



SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)

**SRM INSTITUTE OF SCIENCE AND
TECHNOLOGY**

FACULTY OF ENGINEERING & TECHNOLOGY

(Formerly SRM University, Under section 3 of UGC Act, 1956)

**S.R.M. NAGAR, KATTANKULATHUR –603 203,
KANCHEEPURAM DISTRICT**

**SCHOOL OF COMPUTING
DEPARTMENT OF COMPUTING TECHNOLOGIES**

Course Code: 18CSE305J

Course Name: Artificial Intelligence

Course Project

Title: Text Summarizer

Team Members:

1.Dharva Khambholia - RA1911003010812

2.Pushkar Verma - RA1911003010813

3.Mansi Yadav - RA1911003010817

Date: 19/04/2022

Problem Statement:

To create a Text Summarizer and create your own summarizer in python.

SOURCE CODE:

```
import nltk
from nltk.corpus import stopwords
from nltk.cluster.util import cosine_distance
import numpy as np
import networkx as nx

def read_article(file_name):
    file = open(file_name, "r")
    filedata = file.readlines()
    article = filedata[0].split(". ")
    sentences = []

    for sentence in article:
        print(sentence)
        sentences.append(sentence.replace("[^a-zA-Z]", " ").split(" "))
        sentences.pop()

    return sentences

def sentence_similarity(sent1, sent2, stopwords=None):
    if stopwords is None:
        stopwords = []

    sent1 = [w.lower() for w in sent1]
    sent2 = [w.lower() for w in sent2]

    all_words = list(set(sent1 + sent2))

    vector1 = [0] * len(all_words)
    vector2 = [0] * len(all_words)

    # build the vector for the first sentence
    for w in sent1:
```

```

    if w in stopwords:
        continue
    vector1[all_words.index(w)] += 1

# build the vector for the second sentence
for w in sent2:
    if w in stopwords:
        continue
    vector2[all_words.index(w)] += 1

return 1 - cosine_distance(vector1, vector2)

def build_similarity_matrix(sentences, stop_words):
    # Create an empty similarity matrix
    similarity_matrix = np.zeros((len(sentences), len(sentences)))

    for idx1 in range(len(sentences)):
        for idx2 in range(len(sentences)):
            if idx1 == idx2: #ignore if both are same sentences
                continue
            similarity_matrix[idx1][idx2] =
sentence_similarity(sentences[idx1], sentences[idx2], stop_words)

    return similarity_matrix

def generate_summary(file_name, top_n=5):
    nltk.download("stopwords")
    stop_words = stopwords.words('english')
    summarize_text = []

    # Step 1 - Read text and split it
    sentences = read_article(file_name)

    # Step 2 - Generate Similarity Matrix across sentences
    sentence_similarity_matrix = build_similarity_matrix(sentences,
stop_words)

    # Step 3 - Rank sentences in similarity matrix
    sentence_similarity_graph =
nx.from_numpy_array(sentence_similarity_matrix)

```

```

scores = nx.pagerank(sentence_similarity_graph)

# Step 4 - Sort the rank and pick top sentences
ranked_sentence = sorted(((scores[i],s) for i,s in
enumerate(sentences)), reverse=True)
print("Indexes of top ranked_sentence order are ", ranked_sentence)

for i in range(top_n):
    summarize_text.append(" ".join(ranked_sentence[i][1]))

# Step 5 - Offcourse, output the summarize text
print("Summarize Text: \n", " ".join(summarize_text))

# let's begin
generate_summary( "2.txt", 2)

```

1.txt:

In an attempt to build an AI-ready workforce, Microsoft announced Intelligent Cloud Hub which has been launched to empower the next generation of students with AI-ready skills. Envisioned as a three-year collaborative program, Intelligent Cloud Hub will support around 100 institutions with AI infrastructure, course content and curriculum, developer support, development tools and give students access to cloud and AI services. As part of the program, the Redmond giant which wants to expand its reach and is planning to build a strong developer ecosystem in India with the program will set up the core AI infrastructure and IoT Hub for the selected campuses. The company will provide AI development tools and Azure AI services such as Microsoft Cognitive Services, Bot Services and Azure Machine Learning. According to Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, said, "With AI being the defining technology of our time, it is transforming lives and industry and the jobs of tomorrow will require a different skillset. This will require more collaborations and training and working with AI. That's why it has become more critical than ever for educational institutions to integrate new cloud and AI technologies. The program is an attempt to ramp up the institutional set-up and build capabilities among the educators to educate the workforce of tomorrow." The program aims to build up the cognitive skills and in-depth understanding of developing intelligent cloud connected solutions for applications across industry. Earlier in April this year, the company announced Microsoft Professional Program In AI as a learning track open to the public. The program was developed to provide job ready skills to programmers who wanted to hone their skills in AI and data science with a series of online courses which featured hands-on labs and expert instructors as well. This program also included developer-focused AI school that provided a bunch of assets to help build AI skills.

