```
!pip install pytextrank
```

```
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.1
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.11/dist
```

```
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dis
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/d
Requirement already satisfied: pydantic-core==2.27.2 in /usr/local/lib/python3.11/di
Requirement already satisfied: typing-extensions>=4.12.2 in /usr/local/lib/python3.1
```

```
Downloading executing-2.2.0-py2.py3-none-any.whl (26 kB)
Installing collected packages: executing, colorama, asttokens, icecream, pytextrank
Successfully installed asttokens-3.0.0 colorama-0.4.6 executing-2.2.0 icecream-2.1.4
```

```
import spacy
import pytextrank
```

```
/usr/local/lib/python3.11/dist-packages
```

```
document="""Not only did it only confirm that the film would be
 unfunny and genric,but it also managed to give away the
 ENTIRE movie; and I'm not exaggerating - every moment,every plot point,
 every joke is told in the trailer."""
```

```
import en_core_web_sm
nlp = en_core_web_sm.load()
```

```
en_nlp = spacy.load("en_core_web_sm")
en_nlp.add_pipe("textrank")
doc = en_nlp(document)
```

```
/usr/local/lib/python3.11/dist-packages/spacy/util.py:1740: UserWarning: [W111] Jupyter
  warnings.warn(Warnings.W111)
```

```
tr=doc._.textrank
print(tr.elapsed_time)
```

```
21.96216583251953
```

```
for combination in doc._.phrases:
  print(combination.text,combination.rank,combination.count)
```

```
ENTIRE 0.13514348101679782 1
the
 ENTIRE movie 0.06930139138062855 1
every joke 0.05936552514177136 1
the film 0.05423292745389326 1
the trailer 0.04834919915077192 1
I 0.0 1
it 0.0 2
```

```
from bs4 import BeautifulSoup
from urllib.request import urlopen
```

```
def get_only_text(url):
  page=urlopen(url)
  soup=BeautifulSoup(page)
  text='\t'.join(map(lambda p:p.text,soup.find_all('p')))
  print(text)
  return soup.title.text,text
```

```
url="https://en.wikipedia.org/wiki/Natural_language_processing"
text=get_only_text(url)
```

```
Natural language processing (NLP) is a subfield of computer science and especially arti
        Major tasks in natural language processing are speech recognition, text classif
        Natural language processing has its roots in the 1950s.[1] Already in 1950, Ala
        The premise of symbolic NLP is well-summarized by John Searle's Chinese room ex
        Up until the 1980s, most natural language processing systems were based on comp
        Symbolic approach, i.e., the hand-coding of a set of rules for manipulating syn
        Machine learning approaches, which include both statistical and neural networks
        Although rule-based systems for manipulating symbols were still in use in 2020,
        Before that they were commonly used:
        In the late 1980s and mid-1990s, the statistical approach ended a period of AI
        The earliest decision trees, producing systems of hard if-then rules, were stil
    Only the introduction of hidden Markov models, applied to part-of-speech tagging, annou
        A major drawback of statistical methods is that they require elaborate feature
        Intermediate tasks (e.g., part-of-speech tagging and dependency parsing) are no
        Neural machine translation, based on then-newly invented sequence-to-sequence t
```

```
The following is a list of some of the most commonly researched tasks in natura
Though natural language processing tasks are closely intertwined, they can be s
Based on long-standing trends in the field, it is possible to extrapolate futur
Most higher-level NLP applications involve aspects that emulate intelligent beh
Cognition refers to "the mental action or process of acquiring knowledge and ur
As an example, George Lakoff offers a methodology to build natural language pro
Ties with cognitive linguistics are part of the historical heritage of NLP, but
```

```
len(''.join(text))
```

```
6587
```

```
text[:1000]
```

```
('Natural language processing - Wikipedia',
 'Natural language processing (NLP) is a subfield of computer science and
especially artificial intelligence. It is primarily concerned with providing
computers with the ability to process data encoded in natural language and is thus
closely related to information retrieval, knowledge representation and
computational linguistics, a subfield of linguistics. Typically data is collected
in text corpora, using either rule-based, statistical or neural-based approaches in
machine learning and deep learning.\n\tMajor tasks in natural language processing
are speech recognition, text classification, natural-language understanding, and
