In [2]:
```
1  !pip install pytextrank
```

```
spacy>=3.0->pytextrank) (3.5)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\nihar\anaconda3\lib\site-packages (from requests<3.0.0,>
=2.13.0->spacy>=3.0->pytextrank) (1.26.11)
Requirement already satisfied: blis<1.3.0,>=1.2.0 in c:\users\nihar\appdata\roaming\python\python39\site-packages (from t
hinc<8.4.0,>=8.3.0->spacy>=3.0->pytextrank) (1.2.0)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\nihar\anaconda3\lib\site-packages (from thinc<8.4.0,>
=8.3.0->spacy>=3.0->pytextrank) (0.1.4)
Requirement already satisfied: click<9.0.0,>=7.1.1 in c:\users\nihar\anaconda3\lib\site-packages (from typer<1.0.0,>=0.3.
0->spacy>=3.0->pytextrank) (8.0.4)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in c:\users\nihar\anaconda3\lib\site-packages (from weasel<0.5.0,
>=0.1.0->spacy>=3.0->pytextrank) (5.2.1)
Requirement already satisfied: cloudpathlib<0.17.0,>=0.7.0 in c:\users\nihar\anaconda3\lib\site-packages (from weasel<0.
5.0,>=0.1.0->spacy>=3.0->pytextrank) (0.16.0)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\nihar\anaconda3\lib\site-packages (from jinja2->spacy>=3.0->p
ytextrank) (2.0.1)
Requirement already satisfied: six>=1.5 in c:\users\nihar\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplot
lib>=3.4->networkx[default]>=2.6->pytextrank) (1.16.0)
Installing collected packages: smmap, graphviz, executing, asttokens, icecream, gitdb, GitPython, pytextrank
Successfully installed GitPython-3.1.44 asttokens-3.0.0 executing-2.2.0 gitdb-4.0.12 graphviz-0.20.3 icecream-2.1.4 pytex
trank-3.3.0 smmap-5.0.2
```

In [3]:
```
1  !python -m spacy download en_core_web_sm
```

```
Collecting en-core-web-sm==3.8.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.8.0/en_core_web_sm-3.8.0-py3-non
e-any.whl (https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.8.0/en_core_web_sm-3.8.0-py3-none-a
ny.whl) (12.8 MB)
     ------------------------------------- 12.8/12.8 MB 167.4 kB/s eta 0:00:00
Installing collected packages: en-core-web-sm
  Attempting uninstall: en-core-web-sm
    Found existing installation: en-core-web-sm 3.7.1
    Uninstalling en-core-web-sm-3.7.1:
      Successfully uninstalled en-core-web-sm-3.7.1
Successfully installed en-core-web-sm-3.8.0
[+] Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
```

In [4]:
```
1  import spacy
2  import pytextrank
```

```
C:\Users\nihar\Anaconda3\lib\site-packages
```

In [5]:
```
1  document ="""Not only did it only confirm that the film would be unfunny and generic,but it also managed to give away th
```

In [6]:
```
1  en_nlp = spacy.load("en_core_web_sm")
2  en_nlp.add_pipe("textrank")
3  doc = en_nlp(document)
```

```
C:\Users\nihar\Anaconda3\lib\site-packages\spacy\util.py:910: UserWarning: [W095] Model 'en_core_web_sm' (3.8.0) was traine
d with spaCy v3.8.0 and may not be 100% compatible with the current version (3.7.2). If you see errors or degraded performa
nce, download a newer compatible model or retrain your custom model with the current spaCy version. For more details and av
ailable updates, run: python -m spacy validate
  warnings.warn(warn_msg)
```

In [7]:
```
1  tr = doc._.textrank
2  print(tr.elapsed_time)
```

```
29.731273651123047
```

In [9]:
```
1  for combination in doc._.phrases:
2      print(combination.text, combination.rank,combination.count)
```

```
ENTIRE 0.13514348101679782 1
the ENTIRE movie 0.09548608913294183 1
every joke 0.05936552514177136 1
the film 0.05423292745389326 1
the trailer 0.04834919915077192 1
I 0.0 1
it 0.0 2
```

In [10]:
```
1  from bs4 import BeautifulSoup
2  from urllib.request import urlopen
```

In [11]:
```
1  def get_only_text(url):
2      page = urlopen(url)
3      soup = BeautifulSoup(page)
4      text = '\t'.join(map(lambda p: p.text,soup.find_all('p')))
5      print(text)
6      return soup.title.text, text
```

