Введение в обработку естественного языка

Урок 5. Part-of-Speech разметка, NER, извлечение отношений

Тема "POS-tagger и NER"

Задание 1. Написать теггер на данных с русским языком

1. проверить UnigramTagger, BigramTagger, TrigramTagger и их комбмнации
2. написать свой теггер как на занятии, попробовать разные векторайзеры, добавить знание не только букв но и слов
3. сравнить все реализованные методы сделать выводы

Задание 2. Проверить насколько хорошо работает NER

данные брать из http://www.labinform.ru/pub/named_entities/ (http://www.labinform.ru/pub/named_entities/)

1. проверить NER из nltk/spacy/deeppavlov
2. написать свой нер попробовать разные подходы a. передаём в сетку токен и его соседей b. передаём в сетку только токен c. свой вариант
3. сравнить ваши реализованные подходы на качество (вывести precision/recall/f1_score)

Выполнил *Соковнин ИЛ*

```
В [ ]:
```

```
В [ ]:
```

```python
В [ ]:  # !chcp
        # !chcp 1251
        # !chcp 866
```
```
Active code page: 65001
```

```python
В [ ]:  # !mkdir datasets
        # !wget -O ./datasets/ru_syntagrus-ud-train.conllu https://raw.githubusercontent.com/UniversalDependencies/UD_Russian-
        # !wget -O ./datasets/ru_syntagrus-ud-dev.conllu https://raw.githubusercontent.com/UniversalDependencies/UD_Russian-Sy
```

```python
В [ ]:  # https://raw.githubusercontent.com/UniversalDependencies/UD_Russian-SynTagRus/master/
```

```
В [ ]:
```

```python
В [ ]:  # !wget -O ./datasets/ru_syntagrus-ud-train-a.conllu https://github.com/UniversalDependencies/UD_Russian-SynTagRus/blo
        # !wget -O ./datasets/ru_syntagrus-ud-train-b.conllu https://github.com/UniversalDependencies/UD_Russian-SynTagRus/blo
        # !wget -O ./datasets/ru_syntagrus-ud-train-c.conllu https://github.com/UniversalDependencies/UD_Russian-SynTagRus/blo

        # !wget -O ./datasets/ru_syntagrus-ud-dev.conllu https://raw.githubusercontent.com/UniversalDependencies/UD_Russian-Sy
```

```
В [ ]:
```

```python
В [ ]:  import pandas as pd
```

```python
В [ ]:  # path = 'en-ud-train.conllu'
        # path='./datasets/ru_syntagrus-ud-train-c.conllu'
        # df = conll_df(path, file_index=False)
```

```
В [ ]:
```

```
В [ ]:
```

```python
В [ ]:  import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt

        # from sklearn.feature_extraction.text import CountVectorizer, HashingVectorizer, TfidfVectorizer
        # from sklearn.preprocessing import LabelEncoder
        # from sklearn.linear_model import LogisticRegression
        # from sklearn.metrics import accuracy_score

        # from gensim.models import Word2Vec, FastText

        from nltk.tag import DefaultTagger
```

```python
from nltk.tag import UnigramTagger
from nltk.tag import BigramTagger, TrigramTagger
from nltk.tag import RegexpTagger
from nltk.corpus import names
import nltk
```
```
[nltk_data] Downloading package names to
[nltk_data]     C:\Users\sil\AppData\Roaming\nltk_data...
[nltk_data]   Package names is already up-to-date!
```

Out[11]: True

```python
# full_train_a = pyconll.load_from_file('datasets/ru_syntagrus-ud-train.conllu')
full_train_a = pyconll.load_from_file('datasets/ru_syntagrus-ud-train-a.conllu')
# full_train_b = pyconll.load_from_file('datasets/ru_syntagrus-ud-train-b.conllu')
# full_train_c = pyconll.load_from_file('datasets/ru_syntagrus-ud-train-c.conllu')
```

```python
full_train = full_train_a
# print(type(full_train_a))
```

```python
for sent in full_train[:2]:
    for token in sent:
        print(token.form, token.upos)
```
```
Анкета NOUN
. PUNCT

Начальник NOUN
областного ADJ
управления NOUN
связи NOUN
Семен PROPN
Еремеевич PROPN
был AUX
человек NOUN
простой ADJ
, PUNCT
приходил VERB
на ADP
работу NOUN
всегда ADV
вовремя ADV
, PUNCT
переехал VERB
```