2.txt:

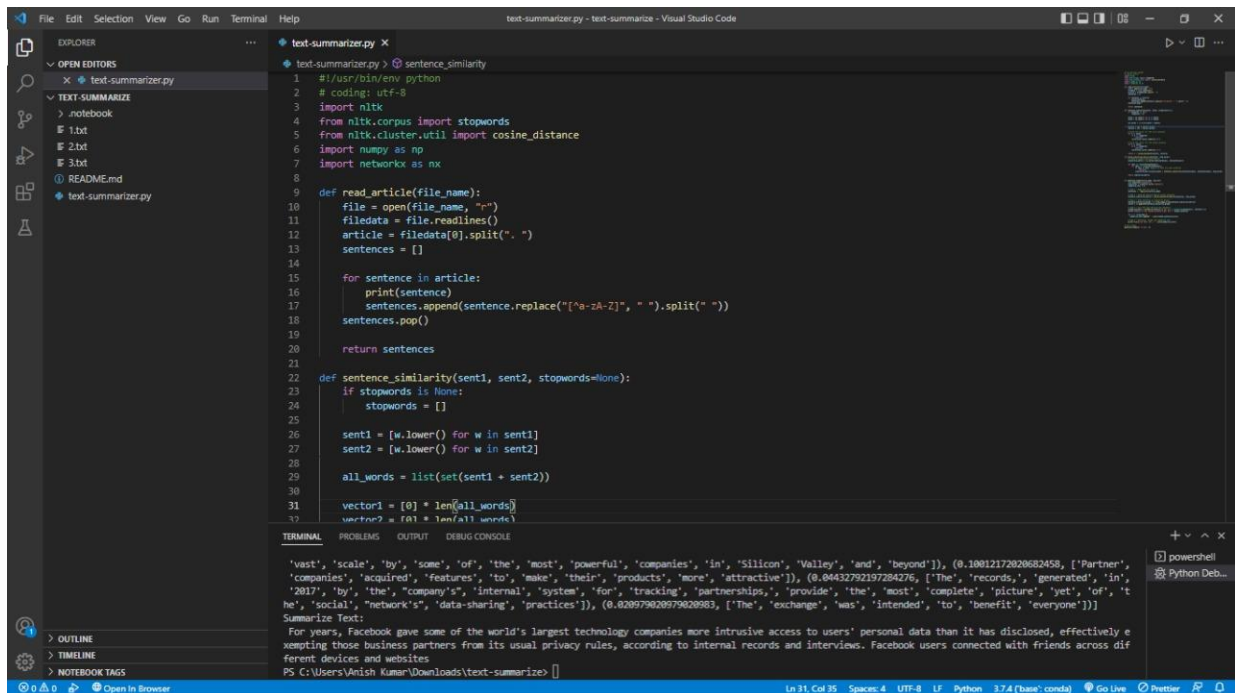
For years, Facebook gave some of the world's largest technology companies more intrusive access to users' personal data than it has disclosed, effectively exempting those business partners from its usual privacy rules, according to internal records and interviews. The special arrangements are detailed in hundreds of pages of Facebook documents obtained by The New York Times. The records, generated in 2017 by the company's internal system for tracking partnerships, provide the most complete picture yet of the social network's data-sharing

practices. They also underscore how personal data has become the most prized commodity of the digital age, traded on a vast scale by some of the most powerful companies in Silicon Valley and beyond. The exchange was intended to benefit everyone. Pushing for explosive growth, Facebook got more users, lifting its advertising revenue. Partner companies acquired features to make their products more attractive. Facebook users connected with friends across different devices and websites. But Facebook also assumed extraordinary power over the personal information of its 2 billion users - control it has wielded with little transparency or outside oversight. Facebook allowed Microsoft's Bing search engine to see the names of virtually all Facebook user's friends without consent, the records show, and gave Netflix and Spotify the ability to read Facebook users' private messages.

