

text-summerization-using-spacy-lib

May 28, 2024

```
[ ]: #!/pip install spacy  
 #(we have 3 type of model is available called small, medium & large)  
#!/python -m spacy download en_core_web_sm
```

```
[1]: import spacy  
nlp = spacy.load("en_core_web_sm")  
doc = nlp("data science and ai has greate career ahead")
```

```
[2]: doc
```

```
[2]: data science and ai has greate career ahead
```

```
[4]: for token in doc:  
    print(token.text)
```

```
data  
science  
and  
ai  
has  
greate  
career  
ahead
```

```
[5]: import spacy  
nlp = spacy.load("en_core_web_sm")  
doc = nlp("Apple is looking at buying U.K. startup for $1 billion")
```

```
[6]: doc
```

```
[6]: Apple is looking at buying U.K. startup for $1 billion
```

```
[11]: for token in doc:  
    print(token.text, token.lemma_, token.pos_, token.tag_, token.dep_, token.  
        ↪shape_, token.is_alpha, token.is_stop)
```

```
Apple Apple PROPN NNP nsubj Xxxxx True False  
is be AUX VBZ aux xx True True
```

```
looking look VERB VBG ROOT xxxx True False
at at ADP IN prep xx True True
buying buy VERB VBG pcomp xxxx True False
U.K. U.K. PROPN NNP dobj X.X. False False
startup startup NOUN NN dep xxxx True False
for for ADP IN prep xxx True True
$ $ SYM $ quantmod $ False False
1 1 NUM CD compound d False False
billion billion NUM CD pobj xxxx True False
```

```
[13]: for token in doc:
      print(token.pos_)
```

```
PROPN
AUX
VERB
ADP
VERB
PROPN
NOUN
ADP
SYM
NUM
NUM
```

```
[14]: for token in doc:
      print(token.text, token.pos_)
```

```
Apple PROPN
is AUX
looking VERB
at ADP
buying VERB
U.K. PROPN
startup NOUN
for ADP
$ SYM
1 NUM
billion NUM
```

```
[15]: for token in doc:
      print(token.text, token.pos_, token.lemma_)
```

```
Apple PROPN Apple
is AUX be
looking VERB look
at ADP at
buying VERB buy
```

U.K. PROP N U.K.
startup NOUN startup
for ADP for
\$ SYM \$
1 NUM 1
billion NUM billion

```
[29]: text = '''There once was a boy who grew bored while watching over the village_
↳sheep. He wanted to make things more exciting.
So, he yelled out that he saw a wolf chasing the sheep. All the villagers came_
↳running to drive the wolf away.
However, they saw no wolf. The boy was amused, but the villagers were not. They_
↳told him not to do it again.
Shortly after, he repeated this antic. The villagers came running again, only_
↳to find that he was lying.
Later that day, the boy really sees a wolf sneaking amongst the flock. He_
↳jumped up and called out for help.
But no one came this time because they thought he was still joking around. At_
↳sunset, the villagers looked for the boy.
He had not returned with their sheep. They found him crying. He told them that_
↳there really was a wolf, and the entire flock was gone.
An old man came to comfort him and told him that nobody would believe a liar_
↳even when they are being honest.'''
```

```
[30]: text
```

```
[30]: 'There once was a boy who grew bored while watching over the village sheep. He
wanted to make things more exciting.\nSo, he yelled out that he saw a wolf
chasing the sheep. All the villagers came running to drive the wolf away.
\nHowever, they saw no wolf. The boy was amused, but the villagers were not.
They told him not to do it again. \nShortly after, he repeated this antic. The
villagers came running again, only to find that he was lying. \nLater that day,
the boy really sees a wolf sneaking amongst the flock. He jumped up and called
out for help.\nBut no one came this time because they thought he was still
joking around. At sunset, the villagers looked for the boy.\nHe had not returned
with their sheep. They found him crying. He told them that there really was a
wolf, and the entire flock was gone. \nAn old man came to comfort him and told
him that nobody would believe a liar even when they are being honest.'
```

```
[31]: import spacy
from spacy.lang.en.stop_words import STOP_WORDS
from string import punctuation
```

```
[32]: stopwords = list(STOP_WORDS)
stopwords
```

```
[32]: ['around',  
      'the',  
      'ours',  
      'must',  
      'more',  
      'name',  
      've',  
      'nothing',  
      'that',  
      'we',  
      'hence',  
      'really',  
      'my',  
      'towards',  
      'latter',  
      'serious',  
      'give',  
      'therefore',  
      'upon',  
      'everything',  
      'hereafter',  
      'nobody',  
      'may',  
      've',  
      'besides',  
      'side',  
      'each',  
      'against',  
      'part',  
      'between',  
      'thereafter',  
      'after',  
      'she',  
      'much',  
      'toward',  
      'well',  
      'seemed',  
      'whereupon',  
      'from',  
      'any',  
      'few',  
      'itself',  
      'enough',  
      'myself',  
      'cannot',  
      'a',  
      'into',
```

"n't",
'regarding',
'had',
"re",
'noone',
'would',
'whether',
'of',
'own',
'm',
"m",
'elsewhere',
'bottom',
'whenever',
'hundred',
'at',
'everyone',
'they',
'm',
'throughout',
'should',
'but',
'above',
'already',
'through',
'by',
'someone',
'further',
'd',
'thence',
'being',
'every',
'less',
'see',
'll',
'move',
'why',
'hereupon',
'though',
'moreover',
'can',
'did',
'six',
'thereby',
'another',
'various',
'many',

'll',
'yourselves',
'whatever',
'himself',
'within',
's',
'sixty',
'done',
'does',
'amount',
'yet',
'somehow',
'them',
'yourself',
'former',
'when',
'used',
'whereas',
'somewhere',
'four',
'something',
'put',
'least',
'made',
'these',
'out',
'what',
'either',
'therein',
'third',
'with',
'among',
'its',
'only',
'seem',
'eight',
'please',
'n't',
'because',
'up',
's',
'amongst',
'none',
'if',
'before',
'full',
'fifty',

'whereafter',
'beforehand',
'i',
'n't',
"'s",
'some',
'first',
'ever',
'keep',
'indeed',
'both',
'anyway',
'together',
'say',
'quite',
'no',
'top',
"'ve",
'he',
'have',
'whoever',
'to',
'across',
'do',
'along',
'anywhere',
'over',
'beside',
'via',
'nevertheless',
'will',
'latterly',
'me',
'due',
'hers',
'for',
'anyhow',
'again',
'all',
'five',
'several',
'mostly',
'could',
'forty',
'her',
'are',
'themselves',

'about',
'were',
'still',
'those',
'in',
'most',
'such',
'twelve',
'sometimes',
'how',
'our',
'below',
'nor',
'wherever',
'mine',
'thru',
'was',
'anyone',
'might',
'than',
'became',
'two',
'same',
'down',
'him',
'll',
'herself',
'neither',
'as',
'ourselves',
'rather',
'behind',
'herein',
'then',
'become',
'call',
're',
'until',
'make',
'others',
'now',
'almost',
'eleven',
'although',
'and',
'ten',
'your',

'afterwards',
'front',
'onto',
'here',
'sometime',
'unless',
'show',
'nowhere',
'next',
'alone',
'us',
'namely',
'whole',
'this',
'not',
'where',
'fifteen',
'it',
'ca',
'take',
'there',
'except',
'becoming',
'who',
'becomes',
'beyond',
'perhaps',
'which',
'their',
'yours',
're',
'using',
'empty',
'last',
'd',
'per',
'under',
'just',
'whence',
'other',
'is',
'or',
'an',
'so',
'whereby',
'since',
'very',

```
'you',  
'everywhere',  
'during',  
'twenty',  
'd',  
'whither',  
'thereupon',  
'however',  
'else',  
'one',  
'hereby',  
'be',  
'am',  
'doing',  
'seems',  
'while',  
'three',  
'seeming',  
'never',  
'off',  
'whose',  
'thus',  
'even',  
'his',  
'wherein',  
'get',  
'go',  
'anything',  
'formerly',  
'too',  
're',  
'without',  
'also',  
'once',  
'otherwise',  
'has',  
'meanwhile',  
'nine',  
'back',  
'been',  
'often',  
'always',  
'on',  
'whom']
```

```
[33]: len(stopwords)
```

[33]: 326

```
[34]: nlp = spacy.load('en_core_web_sm')
```

```
[35]: text
```

```
[35]: 'There once was a boy who grew bored while watching over the village sheep. He
wanted to make things more exciting.\nSo, he yelled out that he saw a wolf
chasing the sheep. All the villagers came running to drive the wolf away.
\nHowever, they saw no wolf. The boy was amused, but the villagers were not.
They told him not to do it again. \nShortly after, he repeated this antic. The
villagers came running again, only to find that he was lying. \nLater that day,
the boy really sees a wolf sneaking amongst the flock. He jumped up and called
out for help.\nBut no one came this time because they thought he was still
joking around. At sunset, the villagers looked for the boy.\nHe had not returned
with their sheep. They found him crying. He told them that there really was a
wolf, and the entire flock was gone. \nAn old man came to comfort him and told
him that nobody would believe a liar even when they are being honest.'
```

```
[36]: doc = nlp(text)
```

```
[37]: doc
```

```
[37]: There once was a boy who grew bored while watching over the village sheep. He
wanted to make things more exciting.
So, he yelled out that he saw a wolf chasing the sheep. All the villagers came
running to drive the wolf away.
However, they saw no wolf. The boy was amused, but the villagers were not. They
told him not to do it again.
Shortly after, he repeated this antic. The villagers came running again, only to
find that he was lying.
Later that day, the boy really sees a wolf sneaking amongst the flock. He jumped
up and called out for help.
But no one came this time because they thought he was still joking around. At
sunset, the villagers looked for the boy.
He had not returned with their sheep. They found him crying. He told them that
there really was a wolf, and the entire flock was gone.
An old man came to comfort him and told him that nobody would believe a liar
even when they are being honest.
```

```
[38]: # lets get the tokens from text
tokens = [token.text for token in doc]
print(tokens)

#when we execute everything we created tokens from the text & not removed any
↳ of the stopwords & didnt cleaned the data
```

```
[ 'There', 'once', 'was', 'a', 'boy', 'who', 'grew', 'bored', 'while',
'watching', 'over', 'the', 'village', 'sheep', '.', 'He', 'wanted', 'to',
'make', 'things', 'more', 'exciting', '.', '\n', 'So', ',', 'he', 'yelled',
'out', 'that', 'he', 'saw', 'a', 'wolf', 'chasing', 'the', 'sheep', '.', 'All',
'the', 'villagers', 'came', 'running', 'to', 'drive', 'the', 'wolf', 'away',
'.', '\n', 'However', ',', 'they', 'saw', 'no', 'wolf', '.', 'The', 'boy',
'was', 'amused', ',', 'but', 'the', 'villagers', 'were', 'not', '.', 'They',
'told', 'him', 'not', 'to', 'do', 'it', 'again', '.', '\n', 'Shortly', 'after',
',', 'he', 'repeated', 'this', 'antic', '.', 'The', 'villagers', 'came',
'running', 'again', ',', 'only', 'to', 'find', 'that', 'he', 'was', 'lying',
'.', '\n', 'Later', 'that', 'day', ',', 'the', 'boy', 'really', 'sees', 'a',
'wolf', 'sneaking', 'amongst', 'the', 'flock', '.', 'He', 'jumped', 'up', 'and',
'called', 'out', 'for', 'help', '.', '\n', 'But', 'no', 'one', 'came', 'this',
'time', 'because', 'they', 'thought', 'he', 'was', 'still', 'joking', 'around',
'.', 'At', 'sunset', ',', 'the', 'villagers', 'looked', 'for', 'the', 'boy',
'.', '\n', 'He', 'had', 'not', 'returned', 'with', 'their', 'sheep', '.',
'They', 'found', 'him', 'crying', '.', 'He', 'told', 'them', 'that', 'there',
'really', 'was', 'a', 'wolf', ',', 'and', 'the', 'entire', 'flock', 'was',
'gone', '.', '\n', 'An', 'old', 'man', 'came', 'to', 'comfort', 'him', 'and',
'told', 'him', 'that', 'nobody', 'would', 'believe', 'a', 'liar', 'even',
'when', 'they', 'are', 'being', 'honest', '.']
```

