

SUMMER TRAINING CSE-2022



**CHANDIGARH
UNIVERSITY**

Discover. Learn. Empower.

Text Summarization

Presented by:

Parth Awasthi
(21bcs1447)

INTRODUCTION

- Automatic text summarization is the task of producing a concise and fluent summary while preserving key information content and overall meaning”-Text Summarization Techniques
- In recent years, the necessity for summarization can be seen in many contexts such as newspaper articles, business documents, search engine results, medical summaries, and online portals to find the most relevant content and track down treatment options. Even as the internet’s information continues to grow, sub-branching has expanded natural language processing limits. Clear information is valuable in search

OBJECTIVE

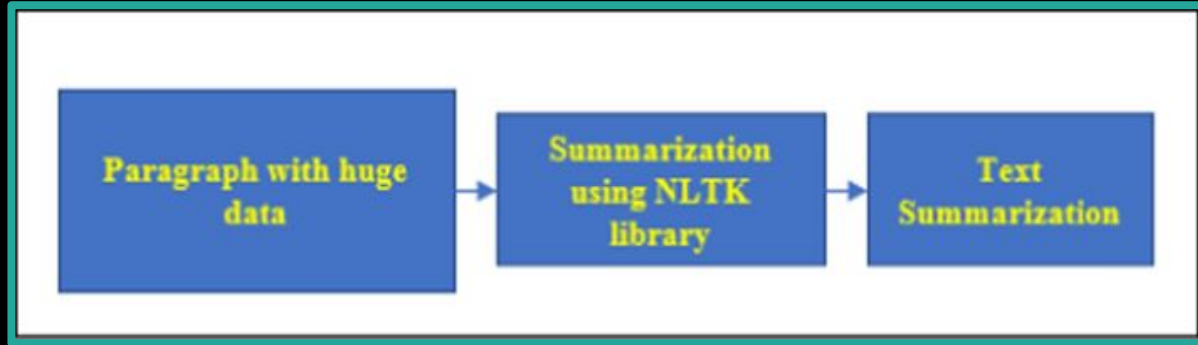
- ▶ This project aims to convert big paragraphs into a summarized text using machine learning.

Importance

- ▶ Can get maximum information by spending minimum time from unstructured textual data.
- ▶ To enhance the readability of the documents.
- ▶ Can eliminate redundant, insignificant text and provide required information
- ▶ Accelerates the process of researching for information



Approach



Steps involved


- 1) Data collection from any website.
- 2) Data clean-up like removing special characters, numeric values, stop words and punctuations.
- 3) Extract Keywords
- 4) Create word cloud
- 5) Summarize


OUTCOME

The outcome of this process is the summary of effective content without changing its meaning. Ultimately users can extract enormous information in the form of an outline. An extended range is essential to business analysts, marketing executives, governments, students, researchers, and teachers. The executive is seen as having to be summarized to allow a maximum amount of information to be processed within a limited time frame

CODE

Downloading required packages, installing them and importing them.

 `pip install rake-nltk`

```
 Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/  
Collecting rake-nltk  
  Downloading rake_nltk-1.0.6-py3-none-any.whl (9.1 kB)  
Requirement already satisfied: nltk<4.0.0,>=3.6.2 in /usr/local/lib/python3.7/dist-packages (from rake-nltk) (3.7)  
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.7/dist-packages (from nltk<4.0.0,>=3.6.2->rake-nltk) (2022.6.2)  
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from nltk<4.0.0,>=3.6.2->rake-nltk) (7.1.2)  
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from nltk<4.0.0,>=3.6.2->rake-nltk) (1.1.0)  
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (from nltk<4.0.0,>=3.6.2->rake-nltk) (4.64.0)  
Installing collected packages: rake-nltk  
Successfully installed rake-nltk-1.0.6
```

```
[ ] import nltk  
    nltk.download("stopwords")  
    nltk.download('punkt')  
    import string  
    from heapq import nlargest
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...  
[nltk_data] Unzipping corpora/stopwords.zip.  
[nltk_data] Downloading package punkt to /root/nltk_data...  
[nltk_data] Unzipping tokenizers/punkt.zip.
```

CODE

Opening and reading the sample text and then printing it in the output shell

```
[ ] with open("sampletext.txt", "r", encoding="utf8") as f:  
    text=f.read()
```

```
[ ] print(text)
```

If the pictures of those towering wildfires in Colorado haven't convinced you, or the size of your AC bill this summer, here are some hard numbers about climate change: June broke or tied 3,215 high-tempe

Meteorologists reported that this spring was the warmest ever recorded for our nation - in fact, it crushed the old record by so much that it represented the "largest temperature departure from average of

RELATED STORIES

The Flaming Lips, Sheryl Crow, the Roots Lead New Climate-Themed Festival This Fall

Crowning Fury: New Mexico Wildfire Reignites Long-Standing Tensions

Not that our leaders seemed to notice. Last month the world's nations, meeting in Rio for the 20th-anniversary reprise of a massive 1992 environmental summit, accomplished nothing. Unlike George H.W. Bush

When we think about global warming at all, the arguments tend to be ideological, theological and economic. But to grasp the seriousness of our predicament, you just need to do a little math. For the past

The First Number: 2° Celsius

If the movie had ended in Hollywood fashion, the Copenhagen climate conference in 2009 would have marked the culmination of the global fight to slow a changing climate. The world's nations had gathered in

In the event, of course, we missed it. Copenhagen failed spectacularly. Neither China nor the United States, which between them are responsible for 40 percent of global carbon emissions, was prepared to o