Задание 1. Написать теггер на данных с русским языком

1. проверить UnigramTagger, BigramTagger, TrigramTagger и их комбинации

```python
fdata_train = []
for sent in full_train[:]:
    fdata_train.append([(token.form, token.upos) for token in sent])

fdata_test = []
for sent in full_test[:]:
    fdata_test.append([(token.form, token.upos) for token in sent])

fdata_sent_test = []
for sent in full_test[:]:
```

```python
MAX_SENT_LEN = max(len(sent) for sent in full_train)
MAX_ORIG_TOKEN_LEN = max(len(token.form) for sent in full_train for token in sent)
print('Наибольшая длина предложения', MAX_SENT_LEN)
```
```
Наибольшая длина предложения 194
Наибольшая длина токена 31
```

Out[17]:

```
[('Начальник', 'NOUN'),
 ('областного', 'ADJ'),
 ('управления', 'NOUN'),
 ('связи', 'NOUN'),
 ('Семен', 'PROPN'),
 ('Еремеевич', 'PROPN'),
 ('был', 'AUX'),
 ('человек', 'NOUN'),
 ('простой', 'ADJ'),
 (',', 'PUNCT'),
 ('приходил', 'VERB'),
 ('на', 'ADP'),
 ('работу', 'NOUN'),
 ('всегда', 'ADV'),
 ('вовремя', 'ADV'),
 (',', 'PUNCT'),
 ('здоровался', 'VERB'),
 ('с', 'ADP'),
 ('секретаршей', 'NOUN'),
 ('за', 'ADP'),
 ('руку', 'NOUN'),
 ('и', 'CCONJ'),
 ('иногда', 'ADV')
```

В [ ]:

Out[18]:
```
['Алгоритм',
 ',',
 'от',
 'имени',
 'учёного',
 'аль',
 '-',
 'Хорезми',
 ',',
 '-',
 'точный',
 'набор',
 'инструкций',
 ',',
 'описывающих',
 'порядок',
 'действий',
 'исполнителя',
 'для',
 'достижения',
 'результата',
 'решения',
 'задачи',
 'за',
 'конечное',
 'время',
 '.']
```

DefaultTagger

В [ ]:
```
default_tagger = nltk.DefaultTagger('NOUN')
display(default_tagger.tag(fdata_sent_test[0]), default_tagger.evaluate(fdata_test))
```

```
[('Алгоритм', 'NOUN'),
 ('.'. 'NOUN').
0.23593983983332248
```

UnigramTagger

```
B [ ]:  unigram_tagger = UnigramTagger(fdata_train)
```

```
[('Алгоритм', None),
 (',', 'PUNCT'),
 ('от', 'ADP'),
 ('имени', 'NOUN'),
 ('учёного', None),
 ('аль', None),
 ('-', 'PUNCT'),
 ('Хорезми', None),
 (',', 'PUNCT'),
 ('-', 'PUNCT'),
 ('точный', 'ADJ'),
 ('набор', 'NOUN'),
 ('инструкций', 'NOUN'),
 (',', 'PUNCT'),
 ('описывающих', 'VERB'),
 ('порядок', 'NOUN'),
 ('действий', 'NOUN'),
 ('исполнителя', None),
 ('для', 'ADP'),
 ('достижения', 'NOUN'),
 ('результата', 'NOUN'),
 ('решения', 'NOUN'),
 ('задачи', 'NOUN'),
 ('за', 'ADP'),
 ('конечное', None),
 ('время', 'NOUN'),
 ('.', 'PUNCT')]
0.823732013802982
```