[39]: tokens

```
[39]: [ 'There',
'once',
'was',
'a',
'boy',
'who',
'grew',
'bored',
'while',
'watching',
'over',
'the',
'village',
'sheep',
'.',
'He',
'wanted',
'to',
'make',
'things',
'more',
'exciting',
'.',
```

'\n',
'So',
,',
'he',
'yelled',
'out',
'that',
'he',
'saw',
'a',
'wolf',
'chasing',
'the',
'sheep',
'.',
'All',
'the',
'villagers',
'came',
'running',
'to',
'drive',
'the',
'wolf',
'away',
'.',
'\n',
'However',
,',
'they',
'saw',
'no',
'wolf',
'.',
'The',
'boy',
'was',
'amused',
,',
'but',
'the',
'villagers',
'were',
'not',
'.',
'They',
'told',

'him',
'not',
'to',
'do',
'it',
'again',
'.',
'\n',
'Shortly',
'after',
'',
'he',
'repeated',
'this',
'antic',
'.',
'The',
'villagers',
'came',
'running',
'again',
'',
'only',
'to',
'find',
'that',
'he',
'was',
'lying',
'.',
'\n',
'Later',
'that',
'day',
'',
'the',
'boy',
'really',
'sees',
'a',
'wolf',
'sneaking',
'amongst',
'the',
'flock',
'.',
'He',

'jumped',
'up',
'and',
'called',
'out',
'for',
'help',
'.',
'\n',
'But',
'no',
'one',
'came',
'this',
'time',
'because',
'they',
'thought',
'he',
'was',
'still',
'joking',
'around',
'.',
'At',
'sunset',
',',
'the',
'villagers',
'looked',
'for',
'the',
'boy',
'.',
'\n',
'He',
'had',
'not',
'returned',
'with',
'their',
'sheep',
'.',
'They',
'found',
'him',
'crying',

```
'.',  
'He',  
'told',  
'them',  
'that',  
'there',  
'really',  
'was',  
'a',  
'wolf',  
,',  
'and',  
'the',  
'entire',  
'flock',  
'was',  
'gone',  
'.',  
'\n',  
'An',  
'old',  
'man',  
'came',  
'to',  
'comfort',  
'him',  
'and',  
'told',  
'him',  
'that',  
'nobody',  
'would',  
'believe',  
'a',  
'liar',  
'even',  
'when',  
'they',  
'are',  
'being',  
'honest',  
'.']
```

```
[40]: len(tokens)
```

```
[40]: 206
```



```
[42]: punctuation      # also called as noisy characters
```

```
[42]: '!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

```
[43]: doc
```

```
[43]: There once was a boy who grew bored while watching over the village sheep. He
      wanted to make things more exciting.
      So, he yelled out that he saw a wolf chasing the sheep. All the villagers came
      running to drive the wolf away.
      However, they saw no wolf. The boy was amused, but the villagers were not. They
      told him not to do it again.
      Shortly after, he repeated this antic. The villagers came running again, only to
      find that he was lying.
      Later that day, the boy really sees a wolf sneaking amongst the flock. He jumped
      up and called out for help.
      But no one came this time because they thought he was still joking around. At
      sunset, the villagers looked for the boy.
      He had not returned with their sheep. They found him crying. He told them that
      there really was a wolf, and the entire flock was gone.
      An old man came to comfort him and told him that nobody would believe a liar
      even when they are being honest.
```

```
[44]: #we have to calculate the frequency of each and every word, how many time word
      ↪is repetition in text
```

```
word_frequencies = {}

for word in doc:
    if word.text.lower() not in stopwords:
        if word.text.lower() not in punctuation:
            if word.text not in word_frequencies.keys():
                word_frequencies[word.text] = 1
            else:
                word_frequencies[word.text] += 1
```

