

## **1. DOCUMENT SUMMARIZATION**

<b>EX.N0 : 1</b>	<b>DEVELOP A SYSTEM TO DO DOCUMENT SUMMARIZATION</b>
<b><u>DATE : /0 /2025</u></b>	

### **AIM:**

To write a program to develop a system to do document summarization

### **ALGORITHM:**

Step 1: Start

Step 2: Import necessary NLP libraries.

Step 3: Read and preprocess the document.

Step 4: Tokenize the text into sentences.

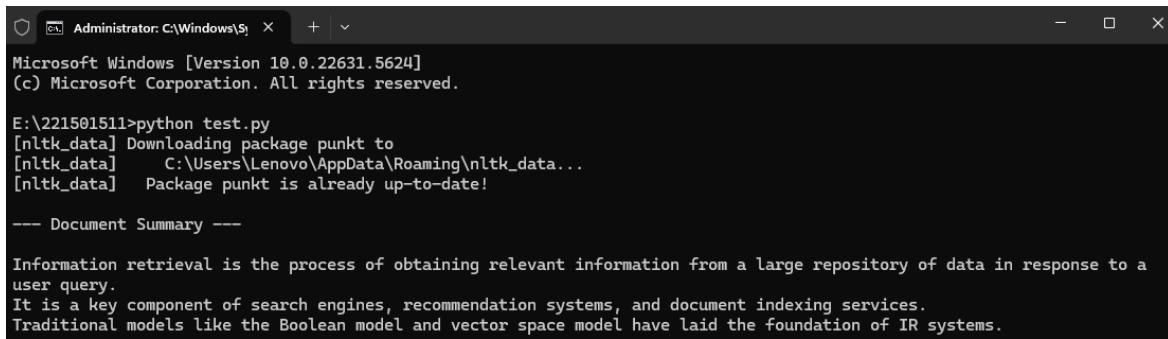
Step 5: Score sentences based on frequency or statistical features.

Step 6: Select top-ranked sentences as the summary.

### **PROGRAM:**

```
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.lex_rank import LexRankSummarizer
text = """
Your input text goes here. You can paste an article or paragraph that you want to summarize.
It should be at least a few sentences long for best results...
parser = PlaintextParser.from_string(text, Tokenizer("english"))
summarizer = LexRankSummarizer()
summary = summarizer(parser.document, 3)
print("Summary:\n")
for sentence in summary:
    print(sentence)
```

## **OUTPUT:**



```
Administrator: C:\Windows\$ > + ▾ Microsoft Windows [Version 10.0.22631.5624]
(c) Microsoft Corporation. All rights reserved.

E:\221501511>python test.py
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\Lenovo\AppData\Roaming\nltk_data...
[nltk_data]     Package punkt is already up-to-date!

--- Document Summary ---

Information retrieval is the process of obtaining relevant information from a large repository of data in response to a user query.
It is a key component of search engines, recommendation systems, and document indexing services.
Traditional models like the Boolean model and vector space model have laid the foundation of IR systems.
```

## **RESULT:**

Thus a program to develop a system to do document summarization has been executed successfully.