BigramTagger

```
B [ ]:  bigram_tagger = BigramTagger(fdata_train, backoff=unigram_tagger)
```

```
[('Алгоритм', None),
 (',', 'PUNCT'),
 ('от', 'ADP'),
 ('имени', 'NOUN'),
 ('учёного', None),
 ('аль', None),
 ('-', 'PUNCT'),
 ('Хорезми', None),
 (',', 'PUNCT'),
 ('-', 'PUNCT'),
 ('точный', 'ADJ'),
 ('набор', 'NOUN'),
 ('инструкций', 'NOUN'),
 (',', 'PUNCT'),
 ('описывающих', 'VERB'),
 ('порядок', 'NOUN'),
 ('действий', 'NOUN'),
 ('исполнителя', None),
 ('для', 'ADP'),
 ('достижения', 'NOUN'),
 ('результата', 'NOUN'),
 ('решения', 'NOUN'),
 ('задачи', 'NOUN'),
 ('за', 'ADP'),
 ('конечное', None),
 ('время', 'NOUN'),
 ('.', 'PUNCT')]
0.8292792499511688
```

TrigramTagger

```
B [ ]:  trigram_tagger = TrigramTagger(fdata_train, backoff=bigram_tagger)
```

```
[('Алгоритм', None),
 (',', 'PUNCT'),
 ('от', 'ADP'),
 ('имени', 'NOUN'),
 ('учёного', None),
 ('аль', None),
 ('-', 'PUNCT'),
 ('Хорезми', None),
 (',', 'PUNCT'),
 ('-', 'PUNCT'),
 ('точный', 'ADJ'),
 ('набор', 'NOUN'),
 ('инструкций', 'NOUN'),
 (',', 'PUNCT'),
 ('описывающих', 'VERB'),
 ('порядок', 'NOUN'),
 ('действий', 'NOUN'),
 ('исполнителя', None),
 ('для', 'ADP'),
 ('достижения', 'NOUN'),
 ('результата', 'NOUN'),
 ('решения', 'NOUN'),
0.8291425222996289
```

Комбинация UnigramTagger, BigramTagger, TrigramTagger

```python
def backoff_tagger(train_sents, tagger_classes, backoff=None):
    for cls in tagger_classes:
        backoff = cls(train_sents, backoff=backoff)
    return backoff


backoff = DefaultTagger('NOUN')
tag = backoff_tagger(fdata_train,
                     [
                      UnigramTagger,
                      BigramTagger,
                      TrigramTagger
                     ],
                     backoff = backoff)

# display(tag[0], tag.evaluate(fdata_test))
```

Out[23]: 0.8787746598085813

Задание 1. Написать теггер на данных с русским языком

2. написать свой теггер как на занятии, попробовать разные векторайзеры, добавить знание не только букв но и слов

3. сравнить все реализованные методы сделать выводы

```python
from sklearn.feature_extraction.text import CountVectorizer, HashingVectorizer, TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#import xgboost as xgb
```

```python
train_tok = []
train_label = []
for sent in fdata_train[:]:
    for tok in sent:
        train_tok.append(tok[0])
        train_label.append('NO_TAG' if tok[1] is None else tok[1])

test_tok = []
test_label = []
for sent in fdata_test[:]:
    for tok in sent:
        test_tok.append(tok[0])
```

```python
le = LabelEncoder()
train_enc_labels = le.fit_transform(train_label)
test_enc_labels = le.transform(test_label)
```

Out[26]: array(['ADJ', 'ADP', 'ADV', 'AUX', 'CCONJ', 'DET', 'INTJ', 'NOUN',
       'NO_TAG', 'NUM', 'PART', 'PRON', 'PROPN', 'PUNCT', 'SCONJ', 'SYM',
       'VERB', 'X'], dtype='<U6')

```python
# если в test_tok есть None значения,
```

```python
# то X_test = hvectorizer.transform(test_tok) выдаём ошибку 'nonetype' object has no attribute 'lower'
# заполним None значения
print(test_tok[:10])
i=0
for text in test_tok:
    if text is None:
        test_tok[i] = 'None'
        print(f'{i}: {text}')
```

```
['Алгоритм', ',', 'от', 'имени', 'учёного', 'аль', '-', 'Хорезми', ',', '-']
118740: None
119181: None
120159: None
120865: None
122152: None
125686: None
126313: None
129117: None
129167: None
129381: None
129836: None
130272: None
130916: None
133264: None
135518: None
136561: None
136812: None
138306: None
138307: None
139000: None
139289: None
139294: None
139311: None
139340: None
139380: None
139547: None
139645: None
139711: None
139837: None
140486: None
140490: None
140604: None
141118: None
141155: None
141494: None
142673: None
142675: None
143423: None
143448: None
143756: None
144112: None
144128: None
144187: None
144269: None
144373: None
144857: None
145143: None
145195: None
145260: None
145637: None
145821: None
145822: None
146023: None
146272: None
148570: None
148680: None
149163: None
149190: None
149603: None
151107: None
153190: None
```

B [ ]:
```python
from sklearn.preprocessing import StandardScaler
# Standardize features by removing the mean and scaling to unit variance.
```

B [ ]:
```python
# data = [[0, 0], [0, 0], [1, 1], [1, 1]]
# print(data)
# scaler = StandardScaler()
# print(scaler.fit(data))
# print(scaler.mean_)
# print(scaler.transform(data))
```

B [ ]:
```python
for vectorizer in [CountVectorizer, HashingVectorizer, TfidfVectorizer]:
    # hvectorizer = HashingVectorizer(ngram_range=(1, 3), analyzer='char', n_features=50)
    hvectorizer = vectorizer(ngram_range=(1, 5), analyzer='char')
```

```
        std_scaler = StandardScaler(with_mean=False)

        X_train = hvectorizer.fit_transform(train_tok)
        X_test = hvectorizer.transform(test_tok)

        X_train = std_scaler.fit_transform(X_train)
        X_test = std_scaler.fit_transform(X_test)


        print('\n', X_train.shape)
        lr = LogisticRegression(random_state=0, max_iter=100, n_jobs=5)
        lr.fit(X_train, train_enc_labels)

        pred = lr.predict(X_test)
```

```
 (426182, 112218)
<class 'sklearn.feature_extraction.text.CountVectorizer'> 0.9295136402109513

 (426182, 1048576)
<class 'sklearn.feature_extraction.text.HashingVectorizer'> 0.9352171365323263

 (426182, 112218)
<class 'sklearn.feature_extraction.text.TfidfVectorizer'> 0.9361221433687089
```

```python
B [ ]: for vectorizer in [CountVectorizer, HashingVectorizer, TfidfVectorizer]:
           # hvectorizer = HashingVectorizer(ngram_range=(1, 3), analyzer='char', n_features=50)
           hvectorizer = vectorizer(ngram_range=(1, 5), analyzer='word')
           std_scaler = StandardScaler(with_mean=False)

           X_train = hvectorizer.fit_transform(train_tok)
           X_test = hvectorizer.transform(test_tok)

           X_train = std_scaler.fit_transform(X_train)
           X_test = std_scaler.fit_transform(X_test)


           print('\n', X_train.shape)
           lr = LogisticRegression(random_state=0, max_iter=100, n_jobs=5)
           lr.fit(X_train, train_enc_labels)

           pred = lr.predict(X_test)
```

```
 (426182, 62595)
<class 'sklearn.feature_extraction.text.CountVectorizer'> 0.7241031317143043

 (426182, 1048576)
<class 'sklearn.feature_extraction.text.HashingVectorizer'> 0.7337196432059379

 (426182, 62595)
<class 'sklearn.feature_extraction.text.TfidfVectorizer'> 0.7274106387134579
```

B [ ]:

Выводы:

1. nltk.tag лучший вариант: Комбинация из UnigramTagger BigramTagger TrigramTagger: 0.8787746598085813

2. Vectorizer лучший вариант: LogisticRegression над TfidfVectorizer при условии analyzer='char': 0.9360895891659613

Задание 2. Проверить насколько хорошо работает NER

данные брать из http://www.labinform.ru/pub/named_entities/ (http://www.labinform.ru/pub/named_entities/)