natural-language generation.\n\tNatural language processing has its roots in the
```

```
1950s.[1] Already in 1950, Alan Turing published an article titled "Computing
Machinery and Intelligence" which proposed what is now called the Turing test as a
criterion of intelligence, though at the time that was not articulated as a problem
separate from artificial intelligence. The proposed test includes a task that
involves the automated interpretation and generation of natural language.\n\tThe
premise of symbolic NLP is well-summarized by John Searle\'s Chinese room
experiment: Given a collection of rules (e.g., a Chinese phrasebook, with questions
and matching answers), the computer emulates natural language understanding (or
other NLP tasks) by applying those rules to the data it confronts.\n\tUp until the
1980s, most natural language processing systems were based on complex sets of hand-
written rules.  Starting in the late 1980s, however, there was a revolution in
natural language processing with the introduction of machine learning algorithms
for language processing.  This was due to both the steady increase in computational
power (see Moore\'s law) and the gradual lessening of the dominance of Chomskyan
theories of linguistics (e.g. transformational grammar), whose theoretical
underpinnings discouraged the sort of corpus linguistics that underlies the
machine-learning approach to language processing.[8]\n\tSymbolic approach, i.e.,
the hand-coding of a set of rules for manipulating symbols, coupled with a
dictionary lookup, was historically the first approach used both by AI in general
and by NLP in particular:[18][19] such as by writing grammars or devising heuristic
rules for stemming.\n\tMachine learning approaches, which include both statistical
and neural networks, on the other hand, have many advantages over the symbolic
approach: \n\tAlthough rule-based systems for manipulating symbols were still in
use in 2020, they have become mostly obsolete with the advance of LLMs in 2023.
\n\tBefore that they were commonly used:\n\tIn the late 1980s and mid-1990s, the
statistical approach ended a period of AI winter, which was caused by the
inefficiencies of the rule-based approaches.[20][21]\n\tThe earliest decision
trees, producing systems of hard if-then rules, were still very similar to the old
```

```
rule-based approaches.\nOnly the introduction of hidden Markov models, applied to
part-of-speech tagging, announced the end of the old rule-based approach.\n\tA
major drawback of statistical methods is that they require elaborate feature
engineering. Since 2015,[22] the statistical approach has been replaced by the
neural networks approach, using semantic networks[23] and word embeddings to
capture semantic properties of words.  \n\tIntermediate tasks (e.g., part-of-speech
tagging and dependency parsing) are not needed anymore. \n\tNeural machine
translation, based on then-newly invented sequence-to-sequence transformations,
made obsolete the intermediate steps, such as word alignment, previously necessary
for statistical machine translation.\n\tThe following is a list of some of the most
commonly researched tasks in natural language processing. Some of these tasks have
direct real-world applications, while others more commonly serve as subtasks that
are used to aid in solving larger tasks.\n\tThough natural language processing
tasks are closely intertwined, they can be subdivided into categories for
convenience. A coarse division is given below.\n\tBased on long-standing trends in
the field, it is possible to extrapolate future directions of NLP. As of 2020,
three trends among the topics of the long-standing series of CoNLL Shared Tasks can
be observed:[46]\n\tMost higher-level NLP applications involve aspects that emulate
intelligent behaviour and apparent comprehension of natural language. More broadly
speaking, the technical operationalization of increasingly advanced aspects of
```

```
!pip install sumy
```

```
Collecting sumy
  Downloading sumy-0.11.0-py2.py3-none-any.whl.metadata (7.5 kB)
Collecting docopt<0.7,>=0.6.1 (from sumy)
  Downloading docopt-0.6.2.tar.gz (25 kB)
  Preparing metadata (setup.py) ... done
```

```
Collecting breadability>=0.1.20 (from sumy)
  Downloading breadability-0.1.20.tar.gz (32 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: requests>=2.7.0 in /usr/local/lib/python3.11/dist-packag
Collecting pycountry>=18.2.23 (from sumy)
  Downloading pycountry-24.6.1-py3-none-any.whl.metadata (12 kB)
Requirement already satisfied: nltk>=3.0.2 in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: chardet in /usr/local/lib/python3.11/dist-packages (fror
