In [14]:
```
!pip install pytextrank
```

```
\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy>=3.0->pytextrank) (2.27.2)
Requirement already satisfied: typing-extensions>=4.12.2 in c:\users\pooja reddy\anaconda3
\lib\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy>=3.0->pytextrank) (4.1
2.2)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\pooja reddy\anaconda3\l
ib\site-packages (from requests<3.0.0,>=2.13.0->spacy>=3.0->pytextrank) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\pooja reddy\anaconda3\lib\site-pack
ages (from requests<3.0.0,>=2.13.0->spacy>=3.0->pytextrank) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\pooja reddy\anaconda3\lib\sit
e-packages (from requests<3.0.0,>=2.13.0->spacy>=3.0->pytextrank) (1.26.16)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\pooja reddy\anaconda3\lib\sit
e-packages (from requests<3.0.0,>=2.13.0->spacy>=3.0->pytextrank) (2023.7.22)
Requirement already satisfied: blis<1.3.0,>=1.2.0 in c:\users\pooja reddy\anaconda3\lib\sit
e-packages (from thinc<8.4.0,>=8.3.4->spacy>=3.0->pytextrank) (1.2.0)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\pooja reddy\anaconda3\l
ib\site-packages (from thinc<8.4.0,>=8.3.4->spacy>=3.0->pytextrank) (0.1.5)
Requirement already satisfied: click>=8.0.0 in c:\users\pooja reddy\anaconda3\lib\site-pack
ages (from typer<1.0.0,>=0.3.0->spacy>=3.0->pytextrank) (8.0.4)
Requirement already satisfied: shellingham>=1.3.0 in c:\users\pooja reddy\anaconda3\lib\sit
e-packages (from typer<1.0.0,>=0.3.0->spacy>=3.0->pytextrank) (1.5.4)
```

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

In [16]:
```
import spacy
import pytextrank
```

```
C:\Users\Pooja Reddy\anaconda3\Lib\site-packages
```

In [17]:
```
document = """Not only did it only confirm that the film would be unfunny and generic , but it
and I'm not exaggerating - every moment, every plot point, every joke is told in the trailer.""
```

In [18]:
```
en_nlp = spacy.load("en_core_web_sm")
en_nlp.add_pipe("textrank")
doc = en_nlp(document)
```

In [21]:
```
tr = doc._.textrank
print(tr.elapsed_time)
```

```
21.811246871948242
```

In [23]:
```
for combination in doc._.phrases:
    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 [24]:
```
from bs4 import BeautifulSoup
from urllib.request import urlopen
```

In [25]:
```
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
```

In [27]:
```
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 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.

Major tasks in natural language processing are speech recognition, text classification, natural-language understanding, and natural-language generation.

Natural 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.

The 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.

Up 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]

Symbolic 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.

Machine learning approaches, which include both statistical and neural networks, on the other hand, have many advantages over the symbolic approach:

Although rule-based systems for manipulating symbols were still in use in 2020, they have become mostly obsolete with the advance of LLMs in 2023.

Before that they were commonly used:

In 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]

The earliest decision trees, producing systems of hard if–then rules, were still very similar to the old rule-based approaches.

Only the introduction of hidden Markov models, applied to part-of-speech tagging, announced the end of the old rule-based approach.

A 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.

Intermediate tasks (e.g., part-of-speech tagging and dependency parsing) are not needed anymore.

Neural 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.

The 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.

Though natural language processing tasks are closely intertwined, they can be subdivided into categories for convenience. A coarse division is given below.

Based 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]

Most 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 cognitive 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, experience, and the senses."[47] Cognitive science is the interdisciplinary, scientific study of the mind and its processes.[48] Cognitive linguistics is an interdisciplinary branch of linguistics, combining knowledge and research from both psychology and linguistics.[49] Especially during the age of symbolic NLP, the area of computational linguistics maintained strong ties with cognitive studies.

As an example, George Lakoff offers a methodology to build natural language processing (NLP) algorithms through the 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 addressed since the statistical turn during the 1990s. Nevertheless, approaches to develop cognitive models towards technically operationalizable frameworks have been pursued in the context of various frameworks, e.g., of cognitive grammar,[53] functional g

```
rammar,[54] construction grammar,[55] computational psycholinguistics and cognitive neuroscien
ce (e.g., ACT-R), however, with limited uptake in mainstream NLP (as measured by presence on m
ajor conferences[56] of the ACL). More recently, ideas of cognitive NLP have been revived as a
n approach to achieve explainability, e.g., under the notion of "cognitive AI".[57] Likewise,
ideas of cognitive NLP are inherent to neural models multimodal NLP (although rarely made expl
icit)[58] and developments in artificial intelligence, specifically tools and technologies usi
ng large language model approaches[59] and new directions in artificial general intelligence b
ased on the free energy principle[60] by British neuroscientist and theoretician at University
College London Karl J. Friston.
```

In [28]:
```
len(''.join(text))
```

Out[28]:
```
6587
```

In [29]:
```
text[:1000]
```

Out[29]:
```
('Natural language processing - Wikipedia',
 'Natural language processing (NLP) is a subfield of computer science and especially artificia
l 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, knowled
ge 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 r
ecognition, text classification, natural-language understanding, and natural-language generati
on.\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 articul
ated as a problem separate from artificial intelligence. The proposed test includes a task tha
t 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 emula
tes 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 compl
ex 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 la
w) and the gradual lessening of the dominance of Chomskyan theories of linguistics (e.g. trans
formational grammar), whose theoretical underpinnings discouraged the sort of corpus linguisti
cs that underlies the machine-learning approach to language processing.[8]\n\tSymbolic approac
h, 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 particula
r:[18][19] such as by writing grammars or devising heuristic rules for stemming.\n\tMachine le
arning 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 s
ymbols 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 statist
ical 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 ru
les, were still very similar to the old rule-based approaches.\nOnly the introduction of hidde
n Markov models, applied to part-of-speech tagging, announced the end of the old rule-based ap
proach.\n\tA major drawback of statistical methods is that they require elaborate feature engi
neering. Since 2015,[22] the statistical approach has been replaced by the neural networks app
roach, using semantic networks[23] and word embeddings to capture semantic properties of word
s.  \n\tIntermediate tasks (e.g., part-of-speech tagging and dependency parsing) are not neede
d anymore. \n\tNeural machine translation, based on then-newly invented sequence-to-sequence t
ransformations, made obsolete the intermediate steps, such as word alignment, previously neces
sary for statistical machine translation.\n\tThe following is a list of some of the most commo
nly researched tasks in natural language processing. Some of these tasks have direct real-worl
d applications, while others more commonly serve as subtasks that are used to aid in solving l
arger 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 2
020, three trends among the topics of the long-standing series of CoNLL Shared Tasks can be ob
served:[46]\n\tMost higher-level NLP applications involve aspects that emulate intelligent beh
aviour and apparent comprehension of natural language. More broadly speaking, the technical op
erationalization of increasingly advanced aspects of cognitive behaviour represents one of the
developmental trajectories of NLP (see trends among CoNLL shared tasks above).\n\tCognition re
fers 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 stu
dy of the mind and its processes.[48] Cognitive linguistics is an interdisciplinary branch of
linguistics, combining knowledge and research from both psychology and linguistics.[49] Especi
ally during the age of symbolic NLP, the area of computational linguistics maintained strong t
ies with cognitive studies.\n\tAs an example, George Lakoff offers a methodology to build natu
ral language processing (NLP) algorithms through the perspective of cognitive science, along w
ith the findings of cognitive linguistics,[50] with two defining aspects:\n\tTies with cogniti
ve linguistics are part of the historical heritage of NLP, but they have been less frequently
addressed since the statistical turn during the 1990s. Nevertheless, approaches to develop cog
nitive models towards technically operationalizable frameworks have been pursued in the contex
t of various frameworks, e.g., of cognitive grammar,[53] functional grammar,[54] construction
grammar,[55] computational psycholinguistics and cognitive neuroscience (e.g., ACT-R), howeve
r, with limited uptake in mainstream NLP (as measured by presence on major conferences[56] of
the ACL). More recently, ideas of cognitive NLP have been revived as an approach to achieve ex
plainability, e.g., under the notion of "cognitive AI".[57] Likewise, ideas of cognitive NLP a
re inherent to neural models multimodal NLP (although rarely made explicit)[58] and developmen
ts 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 neuroscientist and theoretician at University College London Karl J.
Friston.\n')
```

In [30]:
```
!pip install sumy
```

```
Collecting sumy
  Obtaining dependency information for sumy from https://files.pythonhosted.org/packages/19/46/77859104e7c3e12dfa2e5c0e27b5dd1e14cb2409b50f9c936a48f29ceaee/sumy-0.11.0-py2.py3-none-any.whl.metadata (https://files.pythonhosted.org/packages/19/46/77859104e7c3e12dfa2e5c0e27b5dd1e14cb2409b50f9c936a48f29ceaee/sumy-0.11.0-py2.py3-none-any.whl.metadata)
  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): started
  Preparing metadata (setup.py): finished with status 'done'
Collecting breadability>=0.1.20 (from sumy)
  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: requests>=2.7.0 in c:\users\pooja reddy\anaconda3\lib\site-packages (from sumy) (2.31.0)
Collecting pycountry>=18.2.23 (from sumy)
  Obtaining dependency information for pycountry>=18.2.23 from https://files.pythonhosted.org/packages/b1/ec/1fb891d8a2660716aadb2143235481d15ed1cbfe3ad669194690b0604492/pycountry-2
```

In [31]:
```
from sumy.parsers.html import HtmlParser
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.lsa import LsaSummarizer
from sumy.nlp.stemmers import Stemmer
from sumy.utils import get_stop_words
from sumy.summarizers.luhn import LuhnSummarizer
```

In [35]:
```
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):
    print(sentence)
```

```
[ 2] However, real progress was much slower, and after the ALPAC report in 1966, which found t
hat ten years of research had failed to fulfill the expectations, funding for machine translat
ion was dramatically reduced.
However, there is an enormous amount of non-annotated data available (including, among other t
hings, the entire content of the World Wide Web), which can often make up for the worse effici
ency 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 seek
ing 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 t
he rules leading to intractability problems.
[34][35][36] As far as orthography, morphology, syntax and certain aspects of semantics are co
ncerned, 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; S
panish, Dutch since 2002; German since 2003; Bulgarian, Danish, Japanese, Portuguese, Slovenia
n, Swedish, Turkish since 2006; Basque, Catalan, Chinese, Greek, Hungarian, Italian, Turkish s
ince 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 co
gnitive behaviour represents one of the developmental trajectories of NLP (see trends among Co
NLL shared tasks above).
PMID 33736486.^ Lee, Jennifer; Yang, Samuel; Holland-Hall, Cynthia; Sezgin, Emre; Gill, Manjo
t; Linwood, Simon; Huang, Yungui; Hoffman, Jeffrey (2022-06-10).
Retrieved 5 December 2021.^ Yi, Chucai; Tian, Yingli(2012), "Assistive Text Reading from Compl
ex Background for Blind Persons", Camera-Based Document Analysis and Recognition, Lecture Note
s 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, Q
ian (2023).
```

In [40]:
```
text="""A vaccine for the coronavirus will likely be ready by early 2021 but rolling it out saf
The timing of the vaccine is a contentious subject around the world. In the U.S., President Don
"""
```

In [41]:
```
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
```

In [42]:
```
parser = PlaintextParser.from_string(text, Tokenizer("english"))
```

In [43]:
```
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 safely acr
oss 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 i
mmunizing babies and pregnant women, said Gagandeep Kang, professor of microbiology at the Vel
lore-based Christian Medical College and a member of the WHO’s Global Advisory Committee on Va
ccine 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 rese
arch body has since walked back."
```

In [ ]: