preserving essential information. Although it doesn't receive as much attention as other machine learning breakthroughs, **text summarization technology** has seen continuous improvements. By extracting key concepts and maintaining the original meaning, these systems can revolutionize industries such as **banking**, **law**, and **healthcare**, enabling faster decision-making and information retrieval.

There are two primary types of text summarization techniques:

- **Extractive**: Selects important sentences directly from the original text.
- **Abstractive:** Generates new sentences that capture the meaning.

This technique uses **WordNet** to find **semantically similar** words (synsets), building **lexical chains** (groups of related words). The sentences with the **most important chains** are selected for the summary.

Implementing Text Summarization using WordNet

Step 1: Importing Required modules.

```
import nltk
nltk.download('punkt')
nltk.download('wordnet')
nltk.download('omw-1.4')
from nltk.corpus import wordnet as wn
from nltk.tokenize import sent_tokenize, word_tokenize
from collections import defaultdict
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\tdhan\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\tdhan\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package omw-1.4 to
[nltk_data]     C:\Users\tdhan\AppData\Roaming\nltk_data...
[nltk_data]   Package omw-1.4 is already up-to-date!
```

Step 2: Taking Paragraph as inputs to summarize

```
Paragraph  = """
Climate change is one of the most pressing issues facing the world today.
It refers to long-term alterations in temperature, precipitation, wind patterns,
and other elements of the Earth's climate system. Human activities,
particularly the burning of fossil fuels, are the primary cause of climate change.
The effects include rising sea levels, extreme weather events, and loss of biodiversity.
Efforts to combat climate change include reducing greenhouse gas emissions,
switching to renewable energy, and increasing energy efficiency.
"""
# Tokenize sentences and words
sentences = sent_tokenize(Paragraph)
tokenized_sentences = [word_tokenize(sent) for sent in sentences]
```

Step 3: Building Lexical chains Using WordNet.

```
# Create a list of all nouns from the text
nouns = []
for sentence in tokenized_sentences:
    for word in sentence:
        if wn.synsets(word, pos=wn.NOUN):
            nouns.append(word.lower())

# Group semantically similar nouns using WordNet synsets
lexical_chains = defaultdict(list)

for noun in nouns:
    synsets = wn.synsets(noun, pos=wn.NOUN)
    if synsets:
        key = synsets[0].lemmas()[0].name()
        lexical_chains[key].append(noun)
```

Step 4: Score Sentences Based on Chain Density

```
# Score each sentence by counting how many chain words it contains
sentence_scores = []

for i, sentence in enumerate(sentences):
    score = 0
    words = word_tokenize(sentence.lower())
    for chain in lexical_chains.values():
        score += len([w for w in words if w in chain])
    sentence_scores.append((i, score))
```

Step 5: Select Top Sentences for Summary with Output

```
# Sort by score and select top N sentences
top_sentences = sorted(sentence_scores, key=lambda x: x[1], reverse=True)[:2]
top_sentences = sorted(top_sentences)  # to keep the order of appearance

# Generate summary
summary = " ".join([sentences[i] for i, _ in top_sentences])
print("Summary:\n", summary)
```

```
Summary:

Climate change is one of the most pressing issues facing the world today.It refers to long-term alterations in temperature,
precipitation, wind patterns, and other elements of the Earth's climate system. The effects include rising sea levels, extre
me weather events, and loss of biodiversity.Efforts to combat climate change include reducing greenhouse gas emissions, swit
ching to renewable energy, and increasing energy efficiency.
```

Conclusion

In this experiment, we implemented a **basic extractive summarizer** using **WordNet lexical chains** to identify important concepts and score sentences. Although simple, this WordNet-based approach shows how **semantic relationships** can enhance understanding and summarization of a text.