

ICE-4 Text Data: Flattening, Filtering, and Chunking

Alexis Blackwell

Ice 4

(Tutorial) Bag of X

Following is a sample of applying bag of n-grams to Yelp academic dataset review, please download it with following link:

https://github.com/knowitall/yelp-dataset-challenge/blob/master/data/yelp_phoenix_academic_dataset/yelp_academic_dataset_review.json

```
import pandas as pd
import json

f = open(r'C:\Users\black\CSCE5222_Feature_Engineering\
Feature_Engineering_ICE_4\yelp_academic_dataset_review.json')
js = []
for i in range(10000):
    js.append(json.loads(f.readline()))
f.close()
review_df = pd.DataFrame(js)
review_df.shape

(10000, 8)

review_df.head()
```

		votes	user_id	\
0	{'funny': 0, 'useful': 5, 'cool': 2}	rLtl8ZkDX5vH5nAx9C3q5Q		
1	{'funny': 0, 'useful': 0, 'cool': 0}	0a2KyEL0d3Yb1V6aivbIuQ		
2	{'funny': 0, 'useful': 1, 'cool': 0}	0hT2KtFLiobPvh6cDC8JQg		
3	{'funny': 0, 'useful': 2, 'cool': 1}	uZetl9T0NcR0G0yFfughhg		
4	{'funny': 0, 'useful': 0, 'cool': 0}	vYmM4KTsC8ZfQBg-j5MWkw		

	review_id	stars	date	\
0	fWKvX83p0-ka4JS3dc6E5A	5	2011-01-26	
1	IjZ33sJrzXqU-0X6U8NwyA	5	2011-07-27	
2	IESLBzqUCLdSzSqm0eCSxQ	4	2012-06-14	
3	G-WvGaISbqqamHLNnByodA	5	2010-05-27	
4	1uJFq2r5QfJG_6ExMRCaGw	5	2012-01-05	

	text	type	\
0	My wife took me here on my birthday for breakf...	review	
1	I have no idea why some people give bad review...	review	
2	love the gyro plate. Rice is so good and I als...	review	
3	Rosie, Dakota, and I LOVE Chaparral Dog Park!!!...	review	
4	General Manager Scott Petello is a good egg!!!...	review	

```

        business_id
0  9yKzy9PApeiPP0UJEtnvkg
1  ZRJwVLyzEJq1VAihDhYiow
2  6oRAC4uyJCsJl1X0WZpVSA
3  _1QQZuf4zZ0yFCvXc0o6Vg
4  6ozycU1RpktNG2-1BroVtw

```

note: in the default settings of CountVectorizer, the token_pattern = '(?u)\b\w\w+\b', which ignores single-character words. We employ the token_pattern = '(?u)\b\w+\b' to include the single-character words.

```

from sklearn.feature_extraction.text import CountVectorizer
bow_converter = CountVectorizer(token_pattern='(?u)\b\w+\b')
x = bow_converter.fit_transform(review_df['text'])

unigram = bow_converter.get_feature_names_out()

bigram_converter = CountVectorizer(ngram_range=(2,2),
token_pattern='(?u)\b\w+\b')
x2 = bigram_converter.fit_transform(review_df['text'])

bigram = bigram_converter.get_feature_names_out()

trigram_converter = CountVectorizer(ngram_range=(3,3),
token_pattern='(?u)\b\w+\b')
x3 = trigram_converter.fit_transform(review_df['text'])

trigram = trigram_converter.get_feature_names_out()

unigram
array(['0', '00', '000', ..., 'école', 'ém', 'òc'], dtype=object)

bigram
array(['0 0', '0 20', '0 39', ..., 'école lenôtre', 'ém all', 'òc
châm'],
      dtype=object)

trigram
array(['0 0 eye', '0 20 less', '0 39 oz', ..., 'école lenôtre
trained',
      'ém all they', 'òc châm a'], dtype=object)

print (len(unigram), len(bigram), len(trigram))

29222 368943 881620

%matplotlib notebook
import matplotlib.pyplot as plt
import seaborn as sns

```

```

sns.set_style("darkgrid")
counts = [len(unigram), len(bigram), len(trigram)]
plt.plot(counts, color='cornflowerblue')
plt.plot(counts, 'bo')
plt.margins(0.1)
plt.xticks(range(3), ['unigram', 'bigram', 'trigram'])
plt.tick_params(labelsize=14)
plt.title('Number of ngrams in the first 10,000 reviews of the Yelp
dataset', {'fontsize':16})
plt.show()

```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

Task 1. 1 Applying the unigram, bigram, and trigram tokenization methods to the given text below.

```

train_text = """My wife took me here on my birthday for breakfast and
it was excellent. The weather was perfect which made sitting outside
overlooking their grounds an absolute pleasure.
Our waitress was excellent and our food arrived quickly on the semi-
busy Saturday morning. It looked like the place fills up pretty
quickly so the earlier you get here the better.\n\nDo yourself a favor
and get their Bloody Mary. It was phenomenal and simply the best
I\ve ever had.
I\m pretty sure they only use ingredients from their garden and blend
them fresh when you order it. It was amazing.\n\nWhile EVERYTHING on
the menu looks excellent, I had the white truffle scrambled eggs
vegetable skillet and it was tasty and delicious.
It came with 2 pieces of their griddled bread with was amazing and it
absolutely made the meal complete.
It was the best "toast" I\ve ever had.\n\nAnyway, I can\t wait to go
back!"""

```

write your code here

```
df_engram = pd.DataFrame([train_text], columns=['Text'])
```

```
display(df_engram) # check
```

```

                                Text
0  My wife took me here on my birthday for breakf...

```

```

bow_converter = CountVectorizer(token_pattern='(?u)\\b\\w+\\b')
x = bow_converter.fit_transform(df_engram['Text'])

```

```
unigram = bow_converter.get_feature_names_out() # get unigram
```

```

bigram_converter = CountVectorizer(ngram_range=(2,2),
token_pattern='(?u)\\b\\w+\\b')
x2 = bigram_converter.fit_transform(df_engram['Text'])

```

```

bigram = bigram_converter.get_feature_names_out() # get bigram

trigram_converter = CountVectorizer(ngram_range=(3,3),
token_pattern='(?u)\\b\\w+\\b')
x3 = trigram_converter.fit_transform(df_engram['Text'])

trigram = trigram_converter.get_feature_names_out() # get trigram
unigram # check

array(['2', 'a', 'absolute', 'absolutely', 'amazing', 'an', 'and',
      'anyway', 'arrived', 'back', 'best', 'better', 'birthday',
      'blend',
      'bloody', 'bread', 'breakfast', 'busy', 'came', 'can',
      'complete',
      'delicious', 'do', 'earlier', 'eggs', 'ever', 'everything',
      'excellent', 'favor', 'fills', 'food', 'for', 'fresh', 'from',
      'garden', 'get', 'go', 'griddled', 'grounds', 'had', 'here',
      'i',
      'ingredients', 'it', 'like', 'looked', 'looks', 'm', 'made',
      'mary', 'me', 'meal', 'menu', 'morning', 'my', 'of', 'on',
      'only',
      'order', 'our', 'outside', 'overlooking', 'perfect',
      'phenomenal',
      'pieces', 'place', 'pleasure', 'pretty', 'quickly', 'saturday',
      'scrambled', 'semi', 'simply', 'sitting', 'skillet', 'so',
      'sure',
      't', 'tasty', 'the', 'their', 'them', 'they', 'to', 'toast',
      'took', 'truffle', 'up', 'use', 've', 'vegetable', 'wait',
      'waitress', 'was', 'weather', 'when', 'which', 'while',
      'white',
      'wife', 'with', 'you', 'yourself'], dtype=object)

bigram # check

array(['2 pieces', 'a favor', 'absolute pleasure', 'absolutely made',
      'amazing and', 'amazing while', 'an absolute', 'and blend',
      'and delicious', 'and get', 'and it', 'and our', 'and simply',
      'anyway i', 'arrived quickly', 'best i', 'best toast', 'better
do',
      'birthday for', 'blend them', 'bloody mary', 'bread with',
      'breakfast and', 'busy saturday', 'came with', 'can t',
      'complete it', 'delicious it', 'do yourself', 'earlier you',
      'eggs vegetable', 'ever had', 'everything on', 'excellent and',
      'excellent i', 'excellent the', 'favor and', 'fills up',
      'food arrived', 'for breakfast', 'fresh when', 'from their',
      'garden and', 'get here', 'get their', 'go back', 'griddled
bread',
      'grounds an', 'had anyway', 'had i', 'had the', 'here on',
      'here the', 'i can', 'i had', 'i m', 'i ve', 'ingredients

```

```

from',
    'it absolutely', 'it came', 'it it', 'it looked', 'it was',
    'like the', 'looked like', 'looks excellent', 'm pretty',
    'made sitting', 'made the', 'mary it', 'me here', 'meal
complete',
    'menu looks', 'morning it', 'my birthday', 'my wife', 'of
their',
    'on my', 'on the', 'only use', 'order it', 'our food',
    'our waitress', 'outside overlooking', 'overlooking their',
    'perfect which', 'phenomenal and', 'pieces of', 'place fills',
    'pleasure our', 'pretty quickly', 'pretty sure', 'quickly on',
    'quickly so', 'saturday morning', 'scrambled eggs', 'semi
busy',
    'simply the', 'sitting outside', 'skillet and', 'so the',
    'sure they', 't wait', 'tasty and', 'the best', 'the better',
    'the earlier', 'the meal', 'the menu', 'the place', 'the semi',
    'the weather', 'the white', 'their bloody', 'their garden',
    'their griddled', 'their grounds', 'them fresh', 'they only',
    'to go', 'toast i', 'took me', 'truffle scrambled', 'up
pretty',
    'use ingredients', 've ever', 'vegetable skillet', 'wait to',
    'waitress was', 'was amazing', 'was excellent', 'was perfect',
    'was phenomenal', 'was tasty', 'was the', 'weather was',
    'when you', 'which made', 'while everything', 'white truffle',
    'wife took', 'with 2', 'with was', 'you get', 'you order',
    'yourself a'], dtype=object)

```

trigram # *check*

```

array(['2 pieces of', 'a favor and', 'absolute pleasure our',
    'absolutely made the', 'amazing and it',
    'amazing while everything', 'an absolute pleasure',
    'and blend them', 'and delicious it', 'and get their',
    'and it absolutely', 'and it was', 'and our food',
    'and simply the', 'anyway i can', 'arrived quickly on',
    'best i ve', 'best toast i', 'better do yourself',
    'birthday for breakfast', 'blend them fresh', 'bloody mary it',
    'bread with was', 'breakfast and it', 'busy saturday morning',
    'came with 2', 'can t wait', 'complete it was',
    'delicious it came', 'do yourself a', 'earlier you get',
    'eggs vegetable skillet', 'ever had anyway', 'ever had i',
    'everything on the', 'excellent and our', 'excellent i had',
    'excellent the weather', 'favor and get', 'fills up pretty',
    'food arrived quickly', 'for breakfast and', 'fresh when you',
    'from their garden', 'garden and blend', 'get here the',
    'get their bloody', 'griddled bread with', 'grounds an
absolute',
    'had anyway i', 'had i m', 'had the white', 'here on my',
    'here the better', 'i can t', 'i had the', 'i m pretty',
    'i ve ever', 'ingredients from their', 'it absolutely made',
    'it came with', 'it it was', 'it looked like', 'it was

```

```

amazing',
    'it was excellent', 'it was phenomenal', 'it was tasty',
    'it was the', 'like the place', 'looked like the',
    'looks excellent i', 'm pretty sure', 'made sitting outside',
    'made the meal', 'mary it was', 'me here on', 'meal complete
it',
    'menu looks excellent', 'morning it looked', 'my birthday for',
    'my wife took', 'of their griddled', 'on my birthday',
    'on the menu', 'on the semi', 'only use ingredients',
    'order it it', 'our food arrived', 'our waitress was',
    'outside overlooking their', 'overlooking their grounds',
    'perfect which made', 'phenomenal and simply', 'pieces of
their',
    'place fills up', 'pleasure our waitress', 'pretty quickly so',
    'pretty sure they', 'quickly on the', 'quickly so the',
    'saturday morning it', 'scrambled eggs vegetable',
    'semi busy saturday', 'simply the best',
    'sitting outside overlooking', 'skillet and it', 'so the
earlier',
    'sure they only', 't wait to', 'tasty and delicious', 'the best
i',
    'the best toast', 'the better do', 'the earlier you',
    'the meal complete', 'the menu looks', 'the place fills',
    'the semi busy', 'the weather was', 'the white truffle',
    'their bloody mary', 'their garden and', 'their griddled
bread',
    'their grounds an', 'them fresh when', 'they only use',
    'to go back', 'toast i ve', 'took me here',
    'truffle scrambled eggs', 'up pretty quickly',
    'use ingredients from', 've ever had', 'vegetable skillet and',
    'wait to go', 'waitress was excellent', 'was amazing and',
    'was amazing while', 'was excellent and', 'was excellent the',
    'was perfect which', 'was phenomenal and', 'was tasty and',
    'was the best', 'weather was perfect', 'when you order',
    'which made sitting', 'while everything on',
    'white truffle scrambled', 'wife took me', 'with 2 pieces',
    'with was amazing', 'you get here', 'you order it',
    'yourself a favor'], dtype=object)

```

Overall, this step shows the unigrams, bigrams, and trigrams that are present in this piece of text. Each n-gram captures some text context that could be used by a model that is not captured when only considering each word as a single element. For example the bigram vegetable skillet could be useful to capture the context that the meal was vegetarian which could be useful context.

Task 1.2 Create your own naive tokenization method (whitespace-based), and apply it to the text given in the task 1.1

note: 1. do not use the existing tokenization methods given by NLP; 2. split the words by whitespace character, the output is more likely as the unigram; 3. no repeating elements in the output.

```
print(train_text) # check
```

My wife took me here on my birthday for breakfast and it was excellent. The weather was perfect which made sitting outside overlooking their grounds an absolute pleasure. Our waitress was excellent and our food arrived quickly on the semi-busy Saturday morning. It looked like the place fills up pretty quickly so the earlier you get here the better.

Do yourself a favor and get their Bloody Mary. It was phenomenal and simply the best I've ever had. I'm pretty sure they only use ingredients from their garden and blend them fresh when you order it. It was amazing.

While EVERYTHING on the menu looks excellent, I had the white truffle scrambled eggs vegetable skillet and it was tasty and delicious. It came with 2 pieces of their griddled bread with was amazing and it absolutely made the meal complete. It was the best "toast" I've ever had.

Anyway, I can't wait to go back!

```
# write you code here
# using same df_engram text string
# token_pattern='(?u)\b\w+\b' original
token_pattern = '(?u)\w+' # chosen token pattern
bow_converter = CountVectorizer(token_pattern=token_pattern)
x = bow_converter.fit_transform(df_engram['Text'])

unigram = bow_converter.get_feature_names_out() # get unigram

bigram_converter = CountVectorizer(ngram_range=(2,2),
token_pattern=token_pattern)
x2 = bigram_converter.fit_transform(df_engram['Text'])

bigram = bigram_converter.get_feature_names_out() # get bigram

trigram_converter = CountVectorizer(ngram_range=(3,3),
token_pattern=token_pattern)
x3 = trigram_converter.fit_transform(df_engram['Text'])

trigram = trigram_converter.get_feature_names_out() # get trigram
```

unigram # *check*

```
array(['2', 'a', 'absolute', 'absolutely', 'amazing', 'an', 'and',  
      'anyway', 'arrived', 'back', 'best', 'better', 'birthday',  
      'blend',  
      'bloody', 'bread', 'breakfast', 'busy', 'came', 'can',  
      'complete',  
      'delicious', 'do', 'earlier', 'eggs', 'ever', 'everything',  
      'excellent', 'favor', 'fills', 'food', 'for', 'fresh', 'from',  
      'garden', 'get', 'go', 'griddled', 'grounds', 'had', 'here',  
      'i',  
      'ingredients', 'it', 'like', 'looked', 'looks', 'm', 'made',  
      'mary', 'me', 'meal', 'menu', 'morning', 'my', 'of', 'on',  
      'only',  
      'order', 'our', 'outside', 'overlooking', 'perfect',  
      'phenomenal',  
      'pieces', 'place', 'pleasure', 'pretty', 'quickly', 'saturday',  
      'scrambled', 'semi', 'simply', 'sitting', 'skillet', 'so',  
      'sure',  
      't', 'tasty', 'the', 'their', 'them', 'they', 'to', 'toast',  
      'took', 'truffle', 'up', 'use', 've', 'vegetable', 'wait',  
      'waitress', 'was', 'weather', 'when', 'which', 'while',  
      'white',  
      'wife', 'with', 'you', 'yourself'], dtype=object)
```

bigram # *check*

```
array(['2 pieces', 'a favor', 'absolute pleasure', 'absolutely made',  
      'amazing and', 'amazing while', 'an absolute', 'and blend',  
      'and delicious', 'and get', 'and it', 'and our', 'and simply',  
      'anyway i', 'arrived quickly', 'best i', 'best toast', 'better  
do',  
      'birthday for', 'blend them', 'bloody mary', 'bread with',  
      'breakfast and', 'busy saturday', 'came with', 'can t',  
      'complete it', 'delicious it', 'do yourself', 'earlier you',  
      'eggs vegetable', 'ever had', 'everything on', 'excellent and',  
      'excellent i', 'excellent the', 'favor and', 'fills up',  
      'food arrived', 'for breakfast', 'fresh when', 'from their',  
      'garden and', 'get here', 'get their', 'go back', 'griddled  
bread',  
      'grounds an', 'had anyway', 'had i', 'had the', 'here on',  
      'here the', 'i can', 'i had', 'i m', 'i ve', 'ingredients  
from',  
      'it absolutely', 'it came', 'it it', 'it looked', 'it was',  
      'like the', 'looked like', 'looks excellent', 'm pretty',  
      'made sitting', 'made the', 'mary it', 'me here', 'meal  
complete',  
      'menu looks', 'morning it', 'my birthday', 'my wife', 'of  
their',  
      'on my', 'on the', 'only use', 'order it', 'our food',  
      'our waitress', 'outside overlooking', 'overlooking their',
```



```

    'perfect which', 'phenomenal and', 'pieces of', 'place fills',
    'pleasure our', 'pretty quickly', 'pretty sure', 'quickly on',
    'quickly so', 'saturday morning', 'scrambled eggs', 'semi
busy',
    'simply the', 'sitting outside', 'skillet and', 'so the',
    'sure they', 't wait', 'tasty and', 'the best', 'the better',
    'the earlier', 'the meal', 'the menu', 'the place', 'the semi',
    'the weather', 'the white', 'their bloody', 'their garden',
    'their griddled', 'their grounds', 'them fresh', 'they only',
    'to go', 'toast i', 'took me', 'truffle scrambled', 'up
pretty',
    'use ingredients', 've ever', 'vegetable skillet', 'wait to',
    'waitress was', 'was amazing', 'was excellent', 'was perfect',
    'was phenomenal', 'was tasty', 'was the', 'weather was',
    'when you', 'which made', 'while everything', 'white truffle',
    'wife took', 'with 2', 'with was', 'you get', 'you order',
    'yourself a'], dtype=object)

```

trigram # *check*

```

array(['2 pieces of', 'a favor and', 'absolute pleasure our',
    'absolutely made the', 'amazing and it',
    'amazing while everything', 'an absolute pleasure',
    'and blend them', 'and delicious it', 'and get their',
    'and it absolutely', 'and it was', 'and our food',
    'and simply the', 'anyway i can', 'arrived quickly on',
    'best i ve', 'best toast i', 'better do yourself',
    'birthday for breakfast', 'blend them fresh', 'bloody mary it',
    'bread with was', 'breakfast and it', 'busy saturday morning',
    'came with 2', 'can t wait', 'complete it was',
    'delicious it came', 'do yourself a', 'earlier you get',
    'eggs vegetable skillet', 'ever had anyway', 'ever had i',
    'everything on the', 'excellent and our', 'excellent i had',
    'excellent the weather', 'favor and get', 'fills up pretty',
    'food arrived quickly', 'for breakfast and', 'fresh when you',
    'from their garden', 'garden and blend', 'get here the',
    'get their bloody', 'griddled bread with', 'grounds an
absolute',
    'had anyway i', 'had i m', 'had the white', 'here on my',
    'here the better', 'i can t', 'i had the', 'i m pretty',
    'i ve ever', 'ingredients from their', 'it absolutely made',
    'it came with', 'it it was', 'it looked like', 'it was
amazing',
    'it was excellent', 'it was phenomenal', 'it was tasty',
    'it was the', 'like the place', 'looked like the',
    'looks excellent i', 'm pretty sure', 'made sitting outside',
    'made the meal', 'mary it was', 'me here on', 'meal complete
it',
    'menu looks excellent', 'morning it looked', 'my birthday for',
    'my wife took', 'of their griddled', 'on my birthday',
    'on the menu', 'on the semi', 'only use ingredients',

```

```

'order it it', 'our food arrived', 'our waitress was',
'outside overlooking their', 'overlooking their grounds',
'their', 'perfect which made', 'phenomenal and simply', 'pieces of
their', 'place fills up', 'pleasure our waitress', 'pretty quickly so',
'pretty sure they', 'quickly on the', 'quickly so the',
'saturday morning it', 'scrambled eggs vegetable',
'semi busy saturday', 'simply the best',
'sitting outside overlooking', 'skillet and it', 'so the
earlier', 'sure they only', 't wait to', 'tasty and delicious', 'the best
i', 'the best toast', 'the better do', 'the earlier you',
'the meal complete', 'the menu looks', 'the place fills',
'the semi busy', 'the weather was', 'the white truffle',
'their bloody mary', 'their garden and', 'their griddled
bread', 'their grounds an', 'them fresh when', 'they only use',
'to go back', 'toast i ve', 'took me here',
'truffle scrambled eggs', 'up pretty quickly',
'use ingredients from', 've ever had', 'vegetable skillet and',
'wait to go', 'waitress was excellent', 'was amazing and',
'was amazing while', 'was excellent and', 'was excellent the',
'was perfect which', 'was phenomenal and', 'was tasty and',
'was the best', 'weather was perfect', 'when you order',
'which made sitting', 'while everything on',
'white truffle scrambled', 'wife took me', 'with 2 pieces',
'with was amazing', 'you get here', 'you order it',
'yourself a favor'], dtype=object)

```

This next step varies the pattern and produces a similar result and it seems as though the unigrams, bigrams, and trigrams are all the same.