3.txt :

WASHINGTON - The Trump administration has ordered the military to start withdrawing roughly 7,000 troops from Afghanistan in the coming months, two defense officials said Thursday, an abrupt shift in the 17-year-old war there and a decision that stunned Afghan officials, who said they had not been briefed on the plans. President Trump made the decision to pull the troops - about half the number the United States has in Afghanistan now - at the same time he decided to pull American forces out of Syria, one official said. The announcement came hours after Jim Mattis, the secretary of defense, said that he would resign from his position at the end of February after disagreeing with the president over his approach to policy in the Middle East. The whirlwind of troop withdrawals and the resignation of Mr. Mattis leave a murky picture for what is next in the United States' longest war, and they come as Afghanistan has been troubled by spasms of violence afflicting the capital, Kabul, and other important areas. The United States has also been conducting talks with representatives of the Taliban, in what officials have described as discussions that could lead to formal talks to end the conflict. Senior Afghan officials and Western diplomats in Kabul woke up to the shock of the news on Friday morning, and many of them braced for chaos ahead. Several Afghan officials, often in the loop on security planning and decision-making, said they had received no indication in recent days that the Americans would pull troops out. The fear that Mr. Trump might take impulsive actions, however, often loomed in the background of discussions with the United States, they said. They saw the abrupt decision as a further sign that voices from the ground were lacking in the debate over the war and that with Mr. Mattis's resignation, Afghanistan had lost one of the last influential voices in Washington who channeled the reality of the conflict into the White House's deliberations. The president long campaigned on bringing troops home, but in 2017, at the request of Mr. Mattis, he begrudgingly pledged an additional 4,000 troops to the Afghan campaign to try to hasten an end to the conflict. Though Pentagon officials have said the influx of forces - coupled with a more aggressive air campaign - was helping the war effort, Afghan forces continued to take nearly unsustainable levels of casualties and lose ground to the Taliban. The renewed American effort in 2017 was the first step in ensuring Afghan forces could become more independent without a set timeline for a withdrawal. But with plans to quickly reduce the number of American troops in the country, it is unclear if the Afghans can hold their own against an increasingly aggressive Taliban. Currently, American airstrikes are at levels not seen since the height of the war, when tens of thousands of American troops were spread throughout the country. That air support, officials say, consists mostly of propping up Afghan troops while they try to hold territory from a resurgent Taliban.

OUTPUT:

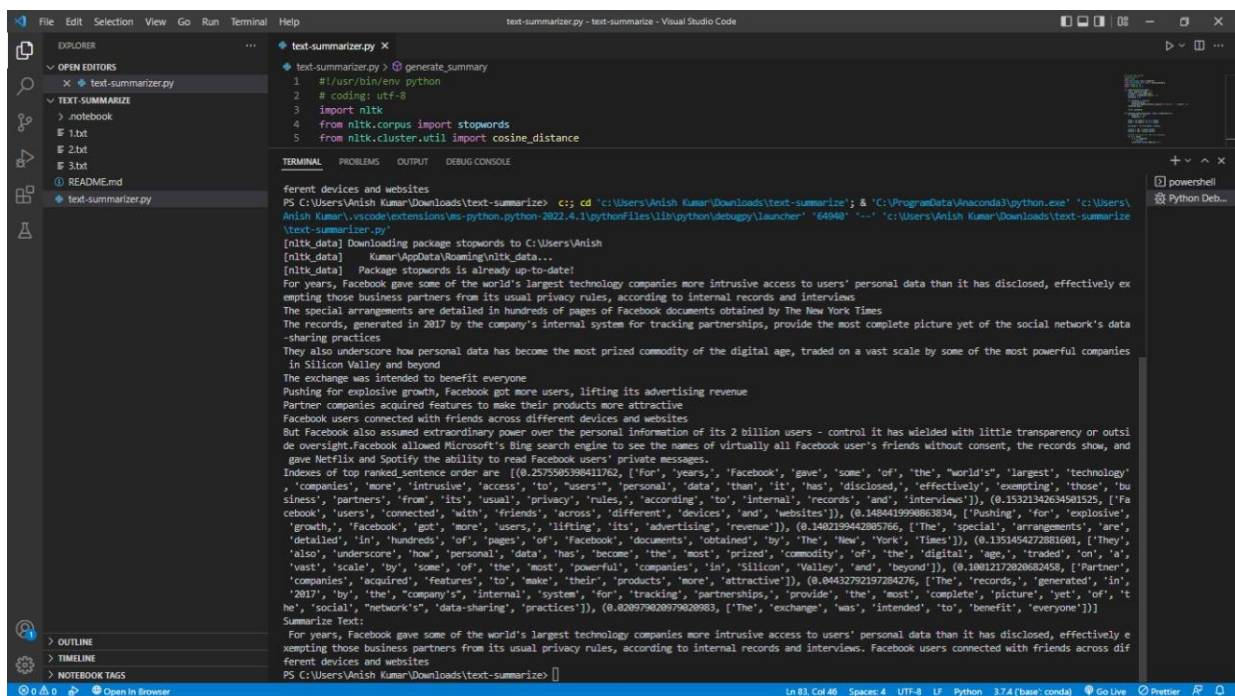


```
text-summarizer.py X
text-summarizer.py > sentence_similarity
1 #!/usr/bin/env python
2 # coding: utf-8
3 import nltk
4 from nltk.corpus import stopwords
5 from nltk.cluster.util import cosine_distance
6 import numpy as np
7 import networkx as nx
8
9 def read_article(file_name):
10     file = open(file_name, "r")
11     filedata = file.readlines()
12     article = filedata[0].split(".")
13     sentences = []
14
15     for sentence in article:
16         print(sentence)
17         sentences.append(sentence.replace("[a-zA-Z]", " ").split(" "))
18     sentences.pop()
19
20     return sentences
21
22 def sentence_similarity(sent1, sent2, stopwords=None):
23     if stopwords is None:
24         stopwords = []
25
26     sent1 = [w.lower() for w in sent1]
27     sent2 = [w.lower() for w in sent2]
28
29     all_words = list(set(sent1 + sent2))
30
31     vector1 = [0] * len(all_words)
32     vector2 = [0] * len(all_words)
```