The accord did contain one important number, however. In Paragraph 1, it formally recognized "the scientific view that the increase in global temperature should be below two degrees Celsius." And in the v

Some context: So far, we've raised the average temperature of the planet just under 0.8 degrees Celsius, and that has caused far more damage than most scientists expected. (A third of summer sea ice in th

Despite such well-founded misgivings, political realism bested scientific data, and the world settled on the two-degree target - indeed, it's fair to say that it's the only thing about climate change the

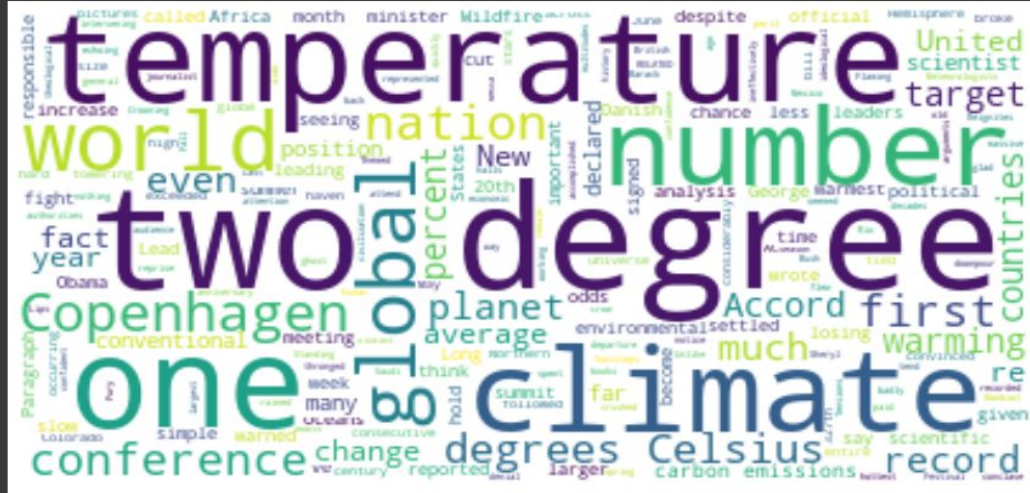
```
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS

#print(STOPWORDS)
print("there are {} words in all text.". format(len(text)))

WC=WordCloud(stopwords=STOPWORDS, background_color="white").generate(text)

plt.figure(figsize=(15,10))
plt.imshow(WC,interpolation='bilinear')
plt.axis("off")
plt.show()
```

there are 7133 words in all text.



CODE

Extract keywords from the paragraph

```
from rake_nltk import Rake  
rk=Rake()
```

```
rk.extract_keywords_from_text(text)  
extract_keyword=rk.get_ranked_phrases()  
extract_keyword
```

```
['confident meeting 20 years ago ','  
'much ." nasa scientist james hansen',  
'gamble ," writes kerry emanuel',  
'accord ratified positions taken earlier',  
'political realism bested scientific data',  
'british journalist george monbiot wrote',  
'danish energy minister connie hedegaard',  
'new mexico wildfire reignites long',  
'saving " copenhagen accord "',  
'crime scene tonight "',  
'purely voluntary agreements committed',  
'intervening decades working ineffectively',  
'airport ." headline writers',  
'since warm air holds',  
'shocking five percent wetter',  
'called major economies forum',  
'" suicide pact "',  
'finally hopeless - position',  
'6 degrees fahrenheit -',  
'massive 1992 environmental summit',  
'arithmetical analysis first published',  
'1995 climate conference chaired',  
'number first gained prominence',  
'angry greenpeace official declared',  
'contain one important number',  
'roots lead new climate',
```

CODE

Process the summary

```
print(text.count("."))
print(string.punctuation)
nopuch=[char for char in text if char not in string.punctuation]
nopuch=" ".join(nopuch)
#print(nopuch)

process_text=[word for word in nopuch.split() if word.lower() not in nltk.corpus.stopwords.words('english')]
#print(process_text)

#create word freq
word_freq={}
for word in process_text:
    if word not in word_freq:
        word_freq[word]=1
    else:
        word_freq[word]=word_freq[word]+1

#dict(sorted(word_freq.items(),key=lambda item:item[1], reverse=True))

max_freq=max(word_freq.values())

for word in word_freq.keys():
    word_freq[word]=(word_freq[word]/max_freq)

#create sent freq
sent_list=nltk.sent_tokenize(text)

sent_score={}
for sent in sent_list:
    for word in nltk.word_tokenize(sent.lower()):
        if word in word_freq.keys():
            if sent not in sent_score.keys():
                sent_score[sent]=word_freq[word]
            else:
                sent_score[sent]=sent_score[sent]+word_freq[word]

#dict(sorted(sent_score.items(),key=lambda item:item[1], reverse=True))

summary_sent=nlargest(3,sent_score, key=sent_score.get)

summary=" ".join(summary_sent)
```


CODE

Output Summary

"Any number much above one degree involves a gamble," writes Kerry Emanuel, chief biodiversity adviser, puts it like this: "If we're seeing what we're seeing." The target that has been talked about in international negotiations for two decades is to limit the rise in global temperature to 1.5 degrees Celsius. But not survive a two-degree rise: "Some countries will flat-out disappear." When the world has settled on. In Paragraph 1, it formally recognized "the scientific consensus that a rise in global temperatures of 1.5 degrees Celsius or more would have significant impacts on the world, and the so-called Major Economies Forum. It was "a ghost of the glad, confident by multitudes." Since I wrote one of the first books for a general audience about climate change, I have been warning that we're losing the fight, badly and quickly - losing it because, most of

THANK YOU