1. проверить NER из nltk/spacy/deeppavlov
2. написать свой нер попробовать разные подходы:
   a. передаём в сетку токен и его соседей
   b. передаём в сетку только токен
   c. свой вариант
3. сравнить ваши реализованные подходы на качество (вывести precision/recall/f1_score)

```
В [ ]: # РАЗМЕЧЕННЫЕ КОЛЛЕКЦИИ ДЛЯ ИЗВЛЕЧЕНИЯ ИМЕНОВАННЫХ СУЩНОСТЕЙ - http://www.labinform.ru/pub/named_entities/index.htm
```

```
В [88]: # !wget  -O ./datasets/collection5.zip http://www.labinform.ru/pub/named_entities/collection5.zip
```

--2022-06-09 13:36:08-- http://www.labinform.ru/pub/named_entities/collection5.zip (http://www.labinform.ru/pub/named_entities/collection5.zip)
Resolving www.labinform.ru (www.labinform.ru)... 95.181.230.181
Connecting to www.labinform.ru (www.labinform.ru)|95.181.230.181|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1899530 (1.8M) [application/zip]
Saving to: 'collection5.zip.1'

collection5.zip.1   100%[===================>]   1.81M 1.47MB/s   in 1.2s

2022-06-09 13:36:10 (1.47 MB/s) - 'collection5.zip.1' saved [1899530/1899530]

```
В [ ]:
```

'ГЄгй п Є®¤®ў п бва Ёж : 1251

```
В [89]:
```

Collection5/ collection5.zip collection5.zip.1 **sample_data/**

```
В [90]: import zipfile

        archive = 'collection5.zip'
        z = zipfile.ZipFile(archive)
        # z.infolist()
        directory_to_extract_to = ''
        with zipfile.ZipFile(archive, 'r') as zip_file:
```

```
В [91]:
```

Collection5 collection5.zip collection5.zip.1 sample_data

2.1 взять нер из nltk

```
В [92]:
```

Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://us-python.pkg.dev/colab-wheels/public/simple/ (https://us-python.pkg.dev/colab-wheels/public/simple/)
Requirement already satisfied: corus in /usr/local/lib/python3.7/dist-packages (0.9.0)

```
В [93]: import corus
```

```
В [94]: import nltk
        nltk.download('maxent_ne_chunker')
        nltk.download('words')
        nltk.download('punkt')
```

[nltk_data] Downloading package maxent_ne_chunker to
[nltk_data]   /root/nltk_data...
[nltk_data]   Package maxent_ne_chunker is already up-to-date!
[nltk_data] Downloading package words to /root/nltk_data...
[nltk_data]   Package words is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]   date!

Out[94]: True

```
В [95]: dir = 'Collection5/'
        records = load_ne5(dir)
```

```
В [96]:
```

Out[96]: '4 октября назначены очередные выборы Верховного Совета Аджарской АР\r\n\r\n\nПо распоряжению президента Грузии Михаила Саакашвили 4 октября 2008 года назначены очередные выборы Верховного Совета Аджарской АР. Об этом Новости-Грузия сообщили в пресс-службе администрации президента во вторник.\r\n\r\n\nВыборы Верховного совета Аджарской автономной республики назначены в соответствии с 241-ой статьей и 4-м пунктом 10-й статьи Конституционного закона Грузии <О статусе Аджарской автономной республики>.'

```
В [97]: {(' '.join(c[0] for c in chunk), chunk.label() ) for chunk in nltk.ne_chunk(nltk.pos_tag(nltk.word_tokenize(document))
```

Out[97]:

```
{('Верховного', 'ORGANIZATION'),
 ('Грузии', 'PERSON'),
 ('Грузии Михаила', 'PERSON'),
```

В [98]:

В [99]:

Out[99]: 'Путин освободил от должности торгпредов в Пакистане и Аргентине\r\n\r\nПремьер-министр России Владимир Путин освобод
ил от занимаемых должностей торгового представителя РФ в Пакистане Ришата Халикова, а также российского торгпреда в А
ргентине и Парагвае (по совместительству) Виктора Кокорева, сообщает сайт российского правительства в понедельник.\r\
n\r\nСогласно размещенным на сайте распоряжениям от 10 марта, Халиков должен сложить свои полномочия 23 марта этого г
ода, Кокорев - 2 апреля.\r\n\r\nОба торгпреда будут освобождены от должностей в связи с истечением срока действия их
срочных служебных контрактов.\r\n\r\nХаликов был назначен на должность торгпреда РФ в Пакистане в мае 2005 года. Ране
е он занимал пост заместителя министра внешнеэкономических связей и торговли Башкирии.\r\n\r\nКокорев занимал должнос
ть торгпреда в Аргентине и Парагвае с июля 2002 года.'