```
[45]: word_frequencies
```

```
[45]: {'boy': 4,
      'grew': 1,
      'bored': 1,
      'watching': 1,
      'village': 1,
      'sheep': 3,
      'wanted': 1,
      'things': 1,
      'exciting': 1,
```

```
'\n': 7,  
'yelled': 1,  
'saw': 2,  
'wolf': 5,  
'chasing': 1,  
'villagers': 4,  
'came': 4,  
'running': 2,  
'drive': 1,  
'away': 1,  
'amused': 1,  
'told': 3,  
'Shortly': 1,  
'repeated': 1,  
'antic': 1,  
'find': 1,  
'lying': 1,  
'Later': 1,  
'day': 1,  
'sees': 1,  
'sneaking': 1,  
'flock': 2,  
'jumped': 1,  
'called': 1,  
'help': 1,  
'time': 1,  
'thought': 1,  
'joking': 1,  
'sunset': 1,  
'looked': 1,  
'returned': 1,  
'found': 1,  
'crying': 1,  
'entire': 1,  
'gone': 1,  
'old': 1,  
'man': 1,  
'comfort': 1,  
'believe': 1,  
'liar': 1,  
'honest': 1}
```

```
[47]: len(word_frequencies)
```

```
[47]: 50
```

```
[49]: max_frequency = max(word_frequencies.values())
max_frequency
```

```
[49]: 7
```

```
[51]: #To get normalized/weighted frequencies you should divide all frequencies with 7
for word in word_frequencies.keys():
    word_frequencies[word] = word_frequencies[word]/max_frequency
```

```
[53]: word_frequencies
```

```
[53]: {'boy': 0.5714285714285714,
      'grew': 0.14285714285714285,
      'bored': 0.14285714285714285,
      'watching': 0.14285714285714285,
      'village': 0.14285714285714285,
      'sheep': 0.42857142857142855,
      'wanted': 0.14285714285714285,
      'things': 0.14285714285714285,
      'exciting': 0.14285714285714285,
      '\n': 1.0,
      'yelled': 0.14285714285714285,
      'saw': 0.2857142857142857,
      'wolf': 0.7142857142857143,
      'chasing': 0.14285714285714285,
      'villagers': 0.5714285714285714,
      'came': 0.5714285714285714,
      'running': 0.2857142857142857,
      'drive': 0.14285714285714285,
      'away': 0.14285714285714285,
      'amused': 0.14285714285714285,
      'told': 0.42857142857142855,
      'Shortly': 0.14285714285714285,
      'repeated': 0.14285714285714285,
      'antic': 0.14285714285714285,
      'find': 0.14285714285714285,
      'lying': 0.14285714285714285,
      'Later': 0.14285714285714285,
      'day': 0.14285714285714285,
      'sees': 0.14285714285714285,
      'sneaking': 0.14285714285714285,
      'flock': 0.2857142857142857,
      'jumped': 0.14285714285714285,
      'called': 0.14285714285714285,
      'help': 0.14285714285714285,
      'time': 0.14285714285714285,
      'thought': 0.14285714285714285,
```

```
'joking': 0.14285714285714285,  
'sunset': 0.14285714285714285,  
'looked': 0.14285714285714285,  
'returned': 0.14285714285714285,  
'found': 0.14285714285714285,  
'crying': 0.14285714285714285,  
'entire': 0.14285714285714285,  
'gone': 0.14285714285714285,  
'old': 0.14285714285714285,  
'man': 0.14285714285714285,  
'comfort': 0.14285714285714285,  
'believe': 0.14285714285714285,  
'liar': 0.14285714285714285,  
'honest': 0.14285714285714285}
```

```
[54]: sentence_tokens = [sent for sent in doc.sents]  
      sentence_tokens
```