Question 1. Given a sentence "He likes cat". In unigram representation, it could be "He", "likes", "cat". In bigram representation, it could be "He likes", "likes cat". In trigram representation, it could be "He likes cat". Explain why the storage and computation cost increase with the growth of n in n-gram methods.

Answer to Q1: type your answer here:

(Tutorial) Stemming and Lemmatization

```

# import PorterStemmer class from nltk.stem.porter module
from nltk.stem.porter import PorterStemmer
stemmer = PorterStemmer()

```

```

stem = stemmer.stem('flowers')
print(f"'flowers' after stemming: {stem}")

```

```

stem = stemmer.stem('zeroes')
print(f"'zeroes' after stemming: {stem}")

stem = stemmer.stem('better')
print(f"'better' after stemming: {stem}")

stem = stemmer.stem('sixties')
print(f"'sixties' after stemming: {stem}")

stem = stemmer.stem('goes')
print(f"'goes' after stemming: {stem}")

stem = stemmer.stem('go')
print(f"'go' after stemming: {stem}")

'flowers' after stemming: flower
'zeroes' after stemming: zero
'better' after stemming: better
'sixties' after stemming: sixti
'goes' after stemming: goe
'go' after stemming: go

# import lemmatizer class from nltk.stem module
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()

lemma = lemmatizer.lemmatize('flowers')
print(f"'flowers' after lemmatization: {lemma}")

lemma = lemmatizer.lemmatize('zeros')
print(f"'zeros' after lemmatization: {lemma}")

lemma = lemmatizer.lemmatize('better')
print(f"'better' after lemmatization: {lemma}")

lemma = lemmatizer.lemmatize('sixties')
print(f"'sixties' after lemmatization: {lemma}")

lemma = lemmatizer.lemmatize('goes')
print(f"'goes' after lemmatization: {lemma}")

lemma = lemmatizer.lemmatize('go')
print(f"'go' after lemmatization: {lemma}")

print("\n\n")
lemma = lemmatizer.lemmatize('better', pos='a')    # 'a' denoted
ADJECTIVE part-of-speech
print(f"'better' (as an adjective) after lemmatization: {lemma}")

```

'flowers' after lemmatization: flower
'zeros' after lemmatization: zero
'better' after lemmatization: better
'sixties' after lemmatization: sixty
'goes' after lemmatization: go
'go' after lemmatization: go

'better' (as an adjective) after lemmatization: good

Task 2. Text filtering for cleaner feature

1. clean the text used in the task 1; 2. remove all punctuations; 3. convert all characters to their lowercase; 4. remove all words in "stopwords"; 5. remove all relatively meaningless words like " 've ", " 's ", etc. 6. after finishing the above operations, apply stemming and lemmatization to the cleaned text respectively.

```
# write your code here
# original text
print(train_text) # check
new_text = train_text # copy text
```

My wife took me here on my birthday for breakfast and it was excellent. The weather was perfect which made sitting outside overlooking their grounds an absolute pleasure. Our waitress was excellent and our food arrived quickly on the semi-busy Saturday morning. It looked like the place fills up pretty quickly so the earlier you get here the better.

Do yourself a favor and get their Bloody Mary. It was phenomenal and simply the best I've ever had. I'm pretty sure they only use ingredients from their garden and blend them fresh when you order it. It was amazing.

While EVERYTHING on the menu looks excellent, I had the white truffle scrambled eggs vegetable skillet and it was tasty and delicious. It came with 2 pieces of their griddled bread with was amazing and it absolutely made the meal complete. It was the best "toast" I've ever had.

Anyway, I can't wait to go back!

```
# remove punctuation
new_text = new_text.replace("\r", " ") # remove \r
new_text = new_text.replace("\n", " ") # remove \n
new_text = new_text.replace(" ", " ") # remove space
new_text = new_text.replace(".", "") # remove .
new_text = new_text.replace("!", "") # remove !
new_text = new_text.replace(",", "") # remove ,
```

```
new_text = new_text.replace("'", "") # remove '  
new_text = new_text.replace("\\", "") # remove \  
  
print(new_text) # check
```

My wife took me here on my birthday for breakfast and it was excellent
The weather was perfect which made sitting outside overlooking their
grounds an absolute pleasure Our waitress was excellent and our food
arrived quickly on the semi-busy Saturday morning It looked like the
place fills up pretty quickly so the earlier you get here the better
Do yourself a favor and get their Bloody Mary It was phenomenal and
simply the best Ive ever had Im pretty sure they only use
ingredients from their garden and blend them fresh when you order it
It was amazing While EVERYTHING on the menu looks excellent I had the
white truffle scrambled eggs vegetable skillet and it was tasty and
delicious It came with 2 pieces of their griddled bread with was
amazing and it absolutely made the meal complete It was the best
toast Ive ever had Anyway I cant wait to go back

```
# convert to lowercase  
new_text = new_text.lower() # change to lowercase  
  
print(new_text) # check
```

my wife took me here on my birthday for breakfast and it was excellent
the weather was perfect which made sitting outside overlooking their
grounds an absolute pleasure our waitress was excellent and our food
arrived quickly on the semi-busy saturday morning it looked like the
place fills up pretty quickly so the earlier you get here the better
do yourself a favor and get their bloody mary it was phenomenal and
simply the best ive ever had im pretty sure they only use
ingredients from their garden and blend them fresh when you order it
it was amazing while everything on the menu looks excellent i had the
white truffle scrambled eggs vegetable skillet and it was tasty and
delicious it came with 2 pieces of their griddled bread with was
amazing and it absolutely made the meal complete it was the best
toast ive ever had anyway i cant wait to go back

```
# remove meaningless words  
new_text = new_text.replace("'s'", "") # remove 's  
new_text = new_text.replace("'ve", "") # remove ve  
new_text = new_text.replace("2", "two") # change the 2 to written two  
  
print(new_text) # check
```

my wife took me here on my birthday for breakfast and it was excellent
the weather was perfect which made sitting outside overlooking their
grounds an absolute pleasure our waitress was excellent and our food
arrived quickly on the semi-busy saturday morning it looked like the
place fills up pretty quickly so the earlier you get here the better
do yourself a favor and get their bloody mary it was phenomenal and
simply the best ive ever had im pretty sure they only use

ingredients from their garden and blend them fresh when you order it
it was amazing while everything on the menu looks excellent i had the
white truffle scrambled eggs vegetable skillet and it was tasty and
delicious it came with two pieces of their griddled bread with was
amazing and it absolutely made the meal complete it was the best
toast ive ever had anyway i cant wait to go back

```
import nltk
nltk.download('punkt')
nltk.download('wordnet')
nltk.download('stopwords')
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data] C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] Package wordnet is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

# remove stopwords
# Loading the stop words in english
stop_text = new_text
stop_words = list(stopwords.words('english'))
for i in stop_words: # iterate through stop words
    regex_stopword = r"\b" + i + r"\b"
    stop_text = stop_text.replace(regex_stopword, '')

print(stop_text) # check
```

my wife took me here on my birthday for breakfast and it was excellent
the weather was perfect which made sitting outside overlooking their
grounds an absolute pleasure our waitress was excellent and our food
arrived quickly on the semi-busy saturday morning it looked like the
place fills up pretty quickly so the earlier you get here the better
do yourself a favor and get their bloody mary it was phenomenal and
simply the best ive ever had im pretty sure they only use
ingredients from their garden and blend them fresh when you order it
it was amazing while everything on the menu looks excellent i had the
white truffle scrambled eggs vegetable skillet and it was tasty and
delicious it came with two pieces of their griddled bread with was
amazing and it absolutely made the meal complete it was the best
toast ive ever had anyway i cant wait to go back

```
# apply stemming and lemmatization
# stemming
from nltk.stem.porter import PorterStemmer
```

```

stemmer = PorterStemmer()

stem_text = ""
temp_text = stop_text.split()
for i in temp_text:
    stem = stemmer.stem(i) # get stem word
    stem_text = stem_text + " " + (stem)

print(stem_text) # check

```

my wife took me here on my birthday for breakfast and it wa excel the weather wa perfect which made sit outsid overlook their ground an absolut pleasur our waitress wa excel and our food arriv quickli on the semi-busi saturday morn it look like the place fill up pretti quickli so the earlier you get here the better do yourself a favor and get their bloodi mari it wa phenomen and simpli the best ive ever had im pretti sure they onli use ingredi from their garden and blend them fresh when you order it it wa amaz while everyth on the menu look excel i had the white truffl scrambl egg veget skillet and it wa tasti and delici it came with two piec of their griddl bread with wa amaz and it absolut made the meal complet it wa the best toast ive ever had anyway i cant wait to go back

```

# import lemmatizer class from nltk.stem module
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()
lem_text = " "
temp_text = stem_text.split()
for i in stem_text:
    lemma = lemmatizer.lemmatize(i) # get lemmatized word
    lem_text = lem_text + (lemma)

print(lem_text) # check

```

my wife took me here on my birthday for breakfast and it wa excel the weather wa perfect which made sit outsid overlook their ground an absolut pleasur our waitress wa excel and our food arriv quickli on the semi-busi saturday morn it look like the place fill up pretti quickli so the earlier you get here the better do yourself a favor and get their bloodi mari it wa phenomen and simpli the best ive ever had im pretti sure they onli use ingredi from their garden and blend them fresh when you order it it wa amaz while everyth on the menu look excel i had the white truffl scrambl egg veget skillet and it wa tasti and delici it came with two piec of their griddl bread with wa amaz and it absolut made the meal complet it wa the best toast ive ever had anyway i cant wait to go back

This task implements text cleaning processes which is similar to previous ICE but with more of a focus on stemming and lemmatization. Stemming being the action of trying to reduce a word to its base form such as walking is transformed to walk. Lemmatization is

reducing a word to its lemma or root word such as became is be. This reduces the number of words that the model needs to consider as it reduces the complexity of the input text.

Question 2. Based on the examples and the output of your code, which one has the better performance, Stemming or Lemmatization? Try to analyze it.

Answer to Q2: type your answer here

(Tutorial) PoS tagging and chunking

note: you need to install spacy and textblob modules first for the following codes If you have problem to install spacy module, try to follow the instruction in the following link: <https://stackoverflow.com/questions/66149878/e053-could-not-read-config-cfg-resumeparser> If you have problem to use textblob module, try to install nltk libraries as shown in the following link: <http://www.nltk.org/data.html>

```
%pip install nltk
```

```
%pip install spacy==2.3.5
```

```
%pip install https://github.com/explosion/spacy-  
models/releases/download/en_core_web_sm-2.3.1/en_core_web_sm-  
2.3.1.tar.gz
```

```
%pip install pyresparser
```

```
Requirement already satisfied: nltk in c:\users\black\anaconda3\lib\  
site-packages (3.7)  
Requirement already satisfied: click in c:\users\black\anaconda3\lib\  
site-packages (from nltk) (8.0.4)  
Requirement already satisfied: regex<=2021.8.3 in c:\users\black\  
anaconda3\lib\site-packages (from nltk) (2022.7.9)  
Requirement already satisfied: tqdm in c:\users\black\anaconda3\lib\  
site-packages (from nltk) (4.64.1)  
Requirement already satisfied: joblib in c:\users\black\anaconda3\lib\  
site-packages (from nltk) (1.1.1)  
Requirement already satisfied: colorama in c:\users\black\anaconda3\  
lib\site-packages (from click->nltk) (0.4.6)  
Note: you may need to restart the kernel to use updated packages.  
Collecting spacy==2.3.5  
  Using cached spacy-2.3.5.tar.gz (5.8 MB)  
  Installing build dependencies: started  
  Installing build dependencies: finished with status 'done'  
  Getting requirements to build wheel: started  
  Getting requirements to build wheel: finished with status 'done'  
  Installing backend dependencies: started  
  Installing backend dependencies: finished with status 'done'  
  Preparing metadata (pyproject.toml): started
```



```
Preparing metadata (pyproject.toml): finished with status 'done'
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\
black\anaconda3\lib\site-packages (from spacy==2.3.5) (3.0.8)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\
black\anaconda3\lib\site-packages (from spacy==2.3.5) (1.0.9)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\black\
anaconda3\lib\site-packages (from spacy==2.3.5) (2.0.7)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in c:\users\
black\anaconda3\lib\site-packages (from spacy==2.3.5) (1.0.2)
Requirement already satisfied: setuptools in c:\users\black\anaconda3\
lib\site-packages (from spacy==2.3.5) (65.6.3)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in c:\users\black\
anaconda3\lib\site-packages (from spacy==2.3.5) (1.1.3)
Requirement already satisfied: thinc<7.5.0,>=7.4.1 in c:\users\black\
anaconda3\lib\site-packages (from spacy==2.3.5) (7.4.6)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\
black\anaconda3\lib\site-packages (from spacy==2.3.5) (2.28.1)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in c:\users\black\
anaconda3\lib\site-packages (from spacy==2.3.5) (0.10.1)
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in c:\users\black\
anaconda3\lib\site-packages (from spacy==2.3.5) (1.0.6)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\black\
anaconda3\lib\site-packages (from spacy==2.3.5) (4.64.1)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in c:\users\black\
anaconda3\lib\site-packages (from spacy==2.3.5) (0.7.9)
Requirement already satisfied: numpy>=1.15.0 in c:\users\black\
anaconda3\lib\site-packages (from spacy==2.3.5) (1.23.5)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\
black\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0-
>spacy==2.3.5) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\black\
anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0-
>spacy==2.3.5) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\
black\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0-
>spacy==2.3.5) (1.26.14)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\black\
anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0-
>spacy==2.3.5) (2023.5.7)
Requirement already satisfied: colorama in c:\users\black\anaconda3\
lib\site-packages (from tqdm<5.0.0,>=4.38.0->spacy==2.3.5) (0.4.6)
Building wheels for collected packages: spacy
  Building wheel for spacy (pyproject.toml): started
  Building wheel for spacy (pyproject.toml): finished with status
'error'
Failed to build spacy
Note: you may need to restart the kernel to use updated packages.
```

```
error: subprocess-exited-with-error
```

Building wheel for spacy (pyproject.toml) did not run successfully.
exit code: 1

```
[1118 lines of output]
running bdist_wheel
running build
running build_py
creating build
creating build\lib.win-amd64-cpython-310
creating build\lib.win-amd64-cpython-310\bin
copying bin\cythonize.py -> build\lib.win-amd64-cpython-310\bin
copying bin\load_reddit.py -> build\lib.win-amd64-cpython-310\bin
copying bin\train_word_vectors.py -> build\lib.win-amd64-cpython-
310\bin
copying bin\__init__.py -> build\lib.win-amd64-cpython-310\bin
creating build\lib.win-amd64-cpython-310\spacy
copying spacy\about.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\analysis.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\compat.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\errors.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\git_info.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\glossary.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\language.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\lemmatizer.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\lookups.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\scorer.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\util.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\_ml.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\__init__.py -> build\lib.win-amd64-cpython-310\spacy
copying spacy\__main__.py -> build\lib.win-amd64-cpython-310\spacy
creating build\lib.win-amd64-cpython-310\bin\ud
copying bin\ud\conll17_ud_eval.py -> build\lib.win-amd64-cpython-
310\bin\ud
copying bin\ud\run_eval.py -> build\lib.win-amd64-cpython-310\bin\ud
copying bin\ud\ud_run_test.py -> build\lib.win-amd64-cpython-310\
bin\ud
copying bin\ud\ud_train.py -> build\lib.win-amd64-cpython-310\bin\ud
copying bin\ud\__init__.py -> build\lib.win-amd64-cpython-310\bin\ud
creating build\lib.win-amd64-cpython-310\spacy\cli
copying spacy\cli\convert.py -> build\lib.win-amd64-cpython-310\
spacy\cli
copying spacy\cli\debug_data.py -> build\lib.win-amd64-cpython-310\
spacy\cli
copying spacy\cli\download.py -> build\lib.win-amd64-cpython-310\
spacy\cli
copying spacy\cli\evaluate.py -> build\lib.win-amd64-cpython-310\
spacy\cli
copying spacy\cli\info.py -> build\lib.win-amd64-cpython-310\spacy\
cli
copying spacy\cli\init_model.py -> build\lib.win-amd64-cpython-310\
```

```
spacy\cli
  copying spacy\cli\link.py -> build\lib.win-amd64-cpython-310\spacy\cli
  copying spacy\cli\package.py -> build\lib.win-amd64-cpython-310\spacy\cli
  copying spacy\cli\pretrain.py -> build\lib.win-amd64-cpython-310\spacy\cli
  copying spacy\cli\profile.py -> build\lib.win-amd64-cpython-310\spacy\cli
  copying spacy\cli\train.py -> build\lib.win-amd64-cpython-310\spacy\cli
  copying spacy\cli\validate.py -> build\lib.win-amd64-cpython-310\spacy\cli
  copying spacy\cli\_schemas.py -> build\lib.win-amd64-cpython-310\spacy\cli
  copying spacy\cli\__init__.py -> build\lib.win-amd64-cpython-310\spacy\cli
  creating build\lib.win-amd64-cpython-310\spacy\data
  copying spacy\data\__init__.py -> build\lib.win-amd64-cpython-310\spacy\data
  creating build\lib.win-amd64-cpython-310\spacy\displacy
  copying spacy\displacy\render.py -> build\lib.win-amd64-cpython-310\spacy\displacy
  copying spacy\displacy\templates.py -> build\lib.win-amd64-cpython-310\spacy\displacy
  copying spacy\displacy\__init__.py -> build\lib.win-amd64-cpython-310\spacy\displacy
  creating build\lib.win-amd64-cpython-310\spacy\lang
  copying spacy\lang\char_classes.py -> build\lib.win-amd64-cpython-310\spacy\lang
  copying spacy\lang\lex_attrs.py -> build\lib.win-amd64-cpython-310\spacy\lang
  copying spacy\lang\norm_exceptions.py -> build\lib.win-amd64-cpython-310\spacy\lang
  copying spacy\lang\punctuation.py -> build\lib.win-amd64-cpython-310\spacy\lang
  copying spacy\lang\tag_map.py -> build\lib.win-amd64-cpython-310\spacy\lang
  copying spacy\lang\tokenizer_exceptions.py -> build\lib.win-amd64-cpython-310\spacy\lang
  copying spacy\lang\__init__.py -> build\lib.win-amd64-cpython-310\spacy\lang
  creating build\lib.win-amd64-cpython-310\spacy\matcher
  copying spacy\matcher\_schemas.py -> build\lib.win-amd64-cpython-310\spacy\matcher
  copying spacy\matcher\__init__.py -> build\lib.win-amd64-cpython-310\spacy\matcher
  creating build\lib.win-amd64-cpython-310\spacy\ml
  copying spacy\ml\common.py -> build\lib.win-amd64-cpython-310\spacy\ml
```

```
    copying spacy\ml\tok2vec.py -> build\lib.win-amd64-cpython-310\
spacy\ml
    copying spacy\ml\_legacy_tok2vec.py -> build\lib.win-amd64-cpython-
310\spacy\ml
    copying spacy\ml\_wire.py -> build\lib.win-amd64-cpython-310\spacy\
ml
    copying spacy\ml\__init__.py -> build\lib.win-amd64-cpython-310\
spacy\ml
    creating build\lib.win-amd64-cpython-310\spacy\pipeline
    copying spacy\pipeline\entityruler.py -> build\lib.win-amd64-
cpython-310\spacy\pipeline
    copying spacy\pipeline\functions.py -> build\lib.win-amd64-cpython-
310\spacy\pipeline
    copying spacy\pipeline\hooks.py -> build\lib.win-amd64-cpython-310\
spacy\pipeline
    copying spacy\pipeline\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\pipeline
    creating build\lib.win-amd64-cpython-310\spacy\syntax
    copying spacy\syntax\__init__.py -> build\lib.win-amd64-cpython-310\
spacy\syntax
    creating build\lib.win-amd64-cpython-310\spacy\tests
    copying spacy\tests\conftest.py -> build\lib.win-amd64-cpython-310\
spacy\tests
    copying spacy\tests\test_architectures.py -> build\lib.win-amd64-
cpython-310\spacy\tests
    copying spacy\tests\test_cli.py -> build\lib.win-amd64-cpython-310\
spacy\tests
    copying spacy\tests\test_displacy.py -> build\lib.win-amd64-cpython-
310\spacy\tests
    copying spacy\tests\test_errors.py -> build\lib.win-amd64-cpython-
310\spacy\tests
    copying spacy\tests\test_gold.py -> build\lib.win-amd64-cpython-310\
spacy\tests
    copying spacy\tests\test_json_schemas.py -> build\lib.win-amd64-
cpython-310\spacy\tests
    copying spacy\tests\test_language.py -> build\lib.win-amd64-cpython-
310\spacy\tests
    copying spacy\tests\test_lemmatizer.py -> build\lib.win-amd64-
cpython-310\spacy\tests
    copying spacy\tests\test_misc.py -> build\lib.win-amd64-cpython-310\
spacy\tests
    copying spacy\tests\test_pickles.py -> build\lib.win-amd64-cpython-
310\spacy\tests
    copying spacy\tests\test_scorer.py -> build\lib.win-amd64-cpython-
310\spacy\tests
    copying spacy\tests\test_tok2vec.py -> build\lib.win-amd64-cpython-
310\spacy\tests
    copying spacy\tests\util.py -> build\lib.win-amd64-cpython-310\
spacy\tests
    copying spacy\tests\__init__.py -> build\lib.win-amd64-cpython-310\
```

```
spacy\tests
  creating build\lib.win-amd64-cpython-310\spacy\tokens
  copying spacy\tokens\underscore.py -> build\lib.win-amd64-cpython-
310\spacy\tokens
  copying spacy\tokens\_serialize.py -> build\lib.win-amd64-cpython-
310\spacy\tokens
  copying spacy\tokens\__init__.py -> build\lib.win-amd64-cpython-310\
spacy\tokens
  creating build\lib.win-amd64-cpython-310\spacy\cli\converters
  copying spacy\cli\converters\conllu2json.py -> build\lib.win-amd64-
cpython-310\spacy\cli\converters
  copying spacy\cli\converters\conll_ner2json.py -> build\lib.win-
amd64-cpython-310\spacy\cli\converters
  copying spacy\cli\converters\iob2json.py -> build\lib.win-amd64-
cpython-310\spacy\cli\converters
  copying spacy\cli\converters\jsonl2json.py -> build\lib.win-amd64-
cpython-310\spacy\cli\converters
  copying spacy\cli\converters\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\cli\converters
  creating build\lib.win-amd64-cpython-310\spacy\lang\af
  copying spacy\lang\af\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\af
  copying spacy\lang\af\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\af
  creating build\lib.win-amd64-cpython-310\spacy\lang\ar
  copying spacy\lang\ar\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ar
  copying spacy\lang\ar\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ar
  copying spacy\lang\ar\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ar
  copying spacy\lang\ar\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ar
  copying spacy\lang\ar\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\ar
  copying spacy\lang\ar\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ar
  creating build\lib.win-amd64-cpython-310\spacy\lang\bg
  copying spacy\lang\bg\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\bg
  copying spacy\lang\bg\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\bg
  copying spacy\lang\bg\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\bg
  creating build\lib.win-amd64-cpython-310\spacy\lang\bn
  copying spacy\lang\bn\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\bn
  copying spacy\lang\bn\morph_rules.py -> build\lib.win-amd64-cpython-
310\spacy\lang\bn
  copying spacy\lang\bn\punctuation.py -> build\lib.win-amd64-cpython-
```

```
310\spacy\lang\bn
  copying spacy\lang\bn\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\bn
  copying spacy\lang\bn\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\bn
  copying spacy\lang\bn\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\bn
  creating build\lib.win-amd64-cpython-310\spacy\lang\ca
  copying spacy\lang\ca\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ca
  copying spacy\lang\ca\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ca
  copying spacy\lang\ca\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ca
  copying spacy\lang\ca\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ca
  copying spacy\lang\ca\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\ca
  copying spacy\lang\ca\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\ca
  copying spacy\lang\ca\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ca
  creating build\lib.win-amd64-cpython-310\spacy\lang\cs
  copying spacy\lang\cs\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\cs
  copying spacy\lang\cs\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\cs
  copying spacy\lang\cs\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\cs
  copying spacy\lang\cs\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\cs
  copying spacy\lang\cs\test_text.py -> build\lib.win-amd64-cpython-
310\spacy\lang\cs
  copying spacy\lang\cs\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\cs
  creating build\lib.win-amd64-cpython-310\spacy\lang\da
  copying spacy\lang\da\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\da
  copying spacy\lang\da\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\da
  copying spacy\lang\da\morph_rules.py -> build\lib.win-amd64-cpython-
310\spacy\lang\da
  copying spacy\lang\da\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\da
  copying spacy\lang\da\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\da
  copying spacy\lang\da\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\da
  copying spacy\lang\da\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\da
```

```

    creating build\lib.win-amd64-cpython-310\spacy\lang\de
    copying spacy\lang\de\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\de
    copying spacy\lang\de\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\de
    copying spacy\lang\de\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\de
    copying spacy\lang\de\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\de
    copying spacy\lang\de\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\de
    copying spacy\lang\de\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\de
    copying spacy\lang\de\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\de
    creating build\lib.win-amd64-cpython-310\spacy\lang\el
    copying spacy\lang\el\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\el
    copying spacy\lang\el\get_pos_from_wiktionary.py -> build\lib.win-
amd64-cpython-310\spacy\lang\el
    copying spacy\lang\el\lemmatizer.py -> build\lib.win-amd64-cpython-
310\spacy\lang\el
    copying spacy\lang\el\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\el
    copying spacy\lang\el\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\el
    copying spacy\lang\el\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\el
    copying spacy\lang\el\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\el
    copying spacy\lang\el\tag_map_fine.py -> build\lib.win-amd64-
cpython-310\spacy\lang\el
    copying spacy\lang\el\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\el
    copying spacy\lang\el\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\el
    creating build\lib.win-amd64-cpython-310\spacy\lang\en
    copying spacy\lang\en\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\en
    copying spacy\lang\en\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\en
    copying spacy\lang\en\morph_rules.py -> build\lib.win-amd64-cpython-
310\spacy\lang\en
    copying spacy\lang\en\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\en
    copying spacy\lang\en\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\en
    copying spacy\lang\en\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\en
    copying spacy\lang\en\tokenizer_exceptions.py -> build\lib.win-

```

```
amd64-cpython-310\spacy\lang\en
  copying spacy\lang\en\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\en
  creating build\lib.win-amd64-cpython-310\spacy\lang\es
  copying spacy\lang\es\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\es
  copying spacy\lang\es\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\es
  copying spacy\lang\es\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\es
  copying spacy\lang\es\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\es
  copying spacy\lang\es\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\es
  copying spacy\lang\es>tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\es
  copying spacy\lang\es\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\es
  copying spacy\lang\es\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\es
  creating build\lib.win-amd64-cpython-310\spacy\lang\et
  copying spacy\lang\et\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\et
  copying spacy\lang\et\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\et
  creating build\lib.win-amd64-cpython-310\spacy\lang\eu
  copying spacy\lang\eu\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\eu
  copying spacy\lang\eu\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\eu
  copying spacy\lang\eu\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\eu
  copying spacy\lang\eu\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\eu
  copying spacy\lang\eu\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\eu
  creating build\lib.win-amd64-cpython-310\spacy\lang\fa
  copying spacy\lang\fa\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fa
  copying spacy\lang\fa\generate_verbs_exc.py -> build\lib.win-amd64-
cpython-310\spacy\lang\fa
  copying spacy\lang\fa\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fa
  copying spacy\lang\fa\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fa
  copying spacy\lang\fa\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fa
  copying spacy\lang\fa\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\fa
  copying spacy\lang\fa>tag_map.py -> build\lib.win-amd64-cpython-310\
```



```

spacy\lang\fa
  copying spacy\lang\fa\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\fa
  copying spacy\lang\fa\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fa
  creating build\lib.win-amd64-cpython-310\spacy\lang\fi
  copying spacy\lang\fi\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fi
  copying spacy\lang\fi\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fi
  copying spacy\lang\fi\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fi
  copying spacy\lang\fi\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fi
  copying spacy\lang\fi\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\fi
  copying spacy\lang\fi\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fi
  creating build\lib.win-amd64-cpython-310\spacy\lang\fr
  copying spacy\lang\fr\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fr
  copying spacy\lang\fr\lemmatizer.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fr
  copying spacy\lang\fr\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fr
  copying spacy\lang\fr\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fr
  copying spacy\lang\fr\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fr
  copying spacy\lang\fr\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\fr
  copying spacy\lang\fr\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\fr
  copying spacy\lang\fr\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\fr
  copying spacy\lang\fr\_tokenizer_exceptions_list.py -> build\
lib.win-amd64-cpython-310\spacy\lang\fr
  copying spacy\lang\fr\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\fr
  creating build\lib.win-amd64-cpython-310\spacy\lang\ga
  copying spacy\lang\ga\irish_morphology_helpers.py -> build\lib.win-
amd64-cpython-310\spacy\lang\ga
  copying spacy\lang\ga\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ga
  copying spacy\lang\ga\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\ga
  copying spacy\lang\ga\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\ga
  copying spacy\lang\ga\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ga

```

```
creating build\lib.win-amd64-cpython-310\spacy\lang\gu
copying spacy\lang\gu\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\gu
copying spacy\lang\gu\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\gu
copying spacy\lang\gu\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\gu
creating build\lib.win-amd64-cpython-310\spacy\lang\he
copying spacy\lang\he\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\he
copying spacy\lang\he\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\he
copying spacy\lang\he\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\he
copying spacy\lang\he\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\he
creating build\lib.win-amd64-cpython-310\spacy\lang\hi
copying spacy\lang\hi\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hi
copying spacy\lang\hi\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hi
copying spacy\lang\hi\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hi
copying spacy\lang\hi\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hi
creating build\lib.win-amd64-cpython-310\spacy\lang\hr
copying spacy\lang\hr\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hr
copying spacy\lang\hr\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hr
copying spacy\lang\hr>tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\hr
copying spacy\lang\hr\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hr
creating build\lib.win-amd64-cpython-310\spacy\lang\hu
copying spacy\lang\hu\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hu
copying spacy\lang\hu\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hu
copying spacy\lang\hu\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hu
copying spacy\lang\hu\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\hu
copying spacy\lang\hu\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hu
creating build\lib.win-amd64-cpython-310\spacy\lang\hy
copying spacy\lang\hy\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hy
copying spacy\lang\hy\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hy
```

```
    copying spacy\lang\hy\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hy
    copying spacy\lang\hy\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\hy
    copying spacy\lang\hy\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\hy
    creating build\lib.win-amd64-cpython-310\spacy\lang\id
    copying spacy\lang\id\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\id
    copying spacy\lang\id\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\id
    copying spacy\lang\id\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\id
    copying spacy\lang\id\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\id
    copying spacy\lang\id\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\id
    copying spacy\lang\id\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\id
    copying spacy\lang\id\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\id
    copying spacy\lang\id\_tokenizer_exceptions_list.py -> build\
lib.win-amd64-cpython-310\spacy\lang\id
    copying spacy\lang\id\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\id
    creating build\lib.win-amd64-cpython-310\spacy\lang\is
    copying spacy\lang\is\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\is
    copying spacy\lang\is\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\is
    creating build\lib.win-amd64-cpython-310\spacy\lang\it
    copying spacy\lang\it\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\it
    copying spacy\lang\it\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\it
    copying spacy\lang\it\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\it
    copying spacy\lang\it\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\it
    copying spacy\lang\it\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\it
    copying spacy\lang\it\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\it
    creating build\lib.win-amd64-cpython-310\spacy\lang\ja
    copying spacy\lang\ja\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ja
    copying spacy\lang\ja\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ja
    copying spacy\lang\ja\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\ja
```

```
    copying spacy\lang\ja\tag_bigram_map.py -> build\lib.win-amd64-
cpython-310\spacy\lang\ja
    copying spacy\lang\ja\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\ja
    copying spacy\lang\ja\tag_orth_map.py -> build\lib.win-amd64-
cpython-310\spacy\lang\ja
    copying spacy\lang\ja\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ja
    creating build\lib.win-amd64-cpython-310\spacy\lang\kn
    copying spacy\lang\kn\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\kn
    copying spacy\lang\kn\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\kn
    copying spacy\lang\kn\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\kn
    creating build\lib.win-amd64-cpython-310\spacy\lang\ko
    copying spacy\lang\ko\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ko
    copying spacy\lang\ko\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ko
    copying spacy\lang\ko\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ko
    copying spacy\lang\ko\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\ko
    copying spacy\lang\ko\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ko
    creating build\lib.win-amd64-cpython-310\spacy\lang\lb
    copying spacy\lang\lb\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lb
    copying spacy\lang\lb\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lb
    copying spacy\lang\lb\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lb
    copying spacy\lang\lb\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lb
    copying spacy\lang\lb\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\lb
    copying spacy\lang\lb\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\lb
    copying spacy\lang\lb\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lb
    creating build\lib.win-amd64-cpython-310\spacy\lang\lij
    copying spacy\lang\lij\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lij
    copying spacy\lang\lij\punctuation.py -> build\lib.win-amd64-
cpython-310\spacy\lang\lij
    copying spacy\lang\lij\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lij
    copying spacy\lang\lij\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\lij
```

```

    copying spacy\lang\lij\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lij
    creating build\lib.win-amd64-cpython-310\spacy\lang\lt
    copying spacy\lang\lt\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lt
    copying spacy\lang\lt\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lt
    copying spacy\lang\lt\morph_rules.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lt
    copying spacy\lang\lt\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lt
    copying spacy\lang\lt\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lt
    copying spacy\lang\lt\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\lt
    copying spacy\lang\lt\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\lt
    copying spacy\lang\lt\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lt
    creating build\lib.win-amd64-cpython-310\spacy\lang\lv
    copying spacy\lang\lv\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lv
    copying spacy\lang\lv\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\lv
    creating build\lib.win-amd64-cpython-310\spacy\lang\mk
    copying spacy\lang\mk\lemmatizer.py -> build\lib.win-amd64-cpython-
310\spacy\lang\mk
    copying spacy\lang\mk\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\mk
    copying spacy\lang\mk\morph_rules.py -> build\lib.win-amd64-cpython-
310\spacy\lang\mk
    copying spacy\lang\mk\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\mk
    copying spacy\lang\mk\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\mk
    copying spacy\lang\mk\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\mk
    creating build\lib.win-amd64-cpython-310\spacy\lang\ml
    copying spacy\lang\ml\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ml
    copying spacy\lang\ml\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ml
    copying spacy\lang\ml\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ml
    copying spacy\lang\ml\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ml
    creating build\lib.win-amd64-cpython-310\spacy\lang\mr
    copying spacy\lang\mr\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\mr
    copying spacy\lang\mr\__init__.py -> build\lib.win-amd64-cpython-

```

```

310\spacy\lang\mr
    creating build\lib.win-amd64-cpython-310\spacy\lang\nb
    copying spacy\lang\nb\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nb
    copying spacy\lang\nb\morph_rules.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nb
    copying spacy\lang\nb\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nb
    copying spacy\lang\nb\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nb
    copying spacy\lang\nb\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\nb
    copying spacy\lang\nb>tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\nb
    copying spacy\lang\nb\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\nb
    copying spacy\lang\nb\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nb
    creating build\lib.win-amd64-cpython-310\spacy\lang\ne
    copying spacy\lang\ne\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ne
    copying spacy\lang\ne\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ne
    copying spacy\lang\ne\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ne
    copying spacy\lang\ne\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ne
    creating build\lib.win-amd64-cpython-310\spacy\lang\nl
    copying spacy\lang\nl\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nl
    copying spacy\lang\nl\lemmatizer.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nl
    copying spacy\lang\nl\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nl
    copying spacy\lang\nl\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nl
    copying spacy\lang\nl\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nl
    copying spacy\lang\nl>tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\nl
    copying spacy\lang\nl\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\nl
    copying spacy\lang\nl\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\nl
    creating build\lib.win-amd64-cpython-310\spacy\lang\pl
    copying spacy\lang\pl\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pl
    copying spacy\lang\pl\lemmatizer.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pl
    copying spacy\lang\pl\lex_attrs.py -> build\lib.win-amd64-cpython-

