```
from nltk.tokenize import sent tokenize, word tokenize
from urllib.request import urlopen, Request
from bs4 import BeautifulSoup as bs
import requests
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from nltk.corpus import stopwords
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from wordcloud import WordCloud
import matplotlib.pyplot as plt
english s = stopwords.words('english')
indonesia_s = stopwords.words('indonesian')
print(len(english_s), "stopwords:", english_s)
print(len(indonesia s), "stopwords:", indonesia s)
#179 stopwords: ['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves',...]
#758 stopwords: ['ada', 'adalah', 'adanya', 'adapun', 'agak', 'agaknya', 'agar', 'akan',...]
     179 stopwords: ['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you
     758 stopwords: ['ada', 'adalah', 'adanya', 'adapun', 'agak', 'agaknya', 'agar', 'akan',
user_agent='Mozilla/5.0 (Macintosh; Intel Mac OS X 10_12_1) AppleWebKit/602.2.14 (KHTML, like
headers = {'User-Agent': user agent,'Accept': 'text/html,application/xhtml+xml,application/xml
alamat = "https://www.kompas.com/tren/read/2020/10/15/163830165/hasil-cpns-2019-diumumkan-30-
req = Request(alamat, headers=headers)
html = urlopen(req)
data = bs(html, 'html.parser')
box = data.find("div",{"class":"read__content"})
hidelabel = box.findAll('strong')
for delete in hidelabel:
   delete.decompose()
items = box.findAll('p')
hasil = [item.get text() for item in items]
```

```
hasil = ['',join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)", "", item).split()
paragraf = ' '.join(hasil)
stop words = set(stopwords.words('indonesian'))
word_tokens = word_tokenize(paragraf)
paragraf_2 = ' '.join([w for w in word_tokens if not w in stop_words])
print (paragraf_2)
factory = StemmerFactory()
stemmer = factory.create_stemmer()
# stemming process
output
         = stemmer.stem(paragraf_2)
print(output)
#kembang kuliner di indonesia cukup bangga dalam 10 tahun akhir
wordcloud = WordCloud(width=1600, height=800, max font size=200).generate(output)
plt.figure(figsize=(12,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



link berita: https://www.kompas.com/tren/read/2020/10/15/163830165/hasil-cpns-2019-diumumkan-30-oktober-2020-ini-jadwal-lengkapnya

Hasil wordcloud mengeluarkan kata2 penting (kata dasar hasil stemming), dan beberapa kata penting dengan ukuran yang lebih besar diantaranya hasil, oktober, penetapan, seleksi, cpns, bkn,

nip. Semua istilah-istilah ini berkaitan dengan PNS dan seleksi PNS

- Hasil = hasil tes CPNS
- Oktober = bulan dimumumkannya hasil tes CPNS
- Seleksi = proses seleksi CPNS
- Penetapan = Penetapan hasil akhir tes CPNS
- CPNS = Calon Pegawai Negeri Sipil
- BKN = Badan Kepegawaian Negara, yang mengadakan seleksi CPNS
- NIP = Nomor Induk PNS