Requirement already satisfied: lxml>=2.0 in /usr/local/lib/python3.11/dist-packages (fr
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from r
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nl
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/di
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-pac
Downloading sumy-0.11.0-py2.py3-none-any.whl (97 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 97.3/97.3 kB 6.0 MB/s eta 0:00:00
Downloading pycountry-24.6.1-py3-none-any.whl (6.3 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6.3/6.3 MB 21.0 MB/s eta 0:00:00
Building wheels for collected packages: breadability, docopt
  Building wheel for breadability (setup.py) ... done
  Created wheel for breadability: filename=breadability-0.1.20-py2.py3-none-any.whl siz
  Stored in directory: /root/.cache/pip/wheels/4d/57/58/7e3d7fedf51fe248b7fcee3df6945ae
  Building wheel for docopt (setup.py) ... done
  Created wheel for docopt: filename=docopt-0.6.2-py2.py3-none-any.whl size=13706 sha25
  Stored in directory: /root/.cache/pip/wheels/1a/b0/8c/4b75c4116c31f83c8f9f047231251e1
```

```
Successfully built breadability docopt
Installing collected packages: docopt, pycountry, breadability, sumy
Successfully installed breadability-0.1.20 docopt-0.6.2 pycountry-24.6.1 sumy-0.11.0
```

```
!pip install lxml-html-clean
from sumy.parsers.html import HtmlParser
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.lsa import LsaSummarizer as Summarizer
from sumy.nlp.stemmers import Stemmer
from sumy.utils import get_stop_words
from sumy.summarizers.lex_rank import LexRankSummarizer as Summarizer
from sumy.summarizers.lsa import LsaSummarizer
```

```
Requirement already satisfied: lxml-html-clean in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: lxml in /usr/local/lib/python3.11/dist-packages (from l
```

```
LANGUAGE="english"
SENTENCES_COUNT=10
url="https://en.wikipedia.org/wiki/Natural_language_processing"
parser=HtmlParser.from_url(url,Tokenizer(LANGUAGE))
summarizer=LsaSummarizer()
```

```
Summarizer=LsaSummarizer(Stemmer(LANGUAGE))
summarizer.stop_words=get_stop_words(LANGUAGE)
for sentence in Summarizer(parser.document,sentences_count=SENTENCES_COUNT):
    print(sentence)
```

```
Generally, this task is much more difficult than supervised learning, and typically pro
[25] For a language like English, this is fairly trivial, since words are usually separ
The applications of sentiment analysis are diverse, extending to tasks such as categori
Coreference resolution Given a sentence or larger chunk of text, determine which words
Natural-language understanding(NLU) Convert chunks of text into more formal representat
[44][45] Based on long-standing trends in the field, it is possible to extrapolate futu
PMID 33736486.^ Lee, Jennifer; Yang, Samuel; Holland-Hall, Cynthia; Sezgin, Emre; Gill,
Retrieved 5 December 2021.^ Yi, Chucai; Tian, Yingli(2012), "Assistive Text Reading fro
Advances in Neural Information Processing Systems.^ Kariampuzha, William; Alyea, Gioco
In some cases, there are few historical records on long-gone civilizations to serve as
```

```
import nltk
nltk.download('punkt_tab')
```

```
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt_tab.zip.
True
```

```
text="""A vaccine for the coronavirus will likely be ready by early 2021 but rolling it out
 a leading vaccine scientist told Bloomberg.India, which is host to some of the front-runne
 currently has no local infrastructure in place to go beyond immunizing babies and pregnant
 said Gagandeep Kang, professor of microbiology at the Vellore-based Christian Medical Coll
The timing of the vaccine is a contentious subject around the world.
 In the U.S., President Donald Trump has contradicted a top administration health expert by
In India, Prime Minister Narendra Modi’s government had promised an indigenous vaccine as e
 a claim the government and its apex medical research body has since walked back."""
```

```
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.lex_rank import LexRankSummarizer
```

```
parser=PlaintextParser.from_string(text,Tokenizer("english"))
```



```
from sumy.summarizers.lex_rank import LexRankSummarizer
from sumy.utils import get_stop_words
summarizer_lex=LexRankSummarizer()
```

```
summarizer_lex.stop_words=get_stop_words('english')
summary =summarizer_lex(parser.document,5)
```

```
lex_summary=""
for sentence in summary:
    lex_summary+=str(sentence)+" "
print(lex_summary)
```

```
A vaccine for the coronavirus will likely be ready by early 2021 but rolling it out sat
```

Start coding or generate with AI.