```
[54]: [There once was a boy who grew bored while watching over the village sheep.,  
      He wanted to make things more exciting.,  
      So, he yelled out that he saw a wolf chasing the sheep.,  
      All the villagers came running to drive the wolf away. ,  
      However, they saw no wolf.,  
      The boy was amused, but the villagers were not.,  
      They told him not to do it again. ,  
      Shortly after, he repeated this antic.,  
      The villagers came running again, only to find that he was lying. ,  
      Later that day, the boy really sees a wolf sneaking amongst the flock.,  
      He jumped up and called out for help.,  
      But no one came this time because they thought he was still joking around.,  
      At sunset, the villagers looked for the boy.,  
      He had not returned with their sheep.,  
      They found him crying.,  
      He told them that there really was a wolf, and the entire flock was gone. ,  
      An old man came to comfort him and told him that nobody would believe a liar  
      even when they are being honest.]
```

```
[55]: len(sentence_tokens)
```

```
[55]: 17
```

```
[56]: sentence_tokens
```

```
[56]: [There once was a boy who grew bored while watching over the village sheep.,  
      He wanted to make things more exciting.,  
      So, he yelled out that he saw a wolf chasing the sheep.,  
      All the villagers came running to drive the wolf away. ,
```

However, they saw no wolf.,
 The boy was amused, but the villagers were not.,
 They told him not to do it again. ,
 Shortly after, he repeated this antic.,
 The villagers came running again, only to find that he was lying. ,
 Later that day, the boy really sees a wolf sneaking amongst the flock.,
 He jumped up and called out for help.,
 But no one came this time because they thought he was still joking around.,
 At sunset, the villagers looked for the boy.,
 He had not returned with their sheep.,
 They found him crying.,
 He told them that there really was a wolf, and the entire flock was gone. ,
 An old man came to comfort him and told him that nobody would believe a liar
 even when they are being honest.]

```

[58]: # we are going to calculate the sentence score, to calculate the sentence score
sentence_scores = {}

for sent in sentence_tokens:
    for word in sent:
        if word.text.lower() in word_frequencies.keys():
            if sent not in sentence_scores.keys():
                sentence_scores[sent] = word_frequencies[word.text.lower()]
            else:
                sentence_scores[sent] += word_frequencies[word.text.lower()]
  
```

```

[59]: sentence_scores
  
```

```

[59]: {There once was a boy who grew bored while watching over the village sheep.:
1.5714285714285712,
He wanted to make things more exciting.: 1.4285714285714286,
So, he yelled out that he saw a wolf chasing the sheep.: 1.7142857142857142,
All the villagers came running to drive the wolf away. : 3.4285714285714284,
However, they saw no wolf.: 1.0,
The boy was amused, but the villagers were not.: 1.2857142857142856,
They told him not to do it again. : 1.4285714285714286,
Shortly after, he repeated this antic.: 0.2857142857142857,
The villagers came running again, only to find that he was lying. :
2.714285714285714,
Later that day, the boy really sees a wolf sneaking amongst the flock.:
1.9999999999999996,
He jumped up and called out for help.: 1.4285714285714286,
But no one came this time because they thought he was still joking around.:
0.9999999999999998,
At sunset, the villagers looked for the boy.: 2.4285714285714284,
He had not returned with their sheep.: 0.5714285714285714,
They found him crying.: 0.2857142857142857,
  
```

He told them that there really was a wolf, and the entire flock was gone. :
2.714285714285714,

An old man came to comfort him and told him that nobody would believe a liar
even when they are being honest.: 1.857142857142857}

```
[60]: #lets say our case study was 30% sentence with maximum scores
      from heapq import nlargest
```

```
[63]: select_length = int(len(sentence_tokens)*0.3)
      select_length
```

```
[63]: 5
```

```
[64]: #we have to select maximum 3 sentences out of all sentences
      summary = nlargest(select_length,sentence_scores, key = sentence_scores.get)
```

```
[65]: summary
```

```
[65]: [All the villagers came running to drive the wolf away. ,
      The villagers came running again, only to find that he was lying. ,
      He told them that there really was a wolf, and the entire flock was gone. ,
      At sunset, the villagers looked for the boy.,
      Later that day, the boy really sees a wolf sneaking amongst the flock.]
```

```
[ ]:
```