```

```
310\spacy\lang\pl
  copying spacy\lang\pl\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pl
  copying spacy\lang\pl\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pl
  copying spacy\lang\pl\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\pl
  copying spacy\lang\pl\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pl
  creating build\lib.win-amd64-cpython-310\spacy\lang\pt
  copying spacy\lang\pt\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pt
  copying spacy\lang\pt\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pt
  copying spacy\lang\pt\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pt
  copying spacy\lang\pt\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pt
  copying spacy\lang\pt\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\pt
  copying spacy\lang\pt\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\pt
  copying spacy\lang\pt\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\pt
  creating build\lib.win-amd64-cpython-310\spacy\lang\ro
  copying spacy\lang\ro\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ro
  copying spacy\lang\ro\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ro
  copying spacy\lang\ro\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ro
  copying spacy\lang\ro\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ro
  copying spacy\lang\ro\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\ro
  copying spacy\lang\ro\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\ro
  copying spacy\lang\ro\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ro
  creating build\lib.win-amd64-cpython-310\spacy\lang\ru
  copying spacy\lang\ru\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ru
  copying spacy\lang\ru\lemmatizer.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ru
  copying spacy\lang\ru\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ru
  copying spacy\lang\ru\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ru
  copying spacy\lang\ru\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\ru
```

```
    copying spacy\lang\ru\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\ru
    copying spacy\lang\ru\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ru
    creating build\lib.win-amd64-cpython-310\spacy\lang\sa
    copying spacy\lang\sa\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sa
    copying spacy\lang\sa\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sa
    copying spacy\lang\sa\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sa
    copying spacy\lang\sa\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sa
    creating build\lib.win-amd64-cpython-310\spacy\lang\si
    copying spacy\lang\si\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\si
    copying spacy\lang\si\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\si
    copying spacy\lang\si\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\si
    copying spacy\lang\si\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\si
    creating build\lib.win-amd64-cpython-310\spacy\lang\sk
    copying spacy\lang\sk\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sk
    copying spacy\lang\sk\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sk
    copying spacy\lang\sk\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sk
    copying spacy\lang\sk>tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\sk
    copying spacy\lang\sk\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sk
    creating build\lib.win-amd64-cpython-310\spacy\lang\sl
    copying spacy\lang\sl\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sl
    copying spacy\lang\sl\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sl
    creating build\lib.win-amd64-cpython-310\spacy\lang\sq
    copying spacy\lang\sq\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sq
    copying spacy\lang\sq\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sq
    copying spacy\lang\sq\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sq
    creating build\lib.win-amd64-cpython-310\spacy\lang\sr
    copying spacy\lang\sr\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sr
    copying spacy\lang\sr\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sr
```



```
    copying spacy\lang\sr\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sr
    copying spacy\lang\sr\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\sr
    copying spacy\lang\sr\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sr
    creating build\lib.win-amd64-cpython-310\spacy\lang\sv
    copying spacy\lang\sv\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sv
    copying spacy\lang\sv\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sv
    copying spacy\lang\sv\morph_rules.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sv
    copying spacy\lang\sv\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sv
    copying spacy\lang\sv\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\sv
    copying spacy\lang\sv\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\sv
    copying spacy\lang\sv\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\sv
    copying spacy\lang\sv\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\sv
    creating build\lib.win-amd64-cpython-310\spacy\lang\ta
    copying spacy\lang\ta\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ta
    copying spacy\lang\ta\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ta
    copying spacy\lang\ta\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ta
    copying spacy\lang\ta\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ta
    creating build\lib.win-amd64-cpython-310\spacy\lang\te
    copying spacy\lang\te\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\te
    copying spacy\lang\te\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\te
    copying spacy\lang\te\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\te
    copying spacy\lang\te\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\te
    creating build\lib.win-amd64-cpython-310\spacy\lang\th
    copying spacy\lang\th\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\th
    copying spacy\lang\th\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\th
    copying spacy\lang\th\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\th
    copying spacy\lang\th\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\th
```

```
    copying spacy\lang\th\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\th
    creating build\lib.win-amd64-cpython-310\spacy\lang\tl
    copying spacy\lang\tl\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tl
    copying spacy\lang\tl\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tl
    copying spacy\lang\tl\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\tl
    copying spacy\lang\tl\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tl
    creating build\lib.win-amd64-cpython-310\spacy\lang\tr
    copying spacy\lang\tr\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    copying spacy\lang\tr\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    copying spacy\lang\tr\morph_rules.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    copying spacy\lang\tr\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    copying spacy\lang\tr\syntax_iterators.py -> build\lib.win-amd64-
cpython-310\spacy\lang\tr
    copying spacy\lang\tr\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\tr
    copying spacy\lang\tr\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    creating build\lib.win-amd64-cpython-310\spacy\lang\tr
    copying spacy\lang\tr\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    copying spacy\lang\tr\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    copying spacy\lang\tr\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    copying spacy\lang\tr\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    copying spacy\lang\tr\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\tr
    copying spacy\lang\tr\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\tr
    creating build\lib.win-amd64-cpython-310\spacy\lang\uk
    copying spacy\lang\uk\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\uk
    copying spacy\lang\uk\lemmatizer.py -> build\lib.win-amd64-cpython-
310\spacy\lang\uk
    copying spacy\lang\uk\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\uk
    copying spacy\lang\uk\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\uk
    copying spacy\lang\uk\tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\uk
```

```
    copying spacy\lang\uk\tokenizer_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\lang\uk
    copying spacy\lang\uk\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\uk
    creating build\lib.win-amd64-cpython-310\spacy\lang\ur
    copying spacy\lang\ur\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ur
    copying spacy\lang\ur\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ur
    copying spacy\lang\ur\punctuation.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ur
    copying spacy\lang\ur\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ur
    copying spacy\lang\ur>tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\ur
    copying spacy\lang\ur\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\ur
    creating build\lib.win-amd64-cpython-310\spacy\lang\vi
    copying spacy\lang\vi\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\vi
    copying spacy\lang\vi\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\vi
    copying spacy\lang\vi>tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\vi
    copying spacy\lang\vi\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\vi
    creating build\lib.win-amd64-cpython-310\spacy\lang\xx
    copying spacy\lang\xx\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\xx
    copying spacy\lang\xx\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\xx
    creating build\lib.win-amd64-cpython-310\spacy\lang\yo
    copying spacy\lang\yo\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\yo
    copying spacy\lang\yo\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\yo
    copying spacy\lang\yo\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\yo
    copying spacy\lang\yo\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\lang\yo
    creating build\lib.win-amd64-cpython-310\spacy\lang\zh
    copying spacy\lang\zh\examples.py -> build\lib.win-amd64-cpython-
310\spacy\lang\zh
    copying spacy\lang\zh\lex_attrs.py -> build\lib.win-amd64-cpython-
310\spacy\lang\zh
    copying spacy\lang\zh\stop_words.py -> build\lib.win-amd64-cpython-
310\spacy\lang\zh
    copying spacy\lang\zh>tag_map.py -> build\lib.win-amd64-cpython-310\
spacy\lang\zh
    copying spacy\lang\zh\__init__.py -> build\lib.win-amd64-cpython-
```

```
310\spacy\lang\zh
  creating build\lib.win-amd64-cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_add_entities.py -> build\lib.win-amd64-
cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_array.py -> build\lib.win-amd64-
cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_creation.py -> build\lib.win-amd64-
cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_doc_api.py -> build\lib.win-amd64-
cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_morphanalysis.py -> build\lib.win-
amd64-cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_pickle_doc.py -> build\lib.win-amd64-
cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_tokenize_merge.py -> build\lib.win-
amd64-cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_tokenize_split.py -> build\lib.win-
amd64-cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_span.py -> build\lib.win-amd64-cpython-
310\spacy\tests\doc
  copying spacy\tests\doc\test_token_api.py -> build\lib.win-amd64-
cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_to_json.py -> build\lib.win-amd64-
cpython-310\spacy\tests\doc
  copying spacy\tests\doc\test_underscore.py -> build\lib.win-amd64-
cpython-310\spacy\tests\doc
  copying spacy\tests\doc\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\tests\doc
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang
  copying spacy\tests\lang\test_attrs.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang
  copying spacy\tests\lang\test_initialize.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang
  copying spacy\tests\lang\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\tests\lang
  creating build\lib.win-amd64-cpython-310\spacy\tests\matcher
  copying spacy\tests\matcher\test_matcher_api.py -> build\lib.win-
amd64-cpython-310\spacy\tests\matcher
  copying spacy\tests\matcher\test_matcher_logic.py -> build\lib.win-
amd64-cpython-310\spacy\tests\matcher
  copying spacy\tests\matcher\test_pattern_validation.py -> build\
lib.win-amd64-cpython-310\spacy\tests\matcher
  copying spacy\tests\matcher\test_phrase_matcher.py -> build\lib.win-
amd64-cpython-310\spacy\tests\matcher
  copying spacy\tests\matcher\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\matcher
  creating build\lib.win-amd64-cpython-310\spacy\tests\morphology
  copying spacy\tests\morphology\test_morph_features.py -> build\
lib.win-amd64-cpython-310\spacy\tests\morphology
  copying spacy\tests\morphology\__init__.py -> build\lib.win-amd64-
```

```
cpython-310\spacy\tests\morphology
  creating build\lib.win-amd64-cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_add_label.py -> build\lib.win-amd64-
cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_arc_eager_oracle.py -> build\
lib.win-amd64-cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_ner.py -> build\lib.win-amd64-
cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_neural_parser.py -> build\lib.win-
amd64-cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_nn_beam.py -> build\lib.win-amd64-
cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_nonproj.py -> build\lib.win-amd64-
cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_parse.py -> build\lib.win-amd64-
cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_parse_navigate.py -> build\lib.win-
amd64-cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_preset_sbd.py -> build\lib.win-
amd64-cpython-310\spacy\tests\parser
  copying spacy\tests\parser\test_space_attachment.py -> build\
lib.win-amd64-cpython-310\spacy\tests\parser
  copying spacy\tests\parser\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\parser
  creating build\lib.win-amd64-cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\test_analysis.py -> build\lib.win-
amd64-cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\test_entity_linker.py -> build\lib.win-
amd64-cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\test_entity_ruler.py -> build\lib.win-
amd64-cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\test_factories.py -> build\lib.win-
amd64-cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\test_functions.py -> build\lib.win-
amd64-cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\test_pipe_methods.py -> build\lib.win-
amd64-cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\test_sentencizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\test_tagger.py -> build\lib.win-amd64-
cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\test_textcat.py -> build\lib.win-amd64-
cpython-310\spacy\tests\pipeline
  copying spacy\tests\pipeline\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\pipeline
  creating build\lib.win-amd64-cpython-310\spacy\tests\regression
  copying spacy\tests\regression\test_issue1-1000.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
  copying spacy\tests\regression\test_issue1001-1500.py -> build\
lib.win-amd64-cpython-310\spacy\tests\regression
```

[illegible]

[illegible]

```
    copying spacy\tests\regression\test_issue5082.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
    copying spacy\tests\regression\test_issue5137.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
    copying spacy\tests\regression\test_issue5152.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
    copying spacy\tests\regression\test_issue5230.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
    copying spacy\tests\regression\test_issue5458.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
    copying spacy\tests\regression\test_issue5838.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
    copying spacy\tests\regression\test_issue5918.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
    copying spacy\tests\regression\test_issue6177.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
    copying spacy\tests\regression\test_issue6207.py -> build\lib.win-
amd64-cpython-310\spacy\tests\regression
    copying spacy\tests\regression\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\regression
    creating build\lib.win-amd64-cpython-310\spacy\tests\serialize
    copying spacy\tests\serialize\test_serialize_doc.py -> build\
lib.win-amd64-cpython-310\spacy\tests\serialize
    copying spacy\tests\serialize\test_serialize_extension_attrs.py ->
build\lib.win-amd64-cpython-310\spacy\tests\serialize
    copying spacy\tests\serialize\test_serialize_kb.py -> build\lib.win-
amd64-cpython-310\spacy\tests\serialize
    copying spacy\tests\serialize\test_serialize_language.py -> build\
lib.win-amd64-cpython-310\spacy\tests\serialize
    copying spacy\tests\serialize\test_serialize_pipeline.py -> build\
lib.win-amd64-cpython-310\spacy\tests\serialize
    copying spacy\tests\serialize\test_serialize_tokenizer.py -> build\
lib.win-amd64-cpython-310\spacy\tests\serialize
    copying spacy\tests\serialize\test_serialize_vocab_strings.py ->
build\lib.win-amd64-cpython-310\spacy\tests\serialize
    copying spacy\tests\serialize\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\serialize
    creating build\lib.win-amd64-cpython-310\spacy\tests\tokenizer
    copying spacy\tests\tokenizer\test_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\tests\tokenizer
    copying spacy\tests\tokenizer\test_explain.py -> build\lib.win-
amd64-cpython-310\spacy\tests\tokenizer
    copying spacy\tests\tokenizer\test_naughty_strings.py -> build\
lib.win-amd64-cpython-310\spacy\tests\tokenizer
    copying spacy\tests\tokenizer\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\tokenizer
    copying spacy\tests\tokenizer\test_urls.py -> build\lib.win-amd64-
cpython-310\spacy\tests\tokenizer
    copying spacy\tests\tokenizer\test_whitespace.py -> build\lib.win-
amd64-cpython-310\spacy\tests\tokenizer
```



```
    copying spacy\tests\tokenizer\__init__.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\tokenizer  
    creating build\lib.win-amd64-cpython-310\spacy\tests\vocab_vectors  
    copying spacy\tests\vocab_vectors\test_lexeme.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\vocab_vectors  
    copying spacy\tests\vocab_vectors\test_lookups.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\vocab_vectors  
    copying spacy\tests\vocab_vectors\test_similarity.py -> build\  
lib.win-amd64-cpython-310\spacy\tests\vocab_vectors  
    copying spacy\tests\vocab_vectors\test_stringstore.py -> build\  
lib.win-amd64-cpython-310\spacy\tests\vocab_vectors  
    copying spacy\tests\vocab_vectors\test_vectors.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\vocab_vectors  
    copying spacy\tests\vocab_vectors\test_vocab_api.py -> build\  
lib.win-amd64-cpython-310\spacy\tests\vocab_vectors  
    copying spacy\tests\vocab_vectors\__init__.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\vocab_vectors  
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ar  
    copying spacy\tests\lang\ar\test_exceptions.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\ar  
    copying spacy\tests\lang\ar\test_text.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\ar  
    copying spacy\tests\lang\ar\__init__.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\ar  
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\bn  
    copying spacy\tests\lang\bn\test_tokenizer.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\bn  
    copying spacy\tests\lang\bn\__init__.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\bn  
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ca  
    copying spacy\tests\lang\ca\test_exception.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\ca  
    copying spacy\tests\lang\ca\test_prefix_suffix_infix.py -> build\  
lib.win-amd64-cpython-310\spacy\tests\lang\ca  
    copying spacy\tests\lang\ca\test_text.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\ca  
    copying spacy\tests\lang\ca\__init__.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\ca  
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\cs  
    copying spacy\tests\lang\cs\test_text.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\cs  
    copying spacy\tests\lang\cs\__init__.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\cs  
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\da  
    copying spacy\tests\lang\da\test_exceptions.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\da  
    copying spacy\tests\lang\da\test_prefix_suffix_infix.py -> build\  
lib.win-amd64-cpython-310\spacy\tests\lang\da  
    copying spacy\tests\lang\da\test_text.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\da
```

```
    copying spacy\tests\lang\da\__init__.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\da  
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\de  
    copying spacy\tests\lang\de\test_exceptions.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\de  
    copying spacy\tests\lang\de\test_noun_chunks.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\de  
    copying spacy\tests\lang\de\test_parser.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\de  
    copying spacy\tests\lang\de\test_prefix_suffix_infix.py -> build\  
lib.win-amd64-cpython-310\spacy\tests\lang\de  
    copying spacy\tests\lang\de\test_text.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\de  
    copying spacy\tests\lang\de\__init__.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\de  
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\el  
    copying spacy\tests\lang\el\test_exception.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\el  
    copying spacy\tests\lang\el\test_noun_chunks.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\el  
    copying spacy\tests\lang\el\test_text.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\el  
    copying spacy\tests\lang\el\__init__.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\el  
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_customized_tokenizer.py -> build\  
lib.win-amd64-cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_exceptions.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_indices.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_noun_chunks.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_parser.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_prefix_suffix_infix.py -> build\  
lib.win-amd64-cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_punct.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_sbd.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_tagger.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\test_text.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\en  
    copying spacy\tests\lang\en\__init__.py -> build\lib.win-amd64-  
cpython-310\spacy\tests\lang\en  
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\es  
    copying spacy\tests\lang\es\test_exception.py -> build\lib.win-  
amd64-cpython-310\spacy\tests\lang\es
```

```
    copying spacy\tests\lang\es\test_noun_chunks.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\es
    copying spacy\tests\lang\es\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\es
    copying spacy\tests\lang\es\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\es
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\eu
    copying spacy\tests\lang\eu\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\eu
    copying spacy\tests\lang\eu\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\eu
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\fa
    copying spacy\tests\lang\fa\test_noun_chunks.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\fa
    copying spacy\tests\lang\fa\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\fa
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\fi
    copying spacy\tests\lang\fi\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\fi
    copying spacy\tests\lang\fi\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\fi
    copying spacy\tests\lang\fi\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\fi
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\fr
    copying spacy\tests\lang\fr\test_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\fr
    copying spacy\tests\lang\fr\test_noun_chunks.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\fr
    copying spacy\tests\lang\fr\test_prefix_suffix_infix.py -> build\
lib.win-amd64-cpython-310\spacy\tests\lang\fr
    copying spacy\tests\lang\fr\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\fr
    copying spacy\tests\lang\fr\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\fr
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ga
    copying spacy\tests\lang\ga\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ga
    copying spacy\tests\lang\ga\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ga
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\gu
    copying spacy\tests\lang\gu\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\gu
    copying spacy\tests\lang\gu\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\gu
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\he
    copying spacy\tests\lang\he\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\he
    copying spacy\tests\lang\he\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\he
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\hi
```

```
    copying spacy\tests\lang\hi\test_lex_attrs.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\hi
    copying spacy\tests\lang\hi\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\hi
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\hu
    copying spacy\tests\lang\hu\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\hu
    copying spacy\tests\lang\hu\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\hu
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\hy
    copying spacy\tests\lang\hy\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\hy
    copying spacy\tests\lang\hy\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\hy
    copying spacy\tests\lang\hy\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\hy
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\id
    copying spacy\tests\lang\id\test_noun_chunks.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\id
    copying spacy\tests\lang\id\test_prefix_suffix_infix.py -> build\
lib.win-amd64-cpython-310\spacy\tests\lang\id
    copying spacy\tests\lang\id\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\id
    copying spacy\tests\lang\id\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\id
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\it
    copying spacy\tests\lang\it\test_prefix_suffix_infix.py -> build\
lib.win-amd64-cpython-310\spacy\tests\lang\it
    copying spacy\tests\lang\it\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\it
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ja
    copying spacy\tests\lang\ja\test_lemmatization.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ja
    copying spacy\tests\lang\ja\test_serialize.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ja
    copying spacy\tests\lang\ja\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ja
    copying spacy\tests\lang\ja\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ja
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ko
    copying spacy\tests\lang\ko\test_lemmatization.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ko
    copying spacy\tests\lang\ko\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ko
    copying spacy\tests\lang\ko\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ko
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\lb
    copying spacy\tests\lang\lb\test_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\lb
    copying spacy\tests\lang\lb\test_prefix_suffix_infix.py -> build\
```

```
lib.win-amd64-cpython-310\spacy\tests\lang\lb
  copying spacy\tests\lang\lb\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\lb
  copying spacy\tests\lang\lb\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\lb
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang\lt
  copying spacy\tests\lang\lt\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\lt
  copying spacy\tests\lang\lt\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\lt
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang\mk
  copying spacy\tests\lang\mk\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\mk
  copying spacy\tests\lang\mk\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\mk
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ml
  copying spacy\tests\lang\ml\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ml
  copying spacy\tests\lang\ml\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ml
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang\nb
  copying spacy\tests\lang\nb\test_noun_chunks.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\nb
  copying spacy\tests\lang\nb\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\nb
  copying spacy\tests\lang\nb\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\nb
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ne
  copying spacy\tests\lang\ne\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ne
  copying spacy\tests\lang\ne\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ne
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang\nl
  copying spacy\tests\lang\nl\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\nl
  copying spacy\tests\lang\nl\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\nl
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang\pl
  copying spacy\tests\lang\pl\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\pl
  copying spacy\tests\lang\pl\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\pl
  copying spacy\tests\lang\pl\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\pl
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang\pt
  copying spacy\tests\lang\pt\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\pt
  copying spacy\tests\lang\pt\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\pt
  creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ro
```

```
    copying spacy\tests\lang\ro\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ro
    copying spacy\tests\lang\ro\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ro
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ru
    copying spacy\tests\lang\ru\test_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ru
    copying spacy\tests\lang\ru\test_lemmatizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ru
    copying spacy\tests\lang\ru\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ru
    copying spacy\tests\lang\ru\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\ru
    copying spacy\tests\lang\ru\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ru
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\sa
    copying spacy\tests\lang\sa\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\sa
    copying spacy\tests\lang\sa\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\sa
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\sr
    copying spacy\tests\lang\sr\test_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\sr
    copying spacy\tests\lang\sr\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\sr
    copying spacy\tests\lang\sr\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\sr
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\sv
    copying spacy\tests\lang\sv\test_exceptions.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\sv
    copying spacy\tests\lang\sv\test_lex_attrs.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\sv
    copying spacy\tests\lang\sv\test_noun_chunks.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\sv
    copying spacy\tests\lang\sv\test_prefix_suffix_infix.py -> build\
lib.win-amd64-cpython-310\spacy\tests\lang\sv
    copying spacy\tests\lang\sv\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\sv
    copying spacy\tests\lang\sv\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\sv
    copying spacy\tests\lang\sv\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\sv
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\th
    copying spacy\tests\lang\th\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\th
    copying spacy\tests\lang\th\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\th
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\tr
    copying spacy\tests\lang\tr\test_noun_chunks.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\tr
```

```
    copying spacy\tests\lang\tr\test_parser.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\tr
    copying spacy\tests\lang\tr\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\tr
    copying spacy\tests\lang\tr\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\tr
    copying spacy\tests\lang\tr\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\tr
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\tr
    copying spacy\tests\lang\tr\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\tr
    copying spacy\tests\lang\tr\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\tr
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\uk
    copying spacy\tests\lang\uk\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\uk
    copying spacy\tests\lang\uk\test_tokenizer_exc.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\uk
    copying spacy\tests\lang\uk\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\uk
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\ur
    copying spacy\tests\lang\ur\test_prefix_suffix_infix.py -> build\
lib.win-amd64-cpython-310\spacy\tests\lang\ur
    copying spacy\tests\lang\ur\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ur
    copying spacy\tests\lang\ur\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\ur
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\yo
    copying spacy\tests\lang\yo\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\yo
    copying spacy\tests\lang\yo\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\yo
    creating build\lib.win-amd64-cpython-310\spacy\tests\lang\zh
    copying spacy\tests\lang\zh\test_serialize.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\zh
    copying spacy\tests\lang\zh\test_text.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\zh
    copying spacy\tests\lang\zh\test_tokenizer.py -> build\lib.win-
amd64-cpython-310\spacy\tests\lang\zh
    copying spacy\tests\lang\zh\__init__.py -> build\lib.win-amd64-
cpython-310\spacy\tests\lang\zh
    running egg_info
    writing spacy.egg-info\PKG-INFO
    writing dependency_links to spacy.egg-info\dependency_links.txt
    writing requirements to spacy.egg-info\requires.txt
    writing top-level names to spacy.egg-info\top_level.txt
    reading manifest file 'spacy.egg-info\SOURCES.txt'
    reading manifest template 'MANIFEST.in'
    warning: no previously-included files matching '*.json' found under
directory 'spacy\lang'
```

```
warning: no files found matching '*.json.gz' under directory 'spacy\lang'
adding license file 'LICENSE'
writing manifest file 'spacy.egg-info\SOURCES.txt'
C:\Users\black\AppData\Local\Temp\pip-build-env-waksoeh_\overlay\
Lib\site-packages\setuptools\command\build_py.py:201: _Warning:
Package 'spacy.matcher' is absent from the `packages` configuration.
!!
```

```
*****
*****
```

```
#####
# Package would be ignored #
#####
Python recognizes 'spacy.matcher' as an importable
package[^1],
but it is absent from setuptools' `packages` configuration.

This leads to an ambiguous overall configuration. If you
want to distribute this
package, please make sure that 'spacy.matcher' is explicitly
added
to the `packages` configuration field.
```

```
Alternatively, you can also rely on setuptools' discovery
methods
(for example by using
`find_namespace_packages(...)`/`find_namespace:`
instead of `find_packages(...)`/`find:`).
```