For years, Facebook gave some of the world's largest technology companies more intrusive access to users' personal data than it has disclosed, effectively exempting those business partners from its usual privacy rules, according to internal records and interviews. Facebook users connected with friends across different devices and websites

Summarize Text:

For years, Facebook gave some of the world's largest technology companies more intrusive access to users' personal data than it has disclosed, effectively exempting those business partners from its usual privacy rules, according to internal records and interviews. Facebook users connected with friends across different devices and websites



```
text-summarizer.py X
text-summarizer.py > generate_summary
1 #!/usr/bin/env python
2 # coding: utf-8
3 import nltk
4 from nltk.corpus import stopwords
5 from nltk.cluster.util import cosine_distance
6
7 def read_article(file_name):
8     file = open(file_name, "r")
9     filedata = file.readlines()
10     article = filedata[0].split(".")
11     sentences = []
12
13     for sentence in article:
14         print(sentence)
15         sentences.append(sentence.replace("[a-zA-Z]", " ").split(" "))
16     sentences.pop()
17
18     return sentences
19
20 def sentence_similarity(sent1, sent2, stopwords=None):
21     if stopwords is None:
22         stopwords = []
23
24     sent1 = [w.lower() for w in sent1]
25     sent2 = [w.lower() for w in sent2]
26
27     all_words = list(set(sent1 + sent2))
28
29     vector1 = [0] * len(all_words)
30     vector2 = [0] * len(all_words)
31
32     for word in sent1:
33         vector1[all_words.index(word)] = 1
34
35     for word in sent2:
36         vector2[all_words.index(word)] = 1
37
38     dot_product = np.dot(vector1, vector2)
39     mag1 = np.linalg.norm(vector1)
40     mag2 = np.linalg.norm(vector2)
41     sim = dot_product / (mag1 * mag2)
42
43     return sim
44
45 def summarize_text(article, num_sentences=10):
46     if num_sentences > len(article):
47         print('Too few sentences: {}'.format(num_sentences))
48         return
49
50     sentences = read_article(article)
51
52     # Sort sentences based on similarity to the first sentence
53     sorted_sentences = sorted(sentences, key=lambda x: sentence_similarity(sentences[0], x))
54
55     # Select the most similar sentences
56     summary_sentences = sorted_sentences[:num_sentences]
57
58     # Join the sentences to form the summary
59     summary = ' '.join(summary_sentences)
60
61     return summary
62
63 if __name__ == '__main__':
64     article = 'ferent devices and websites'
65     summary = summarize_text(article)
66     print(summary)
```

ferent devices and websites

Summarize Text:

For years, Facebook gave some of the world's largest technology companies more intrusive access to users' personal data than it has disclosed, effectively exempting those business partners from its usual privacy rules, according to internal records and interviews. Facebook users connected with friends across different devices and websites

Summarize Text:

For years, Facebook gave some of the world's largest technology companies more intrusive access to users' personal data than it has disclosed, effectively exempting those business partners from its usual privacy rules, according to internal records and interviews. Facebook users connected with friends across different devices and websites

Result: Successfully Executed and Summarized the given three sample text.