In [12]:
```python
1  url="http://en.wikipedia.org/wiki/Natural_language_processing"
2  text = get_only_text(url)
```

of natural language. More broadly speaking, the technical operationalization of increasingly advanced aspects of cognitiv
e behaviour represents one of the developmental trajectories of NLP (see trends among CoNLL shared tasks above).
        Cognition refers to "the mental action or process of acquiring knowledge and understanding through thought, exper
ience, and the senses."[47] Cognitive science is the interdisciplinary, scientific study of the mind and its processes.[4
8] Cognitive linguistics is an interdisciplinary branch of linguistics, combining knowledge and research from both psycho
logy and linguistics.[49] Especially during the age of symbolic NLP, the area of computational linguistics maintained str
ong ties with cognitive studies.
        As an example, George Lakoff offers a methodology to build natural language processing (NLP) algorithms through t
he perspective of cognitive science, along with the findings of cognitive linguistics,[50] with two defining aspects:
        Ties with cognitive linguistics are part of the historical heritage of NLP, but they have been less frequently ad
dressed since the statistical turn during the 1990s. Nevertheless, approaches to develop cognitive models towards technic
ally operationalizable frameworks have been pursued in the context of various frameworks, e.g., of cognitive grammar,[53]
functional grammar,[54] construction grammar,[55] computational psycholinguistics and cognitive neuroscience (e.g., ACT-
R), however, with limited uptake in mainstream NLP (as measured by presence on major conferences[56] of the ACL). More re
cently, ideas of cognitive NLP have been revived as an approach to achieve explainability, e.g., under the notion of "cog
nitive AI".[57] Likewise, ideas of cognitive NLP are inherent to neural models multimodal NLP (although rarely made expli
cit)[58] and developments in artificial intelligence, specifically tools and technologies using large language model appr
oaches[59] and new directions in artificial general intelligence based on the free energy principle[60] by British neuros
cientist and theoretician at University College London Karl J. Friston.

In [13]:
```python
1  len(''.join(text))
```

Out[13]: 6587

In [14]:
```python
1  text[:1000]
```

Out[14]: ('Natural language processing - Wikipedia',
 'Natural language processing (NLP) is a subfield of computer science and especially artificial intelligence. It is primari
ly concerned with providing computers with the ability to process data encoded in natural language and is thus closely rela
ted 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 19
50, Alan Turing published an article titled "Computing Machinery and Intelligence" which proposed what is now called the Tu
ring 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 langua
ge.\n\tThe premise of symbolic NLP is well-summarized by John Searle\'s Chinese room experiment: Given a collection of rule
s (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 process
ing 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 Chomsky
an 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-co
ding 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 stemmin
g.\n\tMachine learning approaches, which include both statistical and neural networks, on the other hand, have many advanta
ges over the symbolic approach: \n\tAlthough rule-based systems for manipulating symbols were still in use in 2020, they ha
ve become mostly obsolete with the advance of LLMs in 2023. \n\tBefore that they were commonly used:\n\tIn the late 1980s a
nd mid-1990s, the statistical approach ended a period of AI winter, which was caused by the inefficiencies of the rule-base
d approaches.[20][21]\n\tThe earliest decision trees, producing systems of hard if–then rules, were still very similar to t
he 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 e
ngineering. Since 2015,[22] the statistical approach has been replaced by the neural networks approach, using semantic netw
orks[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-se
quence 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 processin
g. Some of these tasks have direct real-world applications, while others more commonly serve as subtasks that are used to a
id in solving larger tasks.\n\tThough natural language processing tasks are closely intertwined, they can be subdivided int
o categories for convenience. A coarse division is given below.\n\tBased on long-standing trends in the field, it is possib
le 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 adv
anced aspects of cognitive behaviour represents one of the developmental trajectories of NLP (see trends among CoNLL shared
tasks above).\n\tCognition refers to "the mental action or process of acquiring knowledge and understanding through though
t, experience, and the senses."[47] Cognitive science is the interdisciplinary, scientific study of the mind and its proces
ses.[48] Cognitive linguistics is an interdisciplinary branch of linguistics, combining knowledge and research from both ps
ychology and linguistics.[49] Especially during the age of symbolic NLP, the area of computational linguistics maintained s
trong ties with cognitive studies.\n\tAs an example, George Lakoff offers a methodology to build natural language processin
g (NLP) algorithms through the perspective of cognitive science, along with the findings of cognitive linguistics,[50] with
two defining aspects:\n\tTies with cognitive linguistics are part of the historical heritage of NLP, but they have been les
s frequently addressed since the statistical turn during the 1990s. Nevertheless, approaches to develop cognitive models to
wards technically operationalizable frameworks have been pursued in the context of various frameworks, e.g., of cognitive g
rammar,[53] functional grammar,[54] construction grammar,[55] computational psycholinguistics and cognitive neuroscience
(e.g., ACT-R), however, with limited uptake in mainstream NLP (as measured by presence on major conferences[56] of the AC
L). More recently, ideas of cognitive NLP have been revived as an approach to achieve explainability, e.g., under the notio
n of "cognitive AI".[57] Likewise, ideas of cognitive NLP are inherent to neural models multimodal NLP (although rarely mad
e explicit)[58] and developments in artificial intelligence, specifically tools and technologies using large language model
approaches[59] and new directions in artificial general intelligence based on the free energy principle[60] by British neur
oscientist and theoretician at University College London Karl J. Friston.\n')

In [15]:
```
1  !pip install sumy
```

```
Collecting sumy
  Downloading sumy-0.11.0-py2.py3-none-any.whl (97 kB)
     -------------------------------------- 97.3/97.3 kB 59.3 kB/s eta 0:00:00
Requirement already satisfied: requests>=2.7.0 in c:\users\nihar\anaconda3\lib\site-packages (from sumy) (2.28.1)
Collecting pycountry>=18.2.23
  Downloading pycountry-24.6.1-py3-none-any.whl (6.3 MB)
     -------------------------------------- 6.3/6.3 MB 167.8 kB/s eta 0:00:00
Collecting breadability>=0.1.20
  Downloading breadability-0.1.20.tar.gz (32 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Requirement already satisfied: nltk>=3.0.2 in c:\users\nihar\anaconda3\lib\site-packages (from sumy) (3.9.1)
Collecting docopt<0.7,>=0.6.1
  Downloading docopt-0.6.2.tar.gz (25 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Requirement already satisfied: chardet in c:\users\nihar\anaconda3\lib\site-packages (from breadability>=0.1.20->sumy) (4.
0.0)
Requirement already satisfied: lxml>=2.0 in c:\users\nihar\anaconda3\lib\site-packages (from breadability>=0.1.20->sumy)
(4.9.1)
Requirement already satisfied: joblib in c:\users\nihar\anaconda3\lib\site-packages (from nltk>=3.0.2->sumy) (1.1.0)
Requirement already satisfied: tqdm in c:\users\nihar\anaconda3\lib\site-packages (from nltk>=3.0.2->sumy) (4.64.1)
Requirement already satisfied: click in c:\users\nihar\anaconda3\lib\site-packages (from nltk>=3.0.2->sumy) (8.0.4)
Requirement already satisfied: regex>=2021.8.3 in c:\users\nihar\anaconda3\lib\site-packages (from nltk>=3.0.2->sumy) (202
2.7.9)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\nihar\anaconda3\lib\site-packages (from requests>=2.7.0->s
umy) (1.26.11)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\nihar\anaconda3\lib\site-packages (from requests>=2.7.0
->sumy) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\nihar\anaconda3\lib\site-packages (from requests>=2.7.0->sum
y) (2022.9.14)
Requirement already satisfied: idna<4,>=2.5 in c:\users\nihar\anaconda3\lib\site-packages (from requests>=2.7.0->sumy) (3.
3)
Requirement already satisfied: colorama in c:\users\nihar\anaconda3\lib\site-packages (from click->nltk>=3.0.2->sumy) (0.4.
6)
Building wheels for collected packages: breadability, docopt
  Building wheel for breadability (setup.py): started
  Building wheel for breadability (setup.py): finished with status 'done'
  Created wheel for breadability: filename=breadability-0.1.20-py2.py3-none-any.whl size=21695 sha256=ac14b9679c314cad40443
8fe572eb35549821a4d965d065c645d704437137730
  Stored in directory: c:\users\nihar\appdata\local\pip\cache\wheels\ba\9f\70\7795228568b81b57a8932755938da9fb1f291b0576752
604aa
  Building wheel for docopt (setup.py): started
  Building wheel for docopt (setup.py): finished with status 'done'
  Created wheel for docopt: filename=docopt-0.6.2-py2.py3-none-any.whl size=13706 sha256=9983f2b86ebacaabb4bef14024a1386c01
2f16295483f4bdf72471755ad40c08
  Stored in directory: c:\users\nihar\appdata\local\pip\cache\wheels\70\4a\46\1309fc853b8d395e60bafaf1b6df7845bdd82c95fd59d
d8d2b
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
```

In [16]:
```
1  from sumy.parsers.html import HtmlParser
2  from sumy.parsers.plaintext import PlaintextParser
3  from sumy.nlp.tokenizers import Tokenizer
4  from sumy.summarizers.lsa import LsaSummarizer
5  from sumy.nlp.stemmers import Stemmer
6  from sumy.utils import get_stop_words
7  from sumy.summarizers.luhn import LuhnSummarizer
```

```
In [17]:   1  LANGUAGE = "english"
           2  SENTENCES_COUNT = 10
           3  url ="http://en.wikipedia.org/wiki/Natural_language_processing"
           4  parser = HtmlParser.from_url(url, Tokenizer(LANGUAGE))
           5  summarizer = LsaSummarizer()
           6  summarizer = LsaSummarizer(Stemmer(LANGUAGE))
           7  summarizer.stop_words = get_stop_words(LANGUAGE)
           8  for sentence in summarizer(parser.document, SENTENCES_COUNT):
           9      print(sentence)
```