В [100]:
```python
{(' '.join(c[0] for c in chunk), chunk.label() ) for chunk in nltk.ne_chunk(nltk.pos_tag(nltk.word_tokenize(document))
```

Out[100]:
```
{('Виктора Кокорева', 'PERSON'),
 ('Кокорев', 'GPE'),
 ('Путин', 'PERSON'),
 ('Ришата Халикова', 'PERSON'),
 ('России Владимир Путин', 'PERSON'),
 ('Халиков', 'PERSON')}
```

2.2. проверить spacy

В [101]:

В [156]:

В [158]:

В [105]:
```
============================ Info about spaCy ============================

spaCy version    3.3.1
Location         /usr/local/lib/python3.7/dist-packages/spacy
Platform         Linux-5.4.188+-x86_64-with-Ubuntu-18.04-bionic
Python version   3.7.13
Pipelines        en_core_web_sm (3.3.0), ru_core_news_lg (3.3.0)
```

В [ ]:

В [ ]:
```python
import spacy
from spacy import displacy
```

В [106]:
```python
import requests
from bs4 import BeautifulSoup
import re

def url_to_string(url):
    res = requests.get(url)
    html = res.text
    soup = BeautifulSoup(html, 'html5lib')
    for script in soup(["script", "style", 'aside']):
        script.extract()
    return " ".join(re.split(r'[\n\t]+', soup.get_text()))

document = url_to_string('https://www.nytimes.com/2018/08/13/us/politics/peter-strzok-fired-fbi.html?hp&action=click&p

nltk.pos_tag(nltk.word_tokenize(document))
```

Out[106]:
```
[('nytimes.comPlease', 'NN'),
 ('enable', 'JJ'),
 ('JS', 'NNP'),
 ('and', 'CC'),
 ('disable', 'JJ'),
 ('any', 'DT'),
 ('ad', 'NN'),
 ('blocker', 'NN')]
```

В [177]:
```python
# nlp = en_core_web_md.load()
from spacy.lang.ru import Russian
nlp = Russian()
doc = nlp("Съешь ещё этих мягких французских булок, да выпей чаю.")
```

```python
print(doc)
for ent in doc:
Съешь ещё этих мягких французских булок, да выпей чаю.
Съешь
ещё
этих
мягких
французских
булок
,
да
выпей
чаю
.
```

B [183]:
```python
# from spacy import displacy
```

B [184]:

B [180]:
```python
# Russian spaCy Model Document - https://spacy.io/models/ru
import spacy
from spacy import displacy
import ru_core_news_lg

nlp = spacy.load('ru_core_news_lg')
doc = nlp("Съешь ещё этих мягких французских булок, да выпей чаю.")

print(doc, doc.ents)

for token in doc:
    print(token.text, token.pos_, token.dep_)

displacy.render(doc, jupyter=True, style='dep')
```

```
Съешь ещё этих мягких французских булок, да выпей чаю. ()
Съешь VERB ROOT
ещё ADV advmod
этих DET det
мягких ADJ amod
французских ADJ amod
булок NOUN obj
, PUNCT punct
да CCONJ cc
выпей PROPN conj
чаю PROPN flat:name
. PUNCT punct
```

Съешь VERB ещё ADV этих DET мягких ADJ французских ADJ булок, NOUN да CCONJ выпей PROPN чаю. PROPN advmod det amod amod obj cc conj flat:name

B [167]:
```python
# !python -m spacy download ru_core_news_sm
```

B [182]:
```python
import spacy
from spacy import displacy
from spacy.lang.ru.examples import sentences
import ru_core_news_sm

nlp = spacy.load("ru_core_news_sm")
doc = nlp("Съешь ещё этих мягких французских булок, да выпей чаю.")
print(doc.text)

for token in doc:
    print(token.text, token.pos_, token.dep_)

displacy.render(doc, jupyter=True, style='dep')
```

```
Съешь ещё этих мягких французских булок, да выпей чаю.
Съешь VERB ROOT
ещё ADV advmod
этих DET det
мягких ADJ amod
французских ADJ amod
булок NOUN obj
, PUNCT punct
да CCONJ cc
выпей VERB conj
чаю NOUN obj
. PUNCT punct
```

Съешь VERB ещё ADV этих DET мягких ADJ французских ADJ булок, NOUN да CCONJ выпей VERB чаю. NOUN advmod det amod amod obj cc conj obj

```python
import spacy
from spacy import displacy
import en_core_web_md

nlp = spacy.load('en_core_web_sm')
doc = nlp("Apple is looking at buying U.K. startup for 1$ billion")
displacy.render(doc, jupyter=True, style='ent')
for ent in doc.ents:
```

Apple **ORG** is looking at buying U.K. **GPE** startup for 1$ billion **MONEY**

```
Apple ORG
U.K. GPE
1$ billion MONEY
```

```python
import ru_core_news_lg

nlp = spacy.load('ru_core_news_lg')
doc = nlp("Съешь ещё этих мягких французских булок, да выпей чаю.")
displacy.render(doc, jupyter=True, style='ent')
for ent in doc.ents:
```

```
/usr/local/lib/python3.7/dist-packages/spacy/displacy/__init__.py:205: UserWarning: [W006] No entities to visualize f
ound in Doc object. If this is surprising to you, make sure the Doc was processed using a model that supports named e
ntity recognition, and check the `doc.ents` property manually if necessary.
  warnings.warn(Warnings.W006)
```

Съешь ещё этих мягких французских булок, да выпей чаю.

Out[65]: 'Съешь ещё этих мягких французских булок, да выпей чаю.'

```python
# !pip install spacytextblob
# !python -m textblob.download_corpora
```

```python
ny_bb =  'F.B.I. Agent Peter Strzok, Who Criticized Trump in Texts, Is Fired ' + \
'Give this article ' + \
'1.6K ' + \
'Peter Strzok, a top F.B.I. counterintelligence agent who was taken off the special counsel investigation after his di
'Peter Strzok, a top F.B.I. counterintelligence agent who was taken off the special counsel investigation after his di
'By Adam Goldman and Michael S. Schmidt ' + \
'Aug. 13, 2018 ' + \
'WASHINGTON — Peter Strzok, the F.B.I. senior counterintelligence agent who disparaged President Trump in inflammatory
'Mr. Trump and his allies seized on the texts — exchanged during the 2016 campaign with a former F.B.I. lawyer, Lisa P
'Along with writing the texts, Mr. Strzok was accused of sending a highly sensitive search warrant to his personal ema
'The F.B.I. had been under immense political pressure by Mr. Trump to dismiss Mr. Strzok, who was removed last summer
'Mr. Trump's victory traces back to June, when Mr. Strzok's conduct was laid out in a wide-ranging inspector general's
'But Mr. Strzok's lawyer said the deputy director of the F.B.I., David Bowdich, had overruled the Office of Profession
'A spokeswoman for the F.B.I. did not respond to a message seeking comment about why Mr. Strzok was dismissed rather t
'Editors' Picks ' + \
'13 Potluck Desserts You Can Make in a 9-by-13 Pan ' + \
'With Lifeguards Scarce, It's Beach Season That Needs Rescuing ' + \
''The Facemaker,' a Grisly but Inspiring Story of Surgery During World War I ' + \
'Aitan Goelman, Mr. Strzok's lawyer, denounced his client's dismissal. "The decision to fire Special Agent Strzok is n
'"This decision should be deeply troubling to all Americans," Mr. Goelman added. "A lengthy investigation and multiple
'Mr. Strzