

➤ Newspaper,Article and Essay summarization using deep learning

Installation of the required module.

```
pip install -U spacy
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packages (3.5.2)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.4)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.9)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.7)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.8)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.10/dist-packages (from spacy) (8.1.9)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.1.1)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.4.6)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.8)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.7.0)
Requirement already satisfied: pathy<=0.10.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.10.1)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (6.3.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (4.65.0)
Requirement already satisfied: numpy<=1.15.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.22.4)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.27.1)
Requirement already satisfied: pydantic!=1.8,!<1.8.1,<1.11.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.10.7)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.1.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy) (67.7.2)
Requirement already satisfied: packaging<=20.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (23.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.3.0)
Requirement already satisfied: typing-extensions<=4.2.0 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!<1.8.1,<1.11.0,>=1.7.4) (4.2.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (1.26.15)
Requirement already satisfied: certifi<=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2022.12.7)
Requirement already satisfied: charset-normalizer<=2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2.0.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.4)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8->spacy) (0.7.9)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8->spacy) (0.0.4)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.10/dist-packages (from typer<0.8.0,>=0.3.0->spacy) (8.1.3)
Requirement already satisfied: MarkupSafe<=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2->spacy) (2.1.2)
```

```
pip install sentence-splitter
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: sentence-splitter in /usr/local/lib/python3.10/dist-packages (1.4)
Requirement already satisfied: regex<=2017.12.12 in /usr/local/lib/python3.10/dist-packages (from sentence-splitter) (2022.10.31)
```

```
pip install transformers
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.28.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers) (3.12.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.11.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.14.1)
Requirement already satisfied: numpy<=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (1.22.4)
Requirement already satisfied: packaging<=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (23.1)
Requirement already satisfied: pyyaml<=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (6.0)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2022.10.31)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers) (2.27.1)
Requirement already satisfied: tokenizers!=0.11.3,<0.14,>=0.11.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.13.3)
Requirement already satisfied: tqdm<=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.65.0)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1.0,>=0.11.0->transformers) (2022.11.0)
Requirement already satisfied: typing-extensions<=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1.0,>=0.11.0->transformers) (4.2.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->transformers) (1.26.15)
Requirement already satisfied: certifi<=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->transformers) (2022.12.7)
Requirement already satisfied: charset-normalizer<=2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests->transformers) (2.0.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->transformers) (3.4)
```

```
pip install SentencePiece
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: SentencePiece in /usr/local/lib/python3.10/dist-packages (0.1.98)
```

```
import spacy
from spacy.lang.en.stop_words import STOP_WORDS
from spacy.lang.en.examples import sentences
from string import punctuation
from collections import Counter
from heapq import nlargest
```

```
nlp=spacy.load('en_core_web_sm')
```

```
pip install newspaper3k
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: newspaper3k in /usr/local/lib/python3.10/dist-packages (0.2.8)
Requirement already satisfied: beautifulsoup4>=4.4.1 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (4.11.2)
Requirement already satisfied: Pillow>=3.3.0 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (8.4.0)
Requirement already satisfied: PyYAML>=3.11 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (6.0)
Requirement already satisfied: cssselect>=0.9.2 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (1.2.0)
Requirement already satisfied: lxml>=3.6.0 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (4.9.2)
Requirement already satisfied: nltk>=3.2.1 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (3.8.1)
Requirement already satisfied: requests>=2.10.0 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (2.27.1)
Requirement already satisfied: feedparser>=5.2.1 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (6.0.10)
Requirement already satisfied: tldextract>=2.0.1 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (3.4.1)
Requirement already satisfied: feedfinder2>=0.0.4 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (0.0.4)
Requirement already satisfied: jieba3k>=0.35.1 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (0.35.1)
Requirement already satisfied: python-dateutil>=2.5.3 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (2.8.2)
Requirement already satisfied: tinysegmenter==0.3 in /usr/local/lib/python3.10/dist-packages (from newspaper3k) (0.3)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4>=4.4.1->newspaper3k) (2.4.1)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from feedfinder2>=0.0.4->newspaper3k) (1.16.0)
Requirement already satisfied: sgmlib3k in /usr/local/lib/python3.10/dist-packages (from feedparser>=5.2.1->newspaper3k) (1.0.0)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk>=3.2.1->newspaper3k) (8.1.3)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk>=3.2.1->newspaper3k) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk>=3.2.1->newspaper3k) (2022.10.31)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk>=3.2.1->newspaper3k) (4.65.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.10.0->newspaper3k) (1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.10.0->newspaper3k) (2022.12.7)
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests>=2.10.0->newspaper3k) (2.0.12)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.10.0->newspaper3k) (3.4)
Requirement already satisfied: requests-file>=1.4 in /usr/local/lib/python3.10/dist-packages (from tldextract>=2.0.1->newspaper3k) (1.5.1)
Requirement already satisfied: filelock>=3.0.8 in /usr/local/lib/python3.10/dist-packages (from tldextract>=2.0.1->newspaper3k) (3.12.0)
```

▼ For the content using link

```
from newspaper import Article
```

```
url = 'https://www.indiatoday.in/cities/delhi/story/delhi-tops-list-world-most-polluted-cities-1989359-2022-08-18' article = Article(url)
```

```
article.download() article.parse()
```

```
article.text
```

```
doc=article.text
```

```
doc=''Global warming is the long-term warming of the planet's overall temperature. Though this war
The greenhouse effect is when the sun's rays penetrate the atmosphere, but when that heat is reflec
'''
```

```
doc
```

```
'Global warming is the long-term warming of the planet's overall temperature. Though th
is warming trend has been going on for a long time, its pace has significantly increase
d in the last hundred years due to the burning of fossil fuels. As the human population
has increased, so has the volume of fossil fuels burned. Fossil fuels include coal, oi
l, and natural gas, and burning them causes what is known as the "greenhouse effect" in
```

```
doc=nlp(doc)
```

```
len(list(doc.sents))
```

```
keyword=[]
stopwords=list(STOP_WORDS)
pos_tag=['PROPN','ADJ','NOUN','VERB']
for token in doc:
    if(token.text in stopwords or token.text in punctuation):
        continue
    if(token.pos_ in pos_tag):
        keyword.append(token.text)
```

```
print(stopwords)
```

```
['her', 'and', 'our', 'somehow', 'he', 'one', 'herself', 'although', 'in', 'twelve', 'hundred', 'whoever', 'due', 'still', 'whither', 't
```

```
freq_word = Counter(keyword)
print(freq_word.most_common(7))
```

```
[('fuels', 4), ('warming', 3), ('burning', 3), ('fossil', 3), ('atmosphere', 3), ('long', 2), ('increased', 2)]
```

```
sent_strength={}
for sent in doc.sents:
    for word in sent:
        if word.text in freq_word.keys():
            if sent in sent_strength.keys():
                sent_strength[sent]+=freq_word[word.text]
            else:
                sent_strength[sent]=freq_word[word.text]
print(sent_strength)
```

```
{Global warming is the long-term warming of the planet's overall temperature.: 13, Though this warming trend has been going on for a lon
: 23, The greenhouse effect is when the sun's rays penetrate the atmosphere, but when that heat is reflected off the surface cannot esc
: 19}
```

```
print("")
```

```
summarized_sentences=nlargest(1,sent_strength,key=sent_strength.get)
print(summarized_sentences)
```

```
[Fossil fuels include coal, oil, and natural gas, and burning them causes what is known as the "greenhouse effect" in Earth's atmosphere
]
```

```
final_sentence=(w.text for w in summarized_sentences)
summary=' '.join(final_sentence)
print(summary)
```

```
Fossil fuels include coal, oil, and natural gas, and burning them causes what is known as the "greenhouse effect" in Earth's atmosphere.
```

```
summary
```

```
'Fossil fuels include coal, oil, and natural gas, and burning them causes what is known
as the "greenhouse effect" in Earth's atmosphere.\n\n'
```

▼ Conversion of the text to audio starts from here

```
pip install gtts
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: gtts in /usr/local/lib/python3.10/dist-packages (2.3.2)
Requirement already satisfied: requests<3,>=2.27 in /usr/local/lib/python3.10/dist-packages (from gtts) (2.27.1)
Requirement already satisfied: click<8.2,>=7.1 in /usr/local/lib/python3.10/dist-packages (from gtts) (8.1.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.27->gtts) (1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.27->gtts) (2022.12.7)
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.27->gtts) (2.0.
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.27->gtts) (3.4)
```

Double-click (or enter) to edit

```
from gtts import gtts
```

```
import os
```

```
language='en'
myobj=gtts(text=summary,lang=language,slow=False)
myobj.save("output.mp3")
os.system("start output.mp3")
```

32512

```
pip install torch
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: torch in /usr/local/lib/python3.10/dist-packages (2.0.0+cu118)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from torch) (3.12.0)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from torch) (4.5.0)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch) (1.11.1)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch) (3.1)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from torch) (3.1.2)
Requirement already satisfied: triton==2.0.0 in /usr/local/lib/python3.10/dist-packages (from torch) (2.0.0)
Requirement already satisfied: cmake in /usr/local/lib/python3.10/dist-packages (from triton==2.0.0->torch) (3.25.2)
Requirement already satisfied: lit in /usr/local/lib/python3.10/dist-packages (from triton==2.0.0->torch) (16.0.2)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2->torch) (2.1.2)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy->torch) (1.3.0)
```

```
import torch
```

```
from transformers import PegasusForConditionalGeneration, PegasusTokenizer
```

```
model_name = 'tuner007/pegasus_paraphrase'
torch_device = 'cuda' if torch.cuda.is_available() else 'cpu'
tokenizer = PegasusTokenizer.from_pretrained(model_name)
model = PegasusForConditionalGeneration.from_pretrained(model_name).to(torch_device)
```

```
def get_response(input_text, num_return_sequences):
```

```
    batch = tokenizer.prepare_seq2seq_batch([input_text], truncation=True, padding='longest', max_1
    translated = model.generate(**batch, max_length=60, num_beams=10, num_return_sequences=num_retu
    tgt_text = tokenizer.batch_decode(translated, skip_special_tokens=True)
    return tgt_text
```

```
get_response(summary,5)
```

```
/usr/local/lib/python3.10/dist-packages/transformers/tokenization_utils_base.py:3722: FutureWarning:
`prepare_seq2seq_batch` is deprecated and will be removed in version 5 of HuggingFace Transformers. Use the regular
```

```
`__call__` method to prepare your inputs and targets.
```

Here is a short example:

```
model_inputs = tokenizer(src_texts, text_target=tgt_texts, ...)
```

If you either need to use different keyword arguments for the source and target texts, you should do two calls like this:

```
model_inputs = tokenizer(src_texts, ...)
labels = tokenizer(text_target=tgt_texts, ...)
model_inputs["labels"] = labels["input_ids"]
```

See the documentation of your specific tokenizer for more details on the specific arguments to the tokenizer of choice. For a more complete example, see the implementation of `prepare_seq2seq_batch`.

```
warnings.warn(formatted_warning, FutureWarning)
```

```
from sentence_splitter import SentenceSplitter, split_text_into_sentences
```

```
splitter=SentenceSplitter(language='en')
```

```
sentence_list=splitter.split(summary)
sentence_list
```

```
paraphrase = []
```

```
for i in sentence_list:
    a = get_response(i,1)
    paraphrase.append(a)
```

```
paraphrase
```

```
[[ 'Fossil fuels include coal, oil, and natural gas, and burning them causes what is known as the "greenhouse effect" in Earth\'s atmosphere.' ]]
```

```
paraphrase2=[' '.join(x) for x in paraphrase]
paraphrase2
```

```
[ 'Policies to reduce air pollution offer a win-win strategy for both climate and health, lowering the burden of disease attributable to air pollution, as well as contributing to the near- and long-term mitigation of climate change.' ]
```

```
paraphrase3=[ ' '.join(x for x in paraphrase2)]
paraphrased_text = str(paraphrase3).strip('[]').strip('"')
paraphrased_text
```

```
'Policies to reduce air pollution offer a win-win strategy for both climate and health, lowering the burden of disease attributable to air pollution, as well as contributing to the near- and long-term mitigation of climate change.'
```

```
paraphrased_text
```

```
'Policies to reduce air pollution offer a win-win strategy for both climate and health, lowering the burden of disease attributable to air pollution, as well as contributing to the near- and long-term mitigation of climate change.'
```

▼ pptx module starts from here

```
pip install python-pptx
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting python-pptx
  Downloading python-pptx-0.6.21.tar.gz (10.1 MB)
    10.1/10.1 MB 48.3 MB/s eta 0:00:00
```

```

Preparing metadata (setup.py) ... done
Requirement already satisfied: lxml>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from python-pptx) (4.9.2)
Requirement already satisfied: Pillow>=3.3.2 in /usr/local/lib/python3.10/dist-packages (from python-pptx) (8.4.0)
Collecting XlsxWriter>=0.5.7 (from python-pptx)
  Downloading XlsxWriter-3.1.0-py3-none-any.whl (152 kB)
    |----- 152.7/152.7 kB 13.8 MB/s eta 0:00:00
Building wheels for collected packages: python-pptx
  Building wheel for python-pptx (setup.py) ... done
  Created wheel for python-pptx: filename=python_pptx-0.6.21-py3-none-any.whl size=470949 sha256=ef0f77c4451625a8994d28941cb8b205f9da7d7
  Stored in directory: /root/.cache/pip/wheels/ea/dd/74/01b3ec7256a0800b99384e9a0f7620e358afc3a51a59bf9b49
Successfully built python-pptx
Installing collected packages: XlsxWriter, python-pptx
Successfully installed XlsxWriter-3.1.0 python-pptx-0.6.21

```

```

from pptx import Presentation
root = Presentation()

first_slide_layout = root.slide_layouts[1]

slide = root.slides.add_slide(first_slide_layout)

slide.shapes.title.text = "Summarized text in presentation"

slide.placeholders[1].text = paraphrased_text

root.save("Output.pptx")

print("done")

```

done

```

from pptx import Presentation
from pptx.util import Inches

```

```

Second_Layout = root.slide_layouts[5]
second_slide=root.slides.add_slide(Second_Layout)

```

```

second_slide.shapes.title.text="Second slide"

```

```

textbox=second_slide.shapes.add_textbox(Inches(3),Inches(1.5),Inches(3),Inches(1))

```

```

textframe=textbox.text_frame

```

```

paragraph= textframe.add_paragraph()
paragraph.text="The content for the second slide will be present over here."

```

```

root.save("final.pptx")

```