[2] However, real progress was much slower, and after the ALPAC report in 1966, which found that ten years of research had failed to fulfill the expectations, funding for machine translation was dramatically reduced.
However, there is an enormous amount of non-annotated data available (including, among other things, the entire content of the World Wide Web), which can often make up for the worse efficiency if the algorithm used has a low enough time complexity to be practical.
[14] This is increasingly important in medicine and healthcare, where NLP helps analyze notes and text in electronic health records that would otherwise be inaccessible for study when seeking to improve care[16] or protect patient privacy.
the larger such a (probabilistic) language model is, the more accurate it becomes, in contrast to rule-based systems that can gain accuracy only by increasing the amount and complexity of the rules leading to intractability problems.
[34][35][36] As far as orthography, morphology, syntax and certain aspects of semantics are concerned, and due to the development of powerful neural language models such as GPT-2, this can now (2019) be considered a largely solved problem and is being marketed in various commercial applications.
Increasing interest in multilinguality, and, potentially, multimodality (English since 1999; Spanish, Dutch since 2002; German since 2003; Bulgarian, Danish, Japanese, Portuguese, Slovenian, Swedish, Turkish since 2006; Basque, Catalan, Chinese, Greek, Hungarian, Italian, Turkish since 2007; Czech since 2009; Arabic since 2012; 2017: 40+ languages; 2018: 60+/100+ languages) Elimination of symbolic representations (rule-based over supervised towards weakly supervised methods, representation learning and end-to-end systems)
More broadly speaking, the technical operationalization of increasingly advanced aspects of cognitive behaviour represents one of the developmental trajectories of NLP (see trends among CoNLL shared tasks above).
PMID 33736486.^ Lee, Jennifer; Yang, Samuel; Holland-Hall, Cynthia; Sezgin, Emre; Gill, Manjot; Linwood, Simon; Huang, Yungui; Hoffman, Jeffrey (2022-06-10).
Retrieved 5 December 2021.^ Yi, Chucai; Tian, Yingli(2012), "Assistive Text Reading from Complex Background for Blind Persons", Camera-Based Document Analysis and Recognition, Lecture Notes in Computer Science, vol.
Advances in Neural Information Processing Systems.^ Kariampuzha, William; Alyea, Gioconda; Qu, Sue; Sanjak, Jaleal; Mathé, Ewy; Sid, Eric; Chatelaine, Haley; Yadaw, Arjun; Xu, Yanji; Zhu, Qian (2023).

```
In [18]:   1  text = """A vaccine for the coronavirus will likely be ready by early 2021 but rolling it out safely across India's 1.3
           2
           3  India, which is host to some of the front-runner vaccine clinical trials, currently has no local infrastructure in place
           4
           5  The timing of the vaccine is a contentious subject around the world. In the U.S., President Donald Trump has contradicte
```

```
In [19]:   1  from sumy.parsers.plaintext import PlaintextParser
           2  from sumy.nlp.tokenizers import Tokenizer
```

```
In [20]:   1  parser = PlaintextParser.from_string(text,Tokenizer("english"))
```

```
In [21]:   1  from sumy.summarizers.lex_rank import LexRankSummarizer
           2  from sumy.utils import get_stop_words
           3  summarizer_lex = LexRankSummarizer()
           4  summarizer_lex.stop_words = get_stop_words("english")
           5  summary = summarizer_lex(parser.document,5)
           6  lex_summary =""
           7  for sentence in summary:
           8      lex_summary+=str(sentence)
           9  print(lex_summary)
```

"A vaccine for the coronavirus will likely be ready by early 2021 but rolling it out safely across India's 1.3 billion people will be the country's biggest challenge in fighting its surging epidemic, a leading vaccine scientist told Bloomberg.India, which is host to some of the front-runner vaccine clinical trials, currently has no local infrastructure in place to go beyond immunizing babies and pregnant women, said Gagandeep Kang, professor of microbiology at the Vellore-based Christian Medical College and a member of the WHO's Global Advisory Committee on Vaccine Safety.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 saying a vaccine would be available by October.In India, Prime Minister Narendra Modi's government had promised an indigenous vaccine as early as mid-August, a claim the government and its apex medical research body has since walked back.

```
In [ ]:   1
```