You can read more about "package discovery" on setuptools documentation page:

-
https://setuptools.pypa.io/en/latest/userguide/package_discovery.html

```
If you don't want 'spacy.matcher' to be distributed and are
already explicitly excluding 'spacy.matcher' via
`find_namespace_packages(...)/find_namespace` or
`find_packages(...)/find`,
you can try to use `exclude_package_data`, or `include-
package-data=False` in
combination with a more fine grained `package-data`
configuration.
```

You can read more about "package data files" on setuptools documentation page:

-

<https://setuptools.pypa.io/en/latest/userguide/datafiles.html>

[^1]: For Python, any directory (with suitable naming) can be imported, even if it does not contain any `.py` files. On the other hand, currently there is no concept of package data directory, all directories are treated like packages.

```
*****
*****
```

```
!!
    check.warn(importable)
C:\Users\black\AppData\Local\Temp\pip-build-env-waksoeh\overlay\
Lib\site-packages\setuptools\command\build_py.py:201: _Warning:
Package 'spacy.pipeline' is absent from the `packages` configuration.
!!
```

```
*****
*****
```

```
#####
# Package would be ignored #
#####
Python recognizes 'spacy.pipeline' as an importable
package[^1],
but it is absent from setuptools' `packages` configuration.
```

This leads to an ambiguous overall configuration. If you want to distribute this package, please make sure that 'spacy.pipeline' is explicitly added to the ``packages`` configuration field.

Alternatively, you can also rely on setuptools' discovery methods (for example by using ``find_namespace_packages(...)`/`find_namespace:`` instead of ``find_packages(...)`/`find:``).

You can read more about "package discovery" on setuptools documentation page:

-
https://setuptools.pypa.io/en/latest/userguide/package_discovery.html

If you don't want 'spacy.pipeline' to be distributed and are already explicitly excluding 'spacy.pipeline' via

`find_namespace_packages(...)/find_namespace` or
`find_packages(...)/find`,
you can try to use `exclude_package_data`, or `include-
package-data=False` in
combination with a more fine grained `package-data`
configuration.

You can read more about "package data files" on setuptools
documentation page:

-
<https://setuptools.pypa.io/en/latest/userguide/datafiles.html>

[^1]: For Python, any directory (with suitable naming) can
be imported,
even if it does not contain any `.py` files.
On the other hand, currently there is no concept of
package data
directory, all directories are treated like packages.


```
!!
    check.warn(importable)
C:\Users\black\AppData\Local\Temp\pip-build-env-waksoeh\overlay\
Lib\site-packages\setuptools\command\build_py.py:201: _Warning:
Package 'spacy.syntax' is absent from the `packages` configuration.
!!
```


```
#####
# Package would be ignored #
#####
Python recognizes 'spacy.syntax' as an importable
package[^1],
but it is absent from setuptools' `packages` configuration.
```

This leads to an ambiguous overall configuration. If you
want to distribute this
package, please make sure that 'spacy.syntax' is explicitly
added
to the `packages` configuration field.

Alternatively, you can also rely on setuptools' discovery
methods
(for example by using

```
`find_namespace_packages(...)`/`find_namespace:`  
    instead of `find_packages(...)`/`find:`).
```

You can read more about "package discovery" on setuptools documentation page:

-
https://setuptools.pypa.io/en/latest/userguide/package_discovery.html

If you don't want 'spacy.syntax' to be distributed and are already explicitly excluding 'spacy.syntax' via
`find_namespace_packages(...)/find_namespace` or
`find_packages(...)/find`,
you can try to use `exclude_package_data`, or `include-package-data=False` in
combination with a more fine grained `package-data`
configuration.

You can read more about "package data files" on setuptools documentation page:

-
<https://setuptools.pypa.io/en/latest/userguide/datafiles.html>

[^1]: For Python, any directory (with suitable naming) can be imported,
even if it does not contain any `.py` files.
On the other hand, currently there is no concept of
package data
directory, all directories are treated like packages.

```
*****  
*****
```

```
!!  
    check.warn(importable)  
C:\Users\black\AppData\Local\Temp\pip-build-env-waksoeh_\overlay\  
Lib\site-packages\setuptools\command\build_py.py:201: _Warning:  
Package 'spacy.tokens' is absent from the `packages` configuration.  
!!
```

```
*****  
*****
```

```
#####  
# Package would be ignored #  
#####  
Python recognizes 'spacy.tokens' as an importable  
package[^1],
```

but it is absent from setuptools' `packages` configuration.

This leads to an ambiguous overall configuration. If you want to distribute this package, please make sure that 'spacy.tokens' is explicitly added to the `packages` configuration field.

Alternatively, you can also rely on setuptools' discovery methods

(for example by using
`find_namespace_packages(...)`/`find_namespace:`
instead of `find_packages(...)`/`find:`).

You can read more about "package discovery" on setuptools documentation page:

-
https://setuptools.pypa.io/en/latest/userguide/package_discovery.html

If you don't want 'spacy.tokens' to be distributed and are already explicitly excluding 'spacy.tokens' via
`find_namespace_packages(...)/find_namespace` or
`find_packages(...)/find`,
you can try to use `exclude_package_data`, or `include-package-data=False` in
combination with a more fine grained `package-data`
configuration.

You can read more about "package data files" on setuptools documentation page:

-
<https://setuptools.pypa.io/en/latest/userguide/datafiles.html>

[^1]: For Python, any directory (with suitable naming) can be imported,
even if it does not contain any `.py` files.
On the other hand, currently there is no concept of
package data directory, all directories are treated like packages.

!!

```
check.warn(importable)
copying bin\spacy -> build\lib.win-amd64-cpython-310\bin
copying spacy\__init__.pxd -> build\lib.win-amd64-cpython-310\spacy
```

```

copying spacy\attrs.cpp -> build\lib.win-amd64-cpython-310\spacy
copying spacy\attrs.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\attrs.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\gold.cpp -> build\lib.win-amd64-cpython-310\spacy
copying spacy\gold.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\gold.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\kb.cpp -> build\lib.win-amd64-cpython-310\spacy
copying spacy\kb.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\kb.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\lexeme.cpp -> build\lib.win-amd64-cpython-310\spacy
copying spacy\lexeme.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\lexeme.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\morphology.cpp -> build\lib.win-amd64-cpython-310\
spacy
copying spacy\morphology.pxd -> build\lib.win-amd64-cpython-310\
spacy
copying spacy\morphology.pyx -> build\lib.win-amd64-cpython-310\
spacy
copying spacy\parts_of_speech.cpp -> build\lib.win-amd64-cpython-
310\spacy
copying spacy\parts_of_speech.pxd -> build\lib.win-amd64-cpython-
310\spacy
copying spacy\parts_of_speech.pyx -> build\lib.win-amd64-cpython-
310\spacy
copying spacy\strings.cpp -> build\lib.win-amd64-cpython-310\spacy
copying spacy\strings.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\strings.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\structs.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\symbols.cpp -> build\lib.win-amd64-cpython-310\spacy
copying spacy\symbols.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\symbols.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\tokenizer.cpp -> build\lib.win-amd64-cpython-310\spacy
copying spacy\tokenizer.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\tokenizer.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\typedefs.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\typedefs.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\vectors.cpp -> build\lib.win-amd64-cpython-310\spacy
copying spacy\vectors.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\vocab.cpp -> build\lib.win-amd64-cpython-310\spacy
copying spacy\vocab.pxd -> build\lib.win-amd64-cpython-310\spacy
copying spacy\vocab.pyx -> build\lib.win-amd64-cpython-310\spacy
copying spacy\matcher\dependencymatcher.cpp -> build\lib.win-amd64-
cpython-310\spacy\matcher
copying spacy\matcher\matcher.cpp -> build\lib.win-amd64-cpython-
310\spacy\matcher
copying spacy\matcher\phrasematcher.cpp -> build\lib.win-amd64-
cpython-310\spacy\matcher
copying spacy\pipeline\morphologizer.cpp -> build\lib.win-amd64-
cpython-310\spacy\pipeline
copying spacy\pipeline\pipes.cpp -> build\lib.win-amd64-cpython-310\

```

```
spacy\pipeline
  copying spacy\syntax\_beam_utils.cpp -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\_parser_model.cpp -> build\lib.win-amd64-
cpython-310\spacy\syntax
  copying spacy\syntax\_state.cpp -> build\lib.win-amd64-cpython-310\
spacy\syntax
  copying spacy\syntax\arc_eager.cpp -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\ner.cpp -> build\lib.win-amd64-cpython-310\
spacy\syntax
  copying spacy\syntax\nn_parser.cpp -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\nonproj.cpp -> build\lib.win-amd64-cpython-310\
spacy\syntax
  copying spacy\syntax\stateclass.cpp -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\transition_system.cpp -> build\lib.win-amd64-
cpython-310\spacy\syntax
  copying spacy\tokens\_retokenize.cpp -> build\lib.win-amd64-cpython-
310\spacy\tokens
  copying spacy\tokens\doc.cpp -> build\lib.win-amd64-cpython-310\
spacy\tokens
  copying spacy\tokens\morphanalysis.cpp -> build\lib.win-amd64-
cpython-310\spacy\tokens
  copying spacy\tokens\span.cpp -> build\lib.win-amd64-cpython-310\
spacy\tokens
  copying spacy\tokens\token.cpp -> build\lib.win-amd64-cpython-310\
spacy\tokens
  copying spacy\matcher\dependencymatcher.pyx -> build\lib.win-amd64-
cpython-310\spacy\matcher
  copying spacy\matcher\matcher.pxd -> build\lib.win-amd64-cpython-
310\spacy\matcher
  copying spacy\matcher\matcher.pyx -> build\lib.win-amd64-cpython-
310\spacy\matcher
  copying spacy\matcher\phrasematcher.pxd -> build\lib.win-amd64-
cpython-310\spacy\matcher
  copying spacy\matcher\phrasematcher.pyx -> build\lib.win-amd64-
cpython-310\spacy\matcher
  copying spacy\pipeline\morphologizer.pyx -> build\lib.win-amd64-
cpython-310\spacy\pipeline
  copying spacy\pipeline\pipes.pyx -> build\lib.win-amd64-cpython-310\
spacy\pipeline
  copying spacy\syntax\__init__.pxd -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\_beam_utils.pxd -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\_beam_utils.pyx -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\_parser_model.pxd -> build\lib.win-amd64-
```

```
cpython-310\spacy\syntax
  copying spacy\syntax\_parser_model.pyx -> build\lib.win-amd64-
cpython-310\spacy\syntax
  copying spacy\syntax\_state.pxd -> build\lib.win-amd64-cpython-310\
spacy\syntax
  copying spacy\syntax\_state.pyx -> build\lib.win-amd64-cpython-310\
spacy\syntax
  copying spacy\syntax\arc_eager.pxd -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\arc_eager.pyx -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\ner.pxd -> build\lib.win-amd64-cpython-310\
spacy\syntax
  copying spacy\syntax\ner.pyx -> build\lib.win-amd64-cpython-310\
spacy\syntax
  copying spacy\syntax\nn_parser.pxd -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\nn_parser.pyx -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\nonproj.pxd -> build\lib.win-amd64-cpython-310\
spacy\syntax
  copying spacy\syntax\nonproj.pyx -> build\lib.win-amd64-cpython-310\
spacy\syntax
  copying spacy\syntax\stateclass.pxd -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\stateclass.pyx -> build\lib.win-amd64-cpython-
310\spacy\syntax
  copying spacy\syntax\transition_system.pxd -> build\lib.win-amd64-
cpython-310\spacy\syntax
  copying spacy\syntax\transition_system.pyx -> build\lib.win-amd64-
cpython-310\spacy\syntax
  copying spacy\tokens\__init__.py -> build\lib.win-amd64-cpython-
310\spacy\tokens
  copying spacy\tokens\_retokenize.pyx -> build\lib.win-amd64-cpython-
310\spacy\tokens
  copying spacy\tokens\doc.pxd -> build\lib.win-amd64-cpython-310\
spacy\tokens
  copying spacy\tokens\doc.pyx -> build\lib.win-amd64-cpython-310\
spacy\tokens
  copying spacy\tokens\morphanalysis.pxd -> build\lib.win-amd64-
cpython-310\spacy\tokens
  copying spacy\tokens\morphanalysis.pyx -> build\lib.win-amd64-
cpython-310\spacy\tokens
  copying spacy\tokens\span.pxd -> build\lib.win-amd64-cpython-310\
spacy\tokens
  copying spacy\tokens\span.pyx -> build\lib.win-amd64-cpython-310\
spacy\tokens
  copying spacy\tokens\token.pxd -> build\lib.win-amd64-cpython-310\
spacy\tokens
  copying spacy\tokens\token.pyx -> build\lib.win-amd64-cpython-310\
```

```

spacy\tokens
  copying spacy\lang\hr\lemma_lookup_license.txt -> build\lib.win-
amd64-cpython-310\spacy\lang\hr
  copying spacy\lang\sr\lemma_lookup_licence.txt -> build\lib.win-
amd64-cpython-310\spacy\lang\sr
  copying spacy\tests\tokenizer\sun.txt -> build\lib.win-amd64-
cpython-310\spacy\tests\tokenizer
  running build_ext
  building 'spacy.parts_of_speech' extension
  creating build\temp.win-amd64-cpython-310
  creating build\temp.win-amd64-cpython-310\Release
  creating build\temp.win-amd64-cpython-310\Release\spacy
  "C:\Program Files\Microsoft Visual Studio\2022\Community\VC\Tools\
MSVC\14.33.31629\bin\HostX86\x64\cl.exe" /c /nologo /O2 /W3 /GL
/DNDEBUG /MD -IC:\Users\black\AppData\Local\Temp\pip-build-env-
waksoeh\overlay\Lib\site-packages\numpy\core\include -IC:\Users\
black\anaconda3\include -IC:\Users\black\AppData\Local\Temp\pip-
install-s20x3qrl\spacy_922dba4c6004426590d8919fc7fbb3fd\include -IC:\
Users\black\anaconda3\include -IC:\Users\black\anaconda3\Include "-
IC:\Program Files\Microsoft Visual Studio\2022\Community\VC\Tools\
MSVC\14.33.31629\include" "-IC:\Program Files\Microsoft Visual Studio\
2022\Community\VC\Tools\MSVC\14.33.31629\ATLMFC\include" "-IC:\Program
Files\Microsoft Visual Studio\2022\Community\VC\Auxiliary\VS\include"
"-IC:\Program Files (x86)\Windows Kits\10\include\10.0.19041.0\ucrt"
"-IC:\Program Files (x86)\Windows Kits\10\include\10.0.19041.0\um"
"-IC:\Program Files (x86)\Windows Kits\10\include\10.0.19041.0\
shared" "-IC:\Program Files (x86)\Windows Kits\10\include\
10.0.19041.0\winrt" "-IC:\Program Files (x86)\Windows Kits\10\
include\10.0.19041.0\cppwinrt" "-IC:\Program Files (x86)\Windows
Kits\NETFXSDK\4.8\include\um" /EHsc /Tpspace/parts_of_speech.cpp
/Fobuild\temp.win-amd64-cpython-310\Release\spacy/parts_of_speech.obj
/0x /EHsc
  parts_of_speech.cpp
  "C:\Program Files\Microsoft Visual Studio\2022\Community\VC\Tools\
MSVC\14.33.31629\bin\HostX86\x64\link.exe" /nologo /INCREMENTAL:NO
/LTCG /DLL /MANIFEST:EMBED,ID=2 /MANIFESTUAC:NO /LIBPATH:C:\Users\
black\anaconda3\libs /LIBPATH:C:\Users\black\anaconda3 /LIBPATH:C:\
Users\black\anaconda3\PCbuild\amd64 "/LIBPATH:C:\Program Files\
Microsoft Visual Studio\2022\Community\VC\Tools\MSVC\14.33.31629\
ATLMFC\lib\x64" "/LIBPATH:C:\Program Files\Microsoft Visual Studio\
2022\Community\VC\Tools\MSVC\14.33.31629\lib\x64" "/LIBPATH:C:\Program
Files (x86)\Windows Kits\NETFXSDK\4.8\lib\um\x64" "/LIBPATH:C:\Program
Files (x86)\Windows Kits\10\lib\10.0.19041.0\ucrt\x64" "/LIBPATH:C:\
Program Files (x86)\Windows Kits\10\lib\10.0.19041.0\um\x64"
/EXPORT:PyInit_parts_of_speech build\temp.win-amd64-cpython-310\
Release\spacy/parts_of_speech.obj /OUT:build\lib.win-amd64-cpython-
310\spacy\parts_of_speech.cp310-win_amd64.pyd /IMPLIB:build\temp.win-
amd64-cpython-310\Release\spacy\parts_of_speech.cp310-win_amd64.lib
  Creating library build\temp.win-amd64-cpython-310\Release\spacy\
parts_of_speech.cp310-win_amd64.lib and object build\temp.win-amd64-

```



```

cpython-310\Release\spacy\parts_of_speech.cp310-win_amd64.exp
  Generating code
  Finished generating code
  building 'spacy.strings' extension
  "C:\Program Files\Microsoft Visual Studio\2022\Community\VC\Tools\
MSVC\14.33.31629\bin\HostX86\x64\cl.exe" /c /nologo /O2 /W3 /GL
/DNDEBUG /MD -IC:\Users\black\AppData\Local\Temp\pip-build-env-
waksoeh\overlay\Lib\site-packages\numpy\core\include -IC:\Users\
black\anaconda3\include -IC:\Users\black\AppData\Local\Temp\pip-
install-s20x3qr1\spacy_922dba4c6004426590d8919fc7fbb3fd\include -IC:\
Users\black\anaconda3\include -IC:\Users\black\anaconda3\Include "-
IC:\Program Files\Microsoft Visual Studio\2022\Community\VC\Tools\
MSVC\14.33.31629\include" "-IC:\Program Files\Microsoft Visual Studio\
2022\Community\VC\Tools\MSVC\14.33.31629\ATLMFC\include" "-IC:\Program
Files\Microsoft Visual Studio\2022\Community\VC\Auxiliary\VS\include"
"-IC:\Program Files (x86)\Windows Kits\10\include\10.0.19041.0\ucrt"
"-IC:\Program Files (x86)\Windows Kits\10\include\10.0.19041.0\um"
"-IC:\Program Files (x86)\Windows Kits\10\include\10.0.19041.0\
shared" "-IC:\Program Files (x86)\Windows Kits\10\include\
10.0.19041.0\winrt" "-IC:\Program Files (x86)\Windows Kits\10\
include\10.0.19041.0\cppwinrt" "-IC:\Program Files (x86)\Windows
Kits\NETFXSDK\4.8\include\um" /EHsc /Tp spacy/strings.cpp /Fobuild\
temp.win-amd64-cpython-310\Release\spacy\strings.obj /Ox /EHsc
  strings.cpp
  spacy/strings.cpp(2153): warning C4244: 'argument': conversion from
'Py_ssize_t' to 'int', possible loss of data
  spacy/strings.cpp(2228): warning C4244: 'argument': conversion from
'Py_ssize_t' to 'int', possible loss of data
  spacy/strings.cpp(2769): warning C4244: '=': conversion from 'long'
to 'unsigned char', possible loss of data
  spacy/strings.cpp(3292): warning C4244: 'argument': conversion from
'Py_ssize_t' to 'int', possible loss of data
  spacy/strings.cpp(3863): warning C4244: 'argument': conversion from
'Py_ssize_t' to 'int', possible loss of data
  spacy/strings.cpp(3874): warning C4244: 'argument': conversion from
'Py_ssize_t' to 'int', possible loss of data
  spacy/strings.cpp(4265): warning C4244: 'argument': conversion from
'Py_ssize_t' to 'int', possible loss of data
  spacy/strings.cpp(4280): warning C4018: '<': signed/unsigned
mismatch
  spacy/strings.cpp(5923): warning C4244: 'argument': conversion from
'Py_ssize_t' to 'int', possible loss of data
  spacy/strings.cpp(10123): error C3861: '_PyGen_Send': identifier not
found
  spacy/strings.cpp(10128): error C3861: '_PyGen_Send': identifier not
found
  spacy/strings.cpp(10212): error C3861: '_PyGen_Send': identifier not
found
  error: command 'C:\\Program Files\\Microsoft Visual Studio\\2022\\
Community\\VC\\Tools\\MSVC\\14.33.31629\\bin\\HostX86\\x64\\cl.exe'

```

failed with exit code 2
[end of output]

note: This error originates from a subprocess, and is likely not a problem with pip.

ERROR: Failed building wheel for spacy
ERROR: Could not build wheels for spacy, which is required to install pyproject.toml-based projects

Collecting

https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-2.3.1/en_core_web_sm-2.3.1.tar.gz

Downloading

https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-2.3.1/en_core_web_sm-2.3.1.tar.gz (12.0 MB)

----- 12.0/12.0 MB 34.4 MB/s

eta 0:00:00

Preparing metadata (setup.py): started

Preparing metadata (setup.py): finished with status 'done'

Requirement already satisfied: spacy<2.4.0,>=2.3.0 in c:\users\black\anaconda3\lib\site-packages (from en-core-web-sm==2.3.1) (2.3.9)

Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (2.0.7)

Requirement already satisfied: srsly<1.1.0,>=1.0.2 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (1.0.6)

Requirement already satisfied: numpy>=1.15.0 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (1.23.5)

Requirement already satisfied: setuptools in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (65.6.3)

Requirement already satisfied: plac<1.2.0,>=0.9.6 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (1.1.3)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (1.0.9)

Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (1.0.2)

Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (0.10.1)

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (4.64.1)

Requirement already satisfied: thinc<7.5.0,>=7.4.1 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-

sm==2.3.1) (7.4.6)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (2.28.1)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (3.0.8)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in c:\users\black\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (0.7.9)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\black\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\black\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (1.26.14)
Requirement already satisfied: idna<4,>=2.5 in c:\users\black\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (3.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\black\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (2023.5.7)
Requirement already satisfied: colorama in c:\users\black\anaconda3\lib\site-packages (from tqdm<5.0.0,>=4.38.0->spacy<2.4.0,>=2.3.0->en-core-web-sm==2.3.1) (0.4.6)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: pyresparser in c:\users\black\anaconda3\lib\site-packages (1.0.6)
Requirement already satisfied: pytz>=2019.1 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (2022.7)
Requirement already satisfied: attrs>=19.1.0 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (22.1.0)
Requirement already satisfied: idna>=2.8 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (3.4)
Requirement already satisfied: blis>=0.2.4 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (0.7.9)
Requirement already satisfied: pyrsistent>=0.15.2 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (0.18.0)
Requirement already satisfied: nltk>=3.4.3 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (3.7)
Requirement already satisfied: jsonschema>=3.0.1 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (4.17.3)
Requirement already satisfied: certifi>=2019.6.16 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (2023.5.7)
Requirement already satisfied: requests>=2.22.0 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (2.28.1)
Requirement already satisfied: tqdm>=4.32.2 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (4.64.1)
Requirement already satisfied: python-dateutil>=2.8.0 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (2.8.2)

Requirement already satisfied: pdfminer.six>=20181108 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (20221105)

Requirement already satisfied: thinc>=7.0.4 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (7.4.6)

Requirement already satisfied: urllib3>=1.25.3 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (1.26.14)

Requirement already satisfied: pandas>=0.24.2 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (1.5.3)

Requirement already satisfied: preshed>=2.0.1 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (3.0.8)

Requirement already satisfied: srsly>=0.0.7 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (1.0.6)

Requirement already satisfied: pycryptodome>=3.8.2 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (3.18.0)

Requirement already satisfied: six>=1.12.0 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (1.16.0)

Requirement already satisfied: sortedcontainers>=2.1.0 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (2.4.0)

Requirement already satisfied: chardet>=3.0.4 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (4.0.0)

Requirement already satisfied: cymem>=2.0.2 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (2.0.7)

Requirement already satisfied: wasabi>=0.2.2 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (0.10.1)

Requirement already satisfied: numpy>=1.16.4 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (1.23.5)

Requirement already satisfied: spacy>=2.1.4 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (2.3.9)

Requirement already satisfied: docx2txt>=0.7 in c:\users\black\anaconda3\lib\site-packages (from pyresparser) (0.8)

Requirement already satisfied: joblib in c:\users\black\anaconda3\lib\site-packages (from nltk>=3.4.3->pyresparser) (1.1.1)

Requirement already satisfied: click in c:\users\black\anaconda3\lib\site-packages (from nltk>=3.4.3->pyresparser) (8.0.4)

Requirement already satisfied: regex>=2021.8.3 in c:\users\black\anaconda3\lib\site-packages (from nltk>=3.4.3->pyresparser) (2022.7.9)

Requirement already satisfied: cryptography>=36.0.0 in c:\users\black\anaconda3\lib\site-packages (from pdfminer.six>=20181108->pyresparser) (39.0.1)

Requirement already satisfied: charset-normalizer>=2.0.0 in c:\users\black\anaconda3\lib\site-packages (from pdfminer.six>=20181108->pyresparser) (2.0.4)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\black\anaconda3\lib\site-packages (from preshed>=2.0.1->pyresparser) (1.0.9)

Requirement already satisfied: setuptools in c:\users\black\anaconda3\lib\site-packages (from spacy>=2.1.4->pyresparser) (65.6.3)

Requirement already satisfied: plac<1.2.0,>=0.9.6 in c:\users\black\anaconda3\lib\site-packages (from spacy>=2.1.4->pyresparser) (1.1.3)

Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in c:\users\

```
black\anaconda3\lib\site-packages (from spacy>=2.1.4->pyresparser)
(1.0.2)
Requirement already satisfied: colorama in c:\users\black\anaconda3\
lib\site-packages (from tqdm>=4.32.2->pyresparser) (0.4.6)
Requirement already satisfied: cffi>=1.12 in c:\users\black\anaconda3\
lib\site-packages (from cryptography>=36.0.0->pdfminer.six>=20181108-
>pyresparser) (1.15.1)
Requirement already satisfied: pycparser in c:\users\black\anaconda3\
lib\site-packages (from cffi>=1.12->cryptography>=36.0.0-
>pdfminer.six>=20181108->pyresparser) (2.21)
Note: you may need to restart the kernel to use updated packages.
```

```
# Load the first 10 reviews
```

```
import json
f = open('yelp_academic_dataset_review.json')
js = []
for i in range(10):
    js.append(json.loads(f.readline()))
f.close()
review_df = pd.DataFrame(js)
review_df.shape
```

```
(10, 8)
```

```
review_df.head()
```

		votes	user_id \
0	{'funny': 0, 'useful': 5, 'cool': 2}	rLtl8ZkDX5vH5nAx9C3q5Q	
1	{'funny': 0, 'useful': 0, 'cool': 0}	0a2KyEL0d3Yb1V6aivbIuQ	
2	{'funny': 0, 'useful': 1, 'cool': 0}	0hT2KtfLiobPvh6cDC8JQg	
3	{'funny': 0, 'useful': 2, 'cool': 1}	uZetl9T0NcR0G0yFfughhg	
4	{'funny': 0, 'useful': 0, 'cool': 0}	vYmM4KTsC8ZfQBg-j5MWkw	

	review_id	stars	date \
0	fWKvX83p0-ka4JS3dc6E5A	5	2011-01-26
1	IjZ33sJrzXqU-0X6U8NwyA	5	2011-07-27
2	IESLBzqUCLdSzSqm0eCSxQ	4	2012-06-14
3	G-WvGaISbqqaMHlNnByodA	5	2010-05-27
4	1uJFq2r5QfJG_6ExMRCaGw	5	2012-01-05

	text	type \
0	My wife took me here on my birthday for breakf...	review
1	I have no idea why some people give bad review...	review
2	love the gyro plate. Rice is so good and I als...	review
3	Rosie, Dakota, and I LOVE Chaparral Dog Park!!...	review
4	General Manager Scott Petello is a good egg!!...	review

	business_id
0	9yKzy9PApeiPP0UJEtnvkg
1	ZRJwVLyzEJq1VAihDhYiow
2	6oRAC4uyJCSJl1X0WZpVSA

```
3 _1QQZuf4zZ0yFCvXc0o6Vg
4 6ozycU1RpktNG2-1BroVtw
```

```
# chunking in spaCy
```

```
import spacy
spacy.info('en_core_web_sm')
```

```
===== Info about model 'en_core_web_sm'
=====
```

```
lang          en
name          core_web_sm
license       MIT
author        Explosion
url           https://explosion.ai
email         contact@explosion.ai
description    English multi-task CNN trained on OntoNotes. Assigns
context-specific token vectors, POS tags, dependency parse and named
entities.
sources        [{'name': 'OntoNotes 5', 'url':
'https://catalog.ldc.upenn.edu/LDC2013T19', 'license': 'commercial
(licensed by Explosion)'}]
pipeline       ['tagger', 'parser', 'ner']
version        2.3.1
spacy_version  >=2.3.0,<2.4.0
parent_package spacy
labels         {'tagger': ['$', '"', "'", ',', '-', '-LRB-', '-RRB-', '.',
':', 'ADD', 'AFX', 'CC', 'CD', 'DT', 'EX', 'FW', 'HYPH', 'IN', 'JJ',
'JJR', 'JJS', 'LS', 'MD', 'NFP', 'NN', 'NNP', 'NNPS', 'NNS', 'PDT',
'POS', 'PRP', 'PRP$', 'RB', 'RBR', 'RBS', 'RP', 'SYM', 'TO', 'UH',
'VB', 'VBD', 'VBG', 'VBN', 'VBP', 'VBZ', 'WDT', 'WP', 'WP$', 'WRB',
'XX', '_SP', '`'], 'parser': ['ROOT', 'acl', 'acomp', 'advcl',
'advmod', 'agent', 'amod', 'appos', 'attr', 'aux', 'auxpass', 'case',
'cc', 'ccomp', 'compound', 'conj', 'csubj', 'csubjpass', 'dative',
'dep', 'det', 'dobj', 'expl', 'intj', 'mark', 'meta', 'neg', 'nmod',
'npadvmod', 'nsubj', 'nsubjpass', 'nummod', 'oprd', 'parataxis',
'pcomp', 'pobj', 'poss', 'preconj', 'predet', 'prep', 'prt', 'punct',
'quantmod', 'relcl', 'xcomp'], 'ner': ['CARDINAL', 'DATE', 'EVENT',
'FAC', 'GPE', 'LANGUAGE', 'LAW', 'LOC', 'MONEY', 'NORP', 'ORDINAL',
'ORG', 'PERCENT', 'PERSON', 'PRODUCT', 'QUANTITY', 'TIME',
'WORK_OF_ART']}
```

```
source        C:\Users\black\anaconda3\lib\site-packages\
en_core_web_sm
```

```
{'lang': 'en',
 'name': 'core_web_sm',
 'license': 'MIT',
 'author': 'Explosion',
```

```
'url': 'https://explosion.ai',
'email': 'contact@explosion.ai',
'description': 'English multi-task CNN trained on OntoNotes. Assigns
context-specific token vectors, POS tags, dependency parse and named
entities.',
'sources': [{'name': 'OntoNotes 5',
'url': 'https://catalog.ldc.upenn.edu/LDC2013T19',
'license': 'commercial (licensed by Explosion)'}],
'pipeline': ['tagger', 'parser', 'ner'],
'version': '2.3.1',
'spacy_version': '>=2.3.0,<2.4.0',
'parent_package': 'spacy',
'accuracy': {'las': 89.7572754092,
'uas': 91.6570115569,
'token_acc': 99.756964111,
'las_per_type': {'advmod': {'p': 85.6065101297,
'r': 84.9512113055,
'f': 85.2776018577},
'aux': {'p': 97.9464841319, 'r': 98.0772654442, 'f':
98.0118311613},
'nsubj': {'p': 95.530627567, 'r': 94.7522887555, 'f':
95.1398662913},
'root': {'p': 89.5162856958, 'r': 91.1692936754, 'f':
90.3352283866},
'compound': {'p': 90.4871122761, 'r': 92.2811316552, 'f':
91.3753170839},
'poss': {'p': 97.0346623923, 'r': 97.4838969404, 'f':
97.2587609198},
'case': {'p': 97.927972373, 'r': 99.3493493493, 'f':
98.6335403727},
'dobj': {'p': 92.4513496547, 'r': 93.8729981675, 'f':
93.1567503459},
'prep': {'p': 85.6642170718, 'r': 86.2427438631, 'f':
85.9525069954},
'pobj': {'p': 96.0694769711, 'r': 96.6428459243, 'f':
96.3553084873},
'relcl': {'p': 76.5768958186, 'r': 78.3538796229, 'f':
77.4551971326},
'det': {'p': 97.7105145232, 'r': 97.7901904024, 'f':
97.7503362269},
'amod': {'p': 91.5748754262, 'r': 90.4891480402, 'f':
91.0287743996},
'attr': {'p': 90.4294478528, 'r': 92.9772918419, 'f':
91.6856728177},
'cc': {'p': 83.8244137102, 'r': 83.3532647692, 'f': 83.5881753313},
'mark': {'p': 90.3421052632, 'r': 90.9644939057, 'f':
90.6522313177},
'nmod': {'p': 76.3772954925, 'r': 55.7586837294, 'f':
64.4593166608},
'conj': {'p': 76.8877867328, 'r': 78.0490874764, 'f':
```

77.4640849469},
 'advcl': {'p': 68.9203354298, 'r': 66.2301687232, 'f':
 67.548478233},
 'pcomp': {'p': 85.2515506547, 'r': 86.6246498599, 'f':
 85.9326154915},
 'nummod': {'p': 93.1951089846, 'r': 88.5353535354, 'f':
 90.8054908055},
 'nsubjpass': {'p': 92.4265842349, 'r': 92.0, 'f': 92.2127987664},
 'quantmod': {'p': 85.3463587922, 'r': 78.0666125102, 'f':
 81.5443360204},
 'auxpass': {'p': 94.7204968944, 'r': 97.2665148064, 'f':
 95.9766239604},
 'ccomp': {'p': 79.9260844194, 'r': 83.6863543788, 'f':
 81.7630086559},
 'npadvmod': {'p': 76.8636539204, 'r': 70.6927175844, 'f':
 73.6491487787},
 'appos': {'p': 70.0960219479, 'r': 66.5075921909, 'f':
 68.2546749777},
 'neg': {'p': 94.5082376435, 'r': 94.9824385349, 'f':
 94.7447447447},
 'xcomp': {'p': 88.2854100106, 'r': 89.2677674085, 'f':
 88.7738711405},
 'predet': {'p': 85.2459016393, 'r': 89.2703862661, 'f':
 87.2117400419},
 'acomp': {'p': 90.3553299492, 'r': 88.717716357, 'f':
 89.529035208},
 'acl': {'p': 75.6578947368, 'r': 69.012547736, 'f': 72.182596291},
 'oprd': {'p': 81.0810810811, 'r': 71.6417910448, 'f':
 76.0697305864},
 'dative': {'p': 73.9659367397, 'r': 69.7247706422, 'f':
 71.7827626919},
 'agent': {'p': 88.5328836425, 'r': 94.0860215054, 'f':
 91.2250217202},
 'meta': {'p': 94.7368421053, 'r': 34.615384615400004, 'f':
 50.7042253521},
 'dep': {'p': 40.329218107, 'r': 15.9090909091, 'f': 22.8172293364},
 'prt': {'p': 81.9166666667, 'r': 88.082437276, 'f': 84.8877374784},
 'expl': {'p': 98.3014861996, 'r': 99.1434689507, 'f':
 98.7206823028},
 'parataxis': {'p': 63.9240506329, 'r': 43.8177874187, 'f':
 51.9948519949},
 'intj': {'p': 69.387755102, 'r': 59.7802197802, 'f':
 64.2266824085},
 'csubj': {'p': 70.5882352941, 'r': 71.0059171598, 'f':
 70.796460177},
 'preconj': {'p': 57.4468085106, 'r': 62.7906976744, 'f': 60.0},
 'csubjpass': {'p': 44.4444444444, 'r': 66.6666666667, 'f':
 53.3333333333},
 'tags_acc': 97.056555292,
 'ents_f': 85.4306864065,


```
'ents_p': 85.7239322492,
'ents_r': 85.1394400045,
'ents_per_type': {'ORG': {'p': 83.3194096352,
'r': 82.8808864266,
'f': 83.0995695042},
'CARDINAL': {'p': 83.9554682384, 'r': 86.32996633, 'f':
85.1261620186},
'DATE': {'p': 84.6522781775, 'r': 86.1275705821, 'f':
85.3835521769},
'GPE': {'p': 92.5831202046, 'r': 90.2180685358, 'f':
91.3852950458},
'PERSON': {'p': 88.0239520958, 'r': 92.0908379013, 'f':
90.0114810563},
'MONEY': {'p': 92.9181929182, 'r': 91.4663461538, 'f':
92.1865536039},
'PRODUCT': {'p': 52.1276595745, 'r': 24.1379310345, 'f':
32.9966329966},
'TIME': {'p': 70.1886792453, 'r': 70.9923664122, 'f':
70.5882352941},
'PERCENT': {'p': 91.8566775244, 'r': 88.125, 'f': 89.95215311},
'WORK_OF_ART': {'p': 48.1481481481, 'r': 38.8059701493, 'f':
42.9752066116},
'QUANTITY': {'p': 78.1954887218, 'r': 65.4088050314, 'f':
71.2328767123},
'NORP': {'p': 88.682581786, 'r': 89.2348754448, 'f':
88.9578713969},
'LOC': {'p': 70.8154506438, 'r': 66.0, 'f': 68.3229813665},
'EVENT': {'p': 63.2911392405, 'r': 37.037037037, 'f':
46.7289719626},
'ORDINAL': {'p': 80.0, 'r': 83.9694656489, 'f': 81.9366852886},
'FAC': {'p': 34.8623853211, 'r': 46.9135802469, 'f': 40.0},
'LAW': {'p': 62.962962963, 'r': 56.6666666667, 'f': 59.649122807},
'LANGUAGE': {'p': 75.0, 'r': 65.2173913043, 'f': 69.7674418605}}},
'speed': {'cpu': 6107.3535050376, 'gpu': None, 'nwords': 291315},
'labels': {'tagger': ['$',
'',
'',
'',
'-LRB-',
'-RRB-',
'',
'',
'',
'',
'ADD',
'AFX',
'CC',
'CD',
'DT',
'EX',
'FW',
'HYPH',
'IN',
```

```
'JJ',
'JJR',
'JJS',
'LS',
'MD',
'NFP',
'NN',
'NNP',
'NNPS',
'NNS',
'PDT',
'POS',
'PRP',
'PRP$',
'RB',
'RBR',
'RBS',
'RP',
'SYM',
'TO',
'UH',
'VB',
'VBD',
'VBG',
'VBN',
'VBP',
'VBZ',
'WDT',
'WP',
'WP$',
'WRB',
'XX',
'_SP',
'\`',
],
'parser': ['ROOT',
'acl',
'acomp',
'advcl',
'advmod',
'agent',
'amod',
'appos',
'attr',
'aux',
'auxpass',
'case',
'cc',
'ccomp',
'compound',
'conj',
```

```
'csubj',
'csubjpass',
'dative',
'dep',
'det',
'dobj',
'expl',
'intj',
'mark',
'meta',
'neg',
'nmod',
'npadvmod',
'nsbj',
'nsbjpass',
'nummod',
'oprd',
'parataxis',
'pcomp',
'pobj',
'poss',
'preconj',
'predet',
'prep',
'prt',
'punct',
'quantmod',
'relcl',
'xcomp'],
'ner': ['CARDINAL',
'DATE',
'EVENT',
'FAC',
'GPE',
'LANGUAGE',
'LAW',
'LOC',
'MONEY',
'NORP',
'ORDINAL',
'ORG',
'PERCENT',
'PERSON',
'PRODUCT',
'QUANTITY',
'TIME',
'WORK_OF_ART']},
'source': 'C:\\Users\\black\\anaconda3\\lib\\site-packages\\
en_core_web_sm'}
```

```
nlp = spacy.load("en_core_web_sm")
doc_df = review_df['text'].apply(nlp)
type(doc_df)
```

```
pandas.core.series.Series
```

```
type(doc_df[0])
```

```
spacy.tokens.doc.Doc
```

```
doc_df[4]
```

General Manager Scott Petello is a good egg!!! Not to go into detail, but let me assure you if you have any issues (albeit rare) speak with Scott and treat the guy with some respect as you state your case and I'd be surprised if you don't walk out totally satisfied as I just did. Like I always say..... "Mistakes are inevitable, it's how we recover from them that is important"!!!

Thanks to Scott and his awesome staff. You've got a customer for life!! :^)

```
for doc in doc_df[4]:
    print(doc.text, doc.pos_, doc.tag_)
```

General PROPN NNP
Manager PROPN NNP
Scott PROPN NNP
Petello PROPN NNP
is AUX VBZ
a DET DT
good ADJ JJ
egg NOUN NN
! PUNCT .
! PUNCT .
! PUNCT .
Not PART RB
to PART TO
go VERB VB
into ADP IN
detail NOUN NN
, PUNCT ,
but CCONJ CC
let VERB VB
me PRON PRP
assure VERB VB
you PRON PRP
if SCONJ IN
you PRON PRP
have AUX VBP
any DET DT
issues NOUN NNS

(PUNCT -LRB-
albeit SCONJ IN
rare ADJ JJ
) PUNCT -RRB-
speak VERB VBP
with ADP IN
Scott PROPN NNP
and CCONJ CC
treat VERB VB
the DET DT
guy NOUN NN
with ADP IN
some DET DT
respect NOUN NN
as SCONJ IN
you PRON PRP
state VERB VBP
your DET PRP\$
case NOUN NN
and CCONJ CC
I PRON PRP
'd VERB MD
be AUX VB
surprised ADJ JJ
if SCONJ IN
you PRON PRP
do AUX VBP
n't PART RB
walk VERB VB
out ADP RP
totally ADV RB
satisfied ADJ JJ
as SCONJ IN
I PRON PRP
just ADV RB
did AUX VBD
. PUNCT .
Like INTJ UH
I PRON PRP
always ADV RB
say VERB VBP
..... PUNCT NFP
" PUNCT ``
Mistakes NOUN NNS
are AUX VBP
inevitable ADJ JJ
, PUNCT ,
it PRON PRP
's AUX VBZ
how ADV WRB

```
we PRON PRP
recover VERB VBP
from ADP IN
them PRON PRP
that DET WDT
is AUX VBZ
important ADJ JJ
" PUNCT ''
! PUNCT .
! PUNCT .
! PUNCT .
```

```
SPACE _SP
Thanks NOUN NNS
to ADP IN
Scott PROPN NNP
and CCONJ CC
his DET PRP$
awesome ADJ JJ
staff NOUN NN
. PUNCT .
You PRON PRP
've AUX VB
got VERB VBN
a DET DT
customer NOUN NN
for ADP IN
life NOUN NN
! PUNCT .
! PUNCT .
..... PUNCT NFP
: PUNCT :
^ PROPN NNP
) PUNCT -RRB-
```

```
# spaCy also does some basic noun chunking
print([chunk for chunk in doc_df[4].noun_chunks])
```

```
[General Manager Scott Petello, a good egg, detail, me, you, you, any
issues, Scott, the guy, some respect, you, your case, I, you, I, I,
Mistakes, it, we, them, Thanks, Scott, his awesome staff, You, a
customer, life, :^]
```

```
%pip install textblob
%pip install nltk
```

```
Requirement already satisfied: textblob in c:\users\black\anaconda3\
lib\site-packages (0.17.1)
Requirement already satisfied: nltk>=3.1 in c:\users\black\anaconda3\
lib\site-packages (from textblob) (3.7)
```

Requirement already satisfied: joblib in c:\users\black\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (1.1.1)
 Requirement already satisfied: regex>=2021.8.3 in c:\users\black\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (2022.7.9)
 Requirement already satisfied: click in c:\users\black\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (8.0.4)
 Requirement already satisfied: tqdm in c:\users\black\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (4.64.1)
 Requirement already satisfied: colorama in c:\users\black\anaconda3\lib\site-packages (from click->nltk>=3.1->textblob) (0.4.6)
 Note: you may need to restart the kernel to use updated packages.
 Requirement already satisfied: nltk in c:\users\black\anaconda3\lib\site-packages (3.7)
 Requirement already satisfied: click in c:\users\black\anaconda3\lib\site-packages (from nltk) (8.0.4)
 Requirement already satisfied: joblib in c:\users\black\anaconda3\lib\site-packages (from nltk) (1.1.1)
 Requirement already satisfied: tqdm in c:\users\black\anaconda3\lib\site-packages (from nltk) (4.64.1)
 Requirement already satisfied: regex>=2021.8.3 in c:\users\black\anaconda3\lib\site-packages (from nltk) (2022.7.9)
 Requirement already satisfied: colorama in c:\users\black\anaconda3\lib\site-packages (from click->nltk) (0.4.6)
 Note: you may need to restart the kernel to use updated packages.

chunking in textblob

```
from textblob import TextBlob
blob_df = review_df['text'].apply(TextBlob)
type(blob_df)
```

pandas.core.series.Series

```
type(blob_df[4])
```

textblob.blob.TextBlob

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

```
[nltk_data] Downloading collection 'all'
[nltk_data] |
[nltk_data] |   Downloading package abc to
[nltk_data] |       C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] |   Package abc is already up-to-date!
[nltk_data] |   Downloading package alpino to
[nltk_data] |       C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] |   Package alpino is already up-to-date!
[nltk_data] |   Downloading package averaged_perceptron_tagger to
[nltk_data] |       C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] |   Package averaged_perceptron_tagger is already up-
```

```

[nltk_data] | to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_ru to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_ru is already
[nltk_data] | up-to-date!
[nltk_data] | Downloading package basque_grammars to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package basque_grammars is already up-to-date!
[nltk_data] | Downloading package bcp47 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package bcp47 is already up-to-date!
[nltk_data] | Downloading package biocreative_ppi to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package biocreative_ppi is already up-to-date!
[nltk_data] | Downloading package bllip_wsj_no_aux to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package bllip_wsj_no_aux is already up-to-date!
[nltk_data] | Downloading package book_grammars to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package book_grammars is already up-to-date!
[nltk_data] | Downloading package brown to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package brown is already up-to-date!
[nltk_data] | Downloading package brown_tei to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package brown_tei is already up-to-date!
[nltk_data] | Downloading package cess_cat to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package cess_cat is already up-to-date!
[nltk_data] | Downloading package cess_esp to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package cess_esp is already up-to-date!
[nltk_data] | Downloading package chat80 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package chat80 is already up-to-date!
[nltk_data] | Downloading package city_database to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package city_database is already up-to-date!
[nltk_data] | Downloading package cmudict to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package comparative_sentences to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package comparative_sentences is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package comtrans to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package comtrans is already up-to-date!
[nltk_data] | Downloading package conll2000 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...

```



```
[nltk_data] | Package conll2000 is already up-to-date!
[nltk_data] | Downloading package conll2002 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package conll2002 is already up-to-date!
[nltk_data] | Downloading package conll2007 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package conll2007 is already up-to-date!
[nltk_data] | Downloading package crubadan to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package crubadan is already up-to-date!
[nltk_data] | Downloading package dependency_treebank to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package dependency_treebank is already up-to-date!
[nltk_data] | Downloading package dolch to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package dolch is already up-to-date!
[nltk_data] | Downloading package europarl_raw to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package europarl_raw is already up-to-date!
[nltk_data] | Downloading package extended_omw to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package extended_omw is already up-to-date!
[nltk_data] | Downloading package floresta to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package floresta is already up-to-date!
[nltk_data] | Downloading package framenet_v15 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package framenet_v15 is already up-to-date!
[nltk_data] | Downloading package framenet_v17 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package framenet_v17 is already up-to-date!
[nltk_data] | Downloading package gazetteers to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package gazetteers is already up-to-date!
[nltk_data] | Downloading package genesis to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package genesis is already up-to-date!
[nltk_data] | Downloading package gutenber to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package gutenber is already up-to-date!
[nltk_data] | Downloading package ieer to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package ieer is already up-to-date!
[nltk_data] | Downloading package inaugural to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package inaugural is already up-to-date!
[nltk_data] | Downloading package indian to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package indian is already up-to-date!
[nltk_data] | Downloading package jeita to
```

```

[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package jeita is already up-to-date!
[nltk_data] | Downloading package kimmo to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package kimmo is already up-to-date!
[nltk_data] | Downloading package knbc to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package knbc is already up-to-date!
[nltk_data] | Downloading package large_grammars to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package large_grammars is already up-to-date!
[nltk_data] | Downloading package lin_thesaurus to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package lin_thesaurus is already up-to-date!
[nltk_data] | Downloading package mac_morpho to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package mac_morpho is already up-to-date!
[nltk_data] | Downloading package machado to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package machado is already up-to-date!
[nltk_data] | Downloading package masc_tagged to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package masc_tagged is already up-to-date!
[nltk_data] | Downloading package maxent_ne_chunker to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package maxent_ne_chunker is already up-to-date!
[nltk_data] | Downloading package maxent_treebank_pos_tagger to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package maxent_treebank_pos_tagger is already up-
to-date!
[nltk_data] | Downloading package moses_sample to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package moses_sample is already up-to-date!
[nltk_data] | Downloading package movie_reviews to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...

[nltk_data] | Package movie_reviews is already up-to-date!
[nltk_data] | Downloading package mte_teip5 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package mte_teip5 is already up-to-date!
[nltk_data] | Downloading package mwa_ppdb to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package mwa_ppdb is already up-to-date!
[nltk_data] | Downloading package names to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package names is already up-to-date!
[nltk_data] | Downloading package nombank.1.0 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package nombank.1.0 is already up-to-date!
[nltk_data] | Downloading package nonbreaking_prefixes to

```

```

[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package nonbreaking_prefixes is already up-to-date!
[nltk_data] | Downloading package nps_chat to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package nps_chat is already up-to-date!
[nltk_data] | Downloading package omw to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package omw is already up-to-date!
[nltk_data] | Downloading package omw-1.4 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package omw-1.4 is already up-to-date!
[nltk_data] | Downloading package opinion_lexicon to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package opinion_lexicon is already up-to-date!
[nltk_data] | Downloading package panlex_swadesh to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package panlex_swadesh is already up-to-date!
[nltk_data] | Downloading package paradigms to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package paradigms is already up-to-date!
[nltk_data] | Downloading package pe08 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package pe08 is already up-to-date!
[nltk_data] | Downloading package perluniprops to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package perluniprops is already up-to-date!
[nltk_data] | Downloading package pil to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package pil is already up-to-date!
[nltk_data] | Downloading package pl196x to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package pl196x is already up-to-date!
[nltk_data] | Downloading package porter_test to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package porter_test is already up-to-date!
[nltk_data] | Downloading package ppattach to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package ppattach is already up-to-date!
[nltk_data] | Downloading package problem_reports to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package problem_reports is already up-to-date!
[nltk_data] | Downloading package product_reviews_1 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package product_reviews_1 is already up-to-date!
[nltk_data] | Downloading package product_reviews_2 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package product_reviews_2 is already up-to-date!
[nltk_data] | Downloading package propbank to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package propbank is already up-to-date!

```

```
[nltk_data] | Downloading package pros_cons to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package pros_cons is already up-to-date!
[nltk_data] | Downloading package ptb to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package ptb is already up-to-date!
[nltk_data] | Downloading package punkt to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package punkt is already up-to-date!
[nltk_data] | Downloading package qc to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package qc is already up-to-date!
[nltk_data] | Downloading package reuters to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package reuters is already up-to-date!
[nltk_data] | Downloading package rslp to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package rslp is already up-to-date!
[nltk_data] | Downloading package rte to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package rte is already up-to-date!
[nltk_data] | Downloading package sample_grammars to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package sample_grammars is already up-to-date!
[nltk_data] | Downloading package semcor to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package semcor is already up-to-date!
[nltk_data] | Downloading package senseval to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package senseval is already up-to-date!
[nltk_data] | Downloading package sentence_polarity to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package sentence_polarity is already up-to-date!
[nltk_data] | Downloading package sentiwordnet to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package sentiwordnet is already up-to-date!
[nltk_data] | Downloading package shakespeare to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package shakespeare is already up-to-date!
[nltk_data] | Downloading package sinica_treebank to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package sinica_treebank is already up-to-date!
[nltk_data] | Downloading package smultron to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package smultron is already up-to-date!
[nltk_data] | Downloading package snowball_data to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package snowball_data is already up-to-date!
[nltk_data] | Downloading package spanish_grammars to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
```

```

[nltk_data] | Package spanish_grammars is already up-to-date!
[nltk_data] | Downloading package state_union to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package state_union is already up-to-date!
[nltk_data] | Downloading package stopwords to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package stopwords is already up-to-date!
[nltk_data] | Downloading package subjectivity to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package subjectivity is already up-to-date!
[nltk_data] | Downloading package swadesh to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package swadesh is already up-to-date!
[nltk_data] | Downloading package switchboard to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package switchboard is already up-to-date!
[nltk_data] | Downloading package tagsets to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package tagsets is already up-to-date!
[nltk_data] | Downloading package timit to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package timit is already up-to-date!
[nltk_data] | Downloading package toolbox to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package toolbox is already up-to-date!
[nltk_data] | Downloading package treebank to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package treebank is already up-to-date!
[nltk_data] | Downloading package twitter_samples to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...

[nltk_data] | Package twitter_samples is already up-to-date!
[nltk_data] | Downloading package udhr to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package udhr is already up-to-date!
[nltk_data] | Downloading package udhr2 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package udhr2 is already up-to-date!
[nltk_data] | Downloading package unicode_samples to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package unicode_samples is already up-to-date!
[nltk_data] | Downloading package universal_tagset to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package universal_tagset is already up-to-date!
[nltk_data] | Downloading package universal_treebanks_v20 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package universal_treebanks_v20 is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package vader_lexicon to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...

```

```

[nltk_data] | Package vader_lexicon is already up-to-date!
[nltk_data] | Downloading package verbnet to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package verbnet is already up-to-date!
[nltk_data] | Downloading package verbnet3 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package verbnet3 is already up-to-date!
[nltk_data] | Downloading package webtext to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package webtext is already up-to-date!
[nltk_data] | Downloading package wmt15_eval to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package wmt15_eval is already up-to-date!
[nltk_data] | Downloading package word2vec_sample to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package word2vec_sample is already up-to-date!
[nltk_data] | Downloading package wordnet to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet is already up-to-date!
[nltk_data] | Downloading package wordnet2021 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet2021 is already up-to-date!
[nltk_data] | Downloading package wordnet2022 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet2022 is already up-to-date!
[nltk_data] | Downloading package wordnet31 to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet31 is already up-to-date!
[nltk_data] | Downloading package wordnet_ic to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package words is already up-to-date!
[nltk_data] | Downloading package ycoe to
[nltk_data] | C:\Users\black\AppData\Roaming\nltk_data...
[nltk_data] | Package ycoe is already up-to-date!
[nltk_data] | Done downloading collection all

```

True

blob_df[4].tags

```

[('General', 'NNP'),
 ('Manager', 'NNP'),
 ('Scott', 'NNP'),
 ('Petello', 'NNP'),
 ('is', 'VBZ'),
 ('a', 'DT'),
 ('good', 'JJ'),

```

('egg', 'NN'),
('Not', 'RB'),
('to', 'TO'),
('go', 'VB'),
('into', 'IN'),
('detail', 'NN'),
('but', 'CC'),
('let', 'VB'),
('me', 'PRP'),
('assure', 'VB'),
('you', 'PRP'),
('if', 'IN'),
('you', 'PRP'),
('have', 'VBP'),
('any', 'DT'),
('issues', 'NNS'),
('albeit', 'IN'),
('rare', 'NN'),
('speak', 'NN'),
('with', 'IN'),
('Scott', 'NNP'),
('and', 'CC'),
('treat', 'VB'),
('the', 'DT'),
('guy', 'NN'),
('with', 'IN'),
('some', 'DT'),
('respect', 'NN'),
('as', 'IN'),
('you', 'PRP'),
('state', 'NN'),
('your', 'PRP\$'),
('case', 'NN'),
('and', 'CC'),
('I', 'PRP'),
('d', 'MD'),
('be', 'VB'),
('surprised', 'VBN'),
('if', 'IN'),
('you', 'PRP'),
('do', 'VBP'),
('n't', 'RB'),
('walk', 'VB'),
('out', 'RP'),
('totally', 'RB'),
('satisfied', 'JJ'),
('as', 'IN'),
('I', 'PRP'),
('just', 'RB'),
('did', 'VBD'),

```
( 'Like', 'IN'),
( 'I', 'PRP'),
( 'always', 'RB'),
( 'say', 'VBP'),
( '.....', 'JJ'),
( 'Mistakes', 'NNS'),
( 'are', 'VBP'),
( 'inevitable', 'JJ'),
( 'it', 'PRP'),
( "'s", 'VBZ'),
( 'how', 'WRB'),
( 'we', 'PRP'),
( 'recover', 'VBP'),
( 'from', 'IN'),
( 'them', 'PRP'),
( 'that', 'WDT'),
( 'is', 'VBZ'),
( 'important', 'JJ'),
( 'Thanks', 'NNS'),
( 'to', 'TO'),
( 'Scott', 'NNP'),
( 'and', 'CC'),
( 'his', 'PRP$'),
( 'awesome', 'JJ'),
( 'stuff', 'NN'),
( 'You', 'PRP'),
( "'ve", 'VBP'),
( 'got', 'VBN'),
( 'a', 'DT'),
( 'customer', 'NN'),
( 'for', 'IN'),
( 'life', 'NN'),
( '.....', 'NN'),
( '^', 'NN')]
```

textblob can do some basic noun chunking

```
print([np for np in blob_df[4].noun_phrases])
```

```
['general manager', 'scott petello', 'good egg', 'scott', "n't walk",
'mistakes', 'thanks', 'scott', 'awesome stuff']
```

Task 3. Apply spacy and textblob chunking to the text used in task 1 respectively, and output the noun phrase chunking results

write your code here

use lem_text

Load the first 10 reviewsf

```
f = [lem_text] # use the lem_text
```

```
text_storage = pd.DataFrame(data=f, columns=['Text']) # assign to
dataframe for spacy and textblob
```

```
text_storage.head() # check dataframe
```


Text

0 my wife took me here on my birthday for brea...

```
spacy.info('en_core_web_sm') # load spacy
```

```
===== Info about model 'en_core_web_sm'
=====
```

```
lang          en
name          core_web_sm
license       MIT
author        Explosion
url           https://explosion.ai
email         contact@explosion.ai
description    English multi-task CNN trained on OntoNotes. Assigns
context-specific token vectors, POS tags, dependency parse and named
entities.
sources       [{'name': 'OntoNotes 5', 'url':
'https://catalog.ldc.upenn.edu/LDC2013T19', 'license': 'commercial
(licensed by Explosion)'}]
pipeline      ['tagger', 'parser', 'ner']
version       2.3.1
spacy_version >=2.3.0,<2.4.0
parent_package spacy
labels        {'tagger': ['$', '"', ',', '-', '-LRB-', '-RRB-', '.',
':', 'ADD', 'AFX', 'CC', 'CD', 'DT', 'EX', 'FW', 'HYPH', 'IN', 'JJ',
'JJR', 'JJS', 'LS', 'MD', 'NFP', 'NN', 'NNP', 'NNPS', 'NNS', 'PDT',
'POS', 'PRP', 'PRP$', 'RB', 'RBR', 'RBS', 'RP', 'SYM', 'TO', 'UH',
'VB', 'VBD', 'VBG', 'VBN', 'VBP', 'VBZ', 'WDT', 'WP', 'WP$', 'WRB',
'XX', 'SP', '`'], 'parser': ['ROOT', 'acl', 'acomp', 'advcl',
'advmod', 'agent', 'amod', 'appos', 'attr', 'aux', 'auxpass', 'case',
'cc', 'ccomp', 'compound', 'conj', 'csubj', 'csubjpass', 'dative',
'dep', 'det', 'dobj', 'expl', 'intj', 'mark', 'meta', 'neg', 'nmod',
'npadvmod', 'nsubj', 'nsubjpass', 'nummod', 'oprd', 'parataxis',
'pcomp', 'pobj', 'poss', 'preconj', 'predet', 'prep', 'prt', 'punct',
'quantmod', 'relcl', 'xcomp'], 'ner': ['CARDINAL', 'DATE', 'EVENT',
'FAC', 'GPE', 'LANGUAGE', 'LAW', 'LOC', 'MONEY', 'NORP', 'ORDINAL',
'ORG', 'PERCENT', 'PERSON', 'PRODUCT', 'QUANTITY', 'TIME',
'WORK_OF_ART']}]
source        C:\Users\black\anaconda3\lib\site-packages\
en_core_web_sm
```

```
{'lang': 'en',
 'name': 'core_web_sm',
 'license': 'MIT',
 'author': 'Explosion',
 'url': 'https://explosion.ai',
 'email': 'contact@explosion.ai',
```

```
'description': 'English multi-task CNN trained on OntoNotes. Assigns
context-specific token vectors, POS tags, dependency parse and named
entities.',
'sources': [{'name': 'OntoNotes 5',
'url': 'https://catalog.ldc.upenn.edu/LDC2013T19',
'license': 'commercial (licensed by Explosion)'}],
'pipeline': ['tagger', 'parser', 'ner'],
'version': '2.3.1',
'spacy_version': '>=2.3.0,<2.4.0',
'parent_package': 'spacy',
'accuracy': {'las': 89.7572754092,
'uas': 91.6570115569,
'token_acc': 99.756964111,
'las_per_type': {'advmod': {'p': 85.6065101297,
'r': 84.9512113055,
'f': 85.2776018577},
'aux': {'p': 97.9464841319, 'r': 98.0772654442, 'f':
98.0118311613},
'nsbj': {'p': 95.530627567, 'r': 94.7522887555, 'f':
95.1398662913},
'root': {'p': 89.5162856958, 'r': 91.1692936754, 'f':
90.3352283866},
'compound': {'p': 90.4871122761, 'r': 92.2811316552, 'f':
91.3753170839},
'poss': {'p': 97.0346623923, 'r': 97.4838969404, 'f':
97.2587609198},
'case': {'p': 97.927972373, 'r': 99.3493493493, 'f':
98.6335403727},
'dobj': {'p': 92.4513496547, 'r': 93.8729981675, 'f':
93.1567503459},
'prep': {'p': 85.6642170718, 'r': 86.2427438631, 'f':
85.9525069954},
'pobj': {'p': 96.0694769711, 'r': 96.6428459243, 'f':
96.3553084873},
'relcl': {'p': 76.5768958186, 'r': 78.3538796229, 'f':
77.4551971326},
'det': {'p': 97.7105145232, 'r': 97.7901904024, 'f':
97.7503362269},
'amod': {'p': 91.5748754262, 'r': 90.4891480402, 'f':
91.0287743996},
'attr': {'p': 90.4294478528, 'r': 92.9772918419, 'f':
91.6856728177},
'cc': {'p': 83.8244137102, 'r': 83.3532647692, 'f': 83.5881753313},
'mark': {'p': 90.3421052632, 'r': 90.9644939057, 'f':
90.6522313177},
'nmod': {'p': 76.3772954925, 'r': 55.7586837294, 'f':
64.4593166608},
'conj': {'p': 76.8877867328, 'r': 78.0490874764, 'f':
77.4640849469},
'advcl': {'p': 68.9203354298, 'r': 66.2301687232, 'f':
```

67.548478233},
 'pcomp': {'p': 85.2515506547, 'r': 86.6246498599, 'f':
 85.9326154915},
 'nummod': {'p': 93.1951089846, 'r': 88.5353535354, 'f':
 90.8054908055},
 'nsubjpass': {'p': 92.4265842349, 'r': 92.0, 'f': 92.2127987664},
 'quantmod': {'p': 85.3463587922, 'r': 78.0666125102, 'f':
 81.5443360204},
 'auxpass': {'p': 94.7204968944, 'r': 97.2665148064, 'f':
 95.9766239604},
 'ccomp': {'p': 79.9260844194, 'r': 83.6863543788, 'f':
 81.7630086559},
 'npadvmod': {'p': 76.8636539204, 'r': 70.6927175844, 'f':
 73.6491487787},
 'appos': {'p': 70.0960219479, 'r': 66.5075921909, 'f':
 68.2546749777},
 'neg': {'p': 94.5082376435, 'r': 94.9824385349, 'f':
 94.7447447447},
 'xcomp': {'p': 88.2854100106, 'r': 89.2677674085, 'f':
 88.7738711405},
 'predet': {'p': 85.2459016393, 'r': 89.2703862661, 'f':
 87.2117400419},
 'acomp': {'p': 90.3553299492, 'r': 88.717716357, 'f':
 89.529035208},
 'acl': {'p': 75.6578947368, 'r': 69.012547736, 'f': 72.182596291},
 'oprd': {'p': 81.0810810811, 'r': 71.6417910448, 'f':
 76.0697305864},
 'dative': {'p': 73.9659367397, 'r': 69.7247706422, 'f':
 71.7827626919},
 'agent': {'p': 88.5328836425, 'r': 94.0860215054, 'f':
 91.2250217202},
 'meta': {'p': 94.7368421053, 'r': 34.615384615400004, 'f':
 50.7042253521},
 'dep': {'p': 40.329218107, 'r': 15.9090909091, 'f': 22.8172293364},
 'prt': {'p': 81.9166666667, 'r': 88.082437276, 'f': 84.8877374784},
 'expl': {'p': 98.3014861996, 'r': 99.1434689507, 'f':
 98.7206823028},
 'parataxis': {'p': 63.9240506329, 'r': 43.8177874187, 'f':
 51.9948519949},
 'intj': {'p': 69.387755102, 'r': 59.7802197802, 'f':
 64.2266824085},
 'csubj': {'p': 70.5882352941, 'r': 71.0059171598, 'f':
 70.796460177},
 'preconj': {'p': 57.4468085106, 'r': 62.7906976744, 'f': 60.0},
 'csubjpass': {'p': 44.4444444444, 'r': 66.6666666667, 'f':
 53.3333333333},
 'tags_acc': 97.056555292,
 'ents_f': 85.4306864065,
 'ents_p': 85.7239322492,
 'ents_r': 85.1394400045,

```

    'ents_per_type': {'ORG': {'p': 83.3194096352,
    'r': 82.8808864266,
    'f': 83.0995695042},
    'CARDINAL': {'p': 83.9554682384, 'r': 86.32996633, 'f':
85.1261620186},
    'DATE': {'p': 84.6522781775, 'r': 86.1275705821, 'f':
85.3835521769},
    'GPE': {'p': 92.5831202046, 'r': 90.2180685358, 'f':
91.3852950458},
    'PERSON': {'p': 88.0239520958, 'r': 92.0908379013, 'f':
90.0114810563},
    'MONEY': {'p': 92.9181929182, 'r': 91.4663461538, 'f':
92.1865536039},
    'PRODUCT': {'p': 52.1276595745, 'r': 24.1379310345, 'f':
32.9966329966},
    'TIME': {'p': 70.1886792453, 'r': 70.9923664122, 'f':
70.5882352941},
    'PERCENT': {'p': 91.8566775244, 'r': 88.125, 'f': 89.95215311},
    'WORK_OF_ART': {'p': 48.1481481481, 'r': 38.8059701493, 'f':
42.9752066116},
    'QUANTITY': {'p': 78.1954887218, 'r': 65.4088050314, 'f':
71.2328767123},
    'NORP': {'p': 88.682581786, 'r': 89.2348754448, 'f':
88.9578713969},
    'LOC': {'p': 70.8154506438, 'r': 66.0, 'f': 68.3229813665},
    'EVENT': {'p': 63.2911392405, 'r': 37.037037037, 'f':
46.7289719626},
    'ORDINAL': {'p': 80.0, 'r': 83.9694656489, 'f': 81.9366852886},
    'FAC': {'p': 34.8623853211, 'r': 46.9135802469, 'f': 40.0},
    'LAW': {'p': 62.962962963, 'r': 56.6666666667, 'f': 59.649122807},
    'LANGUAGE': {'p': 75.0, 'r': 65.2173913043, 'f': 69.7674418605}}},
'speed': {'cpu': 6107.3535050376, 'gpu': None, 'nwords': 291315},
'labels': {'tagger': ['$',
'',
',',
'-LRB-',
'-RRB-',
':',
';',
'ADD',
'AFX',
'CC',
'CD',
'DT',
'EX',
'FW',
'HYPH',
'IN',
'JJ',
'JJR',

```

```
'JJS',
'LS',
'MD',
'NFP',
'NN',
'NNP',
'NNPS',
'NNS',
'PDT',
'POS',
'PRP',
'PRP$',
'RB',
'RBR',
'RBS',
'RP',
'SYM',
'TO',
'UH',
'VB',
'VBD',
'VBG',
'VBN',
'VBP',
'VBZ',
'WDT',
'WP',
'WP$',
'WRB',
'XX',
'_SP',
'\`',
],
'parser': ['ROOT',
'acl',
'acomp',
'advcl',
'advmod',
'agent',
'amod',
'appos',
'attr',
'aux',
'auxpass',
'case',
'cc',
'ccomp',
'compound',
'conj',
'csubj',
'csubjpass',
```

```
'dative',
'dep',
'det',
'dobj',
'expl',
'intj',
'mark',
'meta',
'neg',
'nmod',
'npadvmod',
'nsubj',
'nsubjpass',
'nummod',
'oprd',
'parataxis',
'pcomp',
'pobj',
'poss',
'preconj',
'predet',
'prep',
'prt',
'punct',
'quantmod',
'relcl',
'xcomp'],
'ner': ['CARDINAL',
'DATE',
'EVENT',
'FAC',
'GPE',
'LANGUAGE',
'LAW',
'LOC',
'MONEY',
'NORP',
'ORDINAL',
'ORG',
'PERCENT',
'PERSON',
'PRODUCT',
'QUANTITY',
'TIME',
'WORK_OF_ART']},
'source': 'C:\\Users\\black\\anaconda3\\lib\\site-packages\\
en_core_web_sm'}
```

```
nlp = spacy.load("en_core_web_sm") # load spacy
doc_df = text_storage['Text'].apply(nlp) # apply spacy as nlp
type(doc_df) # check
```

```
pandas.core.series.Series
```

```
type(doc_df[0]) # check
```

```
spacy.tokens.doc.Doc
```

```
blob_df[0].tags # get tags
```

```
[('My', 'PRP$'),
 ('wife', 'NN'),
 ('took', 'VBD'),
 ('me', 'PRP'),
 ('here', 'RB'),
 ('on', 'IN'),
 ('my', 'PRP$'),
 ('birthday', 'NN'),
 ('for', 'IN'),
 ('breakfast', 'NN'),
 ('and', 'CC'),
 ('it', 'PRP'),
 ('was', 'VBD'),
 ('excellent', 'JJ'),
 ('The', 'DT'),
 ('weather', 'NN'),
 ('was', 'VBD'),
 ('perfect', 'JJ'),
 ('which', 'WDT'),
 ('made', 'VBD'),
 ('sitting', 'VBG'),
 ('outside', 'IN'),
 ('overlooking', 'VBG'),
 ('their', 'PRP$'),
 ('grounds', 'NNS'),
 ('an', 'DT'),
 ('absolute', 'JJ'),
 ('pleasure', 'NN'),
 ('Our', 'PRP$'),
 ('waitress', 'NN'),
 ('was', 'VBD'),
 ('excellent', 'JJ'),
 ('and', 'CC'),
 ('our', 'PRP$'),
 ('food', 'NN'),
 ('arrived', 'VBD'),
 ('quickly', 'RB'),
 ('on', 'IN'),
 ('the', 'DT'),
```

('semi-busy', 'JJ'),
('Saturday', 'NNP'),
('morning', 'NN'),
('It', 'PRP'),
('looked', 'VBD'),
('like', 'IN'),
('the', 'DT'),
('place', 'NN'),
('fills', 'VBZ'),
('up', 'RP'),
('pretty', 'RB'),
('quickly', 'RB'),
('so', 'IN'),
('the', 'DT'),
('earlier', 'JJR'),
('you', 'PRP'),
('get', 'VBP'),
('here', 'RB'),
('the', 'DT'),
('better', 'JJR'),
('Do', 'VB'),
('yourself', 'PRP'),
('a', 'DT'),
('favor', 'NN'),
('and', 'CC'),
('get', 'VB'),
('their', 'PRP\$'),
('Bloody', 'NNP'),
('Mary', 'NNP'),
('It', 'PRP'),
('was', 'VBD'),
('phenomenal', 'JJ'),
('and', 'CC'),
('simply', 'RB'),
('the', 'DT'),
('best', 'JJS'),
('I', 'PRP'),
("'ve", 'VBP'),
('ever', 'RB'),
('had', 'VBN'),
('I', 'PRP'),
("'m", 'VBP'),
('pretty', 'JJ'),
('sure', 'NN'),
('they', 'PRP'),
('only', 'RB'),
('use', 'VBP'),
('ingredients', 'NNS'),
('from', 'IN'),
('their', 'PRP\$'),

('garden', 'NN'),
('and', 'CC'),
('blend', 'VB'),
('them', 'PRP'),
('fresh', 'JJ'),
('when', 'WRB'),
('you', 'PRP'),
('order', 'NN'),
('it', 'PRP'),
('It', 'PRP'),
('was', 'VBD'),
('amazing', 'VBG'),
('While', 'IN'),
('EVERYTHING', 'NN'),
('on', 'IN'),
('the', 'DT'),
('menu', 'NN'),
('looks', 'VBZ'),
('excellent', 'JJ'),
('I', 'PRP'),
('had', 'VBD'),
('the', 'DT'),
('white', 'JJ'),
('truffle', 'NN'),
('scrambled', 'VBD'),
('eggs', 'NNS'),
('vegetable', 'JJ'),
('skillet', 'NN'),
('and', 'CC'),
('it', 'PRP'),
('was', 'VBD'),
('tasty', 'JJ'),
('and', 'CC'),
('delicious', 'JJ'),
('It', 'PRP'),
('came', 'VBD'),
('with', 'IN'),
('2', 'CD'),
('pieces', 'NNS'),
('of', 'IN'),
('their', 'PRP\$'),
('griddled', 'VBN'),
('bread', 'NN'),
('with', 'IN'),
('was', 'VBD'),
('amazing', 'VBG'),
('and', 'CC'),
('it', 'PRP'),
('absolutely', 'RB'),
('made', 'VBD'),

```
('the', 'DT'),
('meal', 'JJ'),
('complete', 'JJ'),
('It', 'PRP'),
('was', 'VBD'),
('the', 'DT'),
('best', 'JJS'),
('toast', 'NN'),
('I', 'PRP'),
("'ve", 'VBP'),
('ever', 'RB'),
('had', 'VBN'),
('Anyway', 'RB'),
('I', 'PRP'),
('ca', 'MD'),
("n't", 'RB'),
('wait', 'VB'),
('to', 'TO'),
('go', 'VB'),
('back', 'RB')]
```

```
for d in doc_df[0]:
    print(d.text, d.pos_, d.tag_)
```

```
SPACE _SP
my DET PRP$
wife NOUN NN
took VERB VBD
me PRON PRP
here ADV RB
on ADP IN
my DET PRP$
birthday NOUN NN
for ADP IN
breakfast NOUN NN
and CCONJ CC
it PRON PRP
wa ADV RB
excel VERB VBP
the DET DT
weather NOUN NN
wa PROPN NNP
perfect ADJ JJ
which DET WDT
made VERB VBD
sit VERB VB
outsid ADV RB
overlook VERB VB
their DET PRP$
ground NOUN NN
an DET DT
```

absolut ADJ JJ
pleasur NOUN NN
our DET PRP\$
waitress NOUN NN
wa PROPN NNP
excel PROPN NNP
and CCONJ CC
our DET PRP\$
food NOUN NN
arriv PROPN NNP
quickli PROPN NNP
on ADP IN
the DET DT
semi ADJ JJ
- ADJ JJ
busi PROPN NNP
saturday PROPN NNP
morn VERB VBD
it PRON PRP
look VERB VB
like SCONJ IN
the DET DT
place NOUN NN
fill VERB VBP
up ADP RP
pretti PROPN NNP
quickli PROPN NNP
so SCONJ IN
the DET DT
earlier ADV RBR
you PRON PRP
get VERB VBP
here ADV RB
the DET DT
better ADJ JJR
do AUX VBP
yourself PRON PRP
a DET DT
favor NOUN NN
and CCONJ CC
get AUX VB
their DET PRP\$
bloodi NOUN NNS
mari PROPN NNP
it PRON PRP
wa VERB VBP
phenomen VERB VBZ
and CCONJ CC
simpli PROPN NNP
the DET DT

best ADJ JJS
i PRON PRP
ve VERB VB
ever ADV RB
had VERB VBN
i PRON PRP
m VERB VBP
pretti VERB VB
sure ADJ JJ
they PRON PRP
onli VERB VBP
use VERB VBP
ingredi NOUN NN
from ADP IN
their DET PRP\$
garden NOUN NN
and CCONJ CC
blend VERB VB
them PRON PRP
fresh ADJ JJ
when ADV WRB
you PRON PRP
order VERB VBP
it PRON PRP
it PRON PRP
wa VERB VBP
amaz VERB VBN
while SCONJ IN
everyth NOUN NN
on ADP IN
the DET DT
menu NOUN NN
look VERB VB
excel PROPN NNP
i PRON PRP
had AUX VBD
the DET DT
white PROPN NNP
truffl PROPN NNP
scrambl PROPN NNP
egg PROPN NNP
veget PROPN NNP
skillet PROPN NNP
and CCONJ CC
it PRON PRP
wa VERB VBP
tasti NOUN NN
and CCONJ CC
delici VERB VBZ
it PRON PRP

came VERB VBD
with ADP IN
two NUM CD
piec NOUN NNS
of ADP IN
their DET PRP\$
griddl NOUN NN
bread NOUN NN
with ADP IN
wa PROPN NNP
amaz PROPN NNP
and CCONJ CC
it PRON PRP
absolut ADJ JJ
made VERB VBD
the DET DT
meal NOUN NN
complet VERB VB
it PRON PRP
wa VERB VB
the DET DT
best ADJ JJS
toast NOUN NN
i PRON PRP
ve VERB VB
ever ADV RB
had VERB VBN
anyway INTJ UH
i PRON PRP
ca VERB MD
nt PART RB
wait VERB VB
to PART TO
go VERB VB
back ADV RB

spaCy also does some basic noun chunking

```
print([chunk for chunk in doc_df[0].noun_chunks]) # spacy noun chunking
```

```
[ my wife, me, my birthday, breakfast, it, the weather, wa, their  
ground, our waitress, wa excel, our food arriv quickli, the semi-busi  
saturday, it, the place, pretti quickli, you, yourself, a favor, their  
bloodi, it, simpli, i, i, they, ingredi, their garden, them, you, it,  
it, everyth, the menu, excel, i, the white truffl, scrambl egg veget  
skillet, it, it, two piec, their griddl bread, wa amaz, it, the meal,  
it, i, i]
```

textblob can do some basic noun chunking

```
print([np for np in blob_df[0].noun_phrases]) # textblob noun chunking
```

```
['absolute pleasure', 'place fills', 'bloody mary', 'everything',  
'white truffle', 'vegetable skillet', 'anyway', "ca n't"]
```

Overall, this task is interesting as it compares the noun chunking performed by spaCy with the noun chunking performed by textblob. Chunking is essentially grouping similar words which in this case could be a pair such as white truffle and scrambled eggs. Based on the results, textblob seemed to perform better as it captured more correct words whereas spaCy had a couple errors such as 'the white truffl'.

Question 3. Comparing the outputs of spacy and textblob chunking in task 3, which one would you like to use in your application? Explain it.

Answer to Q3: type your answer here

Question 4. Whats the disadvantage in bag of words . Please explain in your own words with an example .

Write code for the example.

Answer to Q4: type your answer here and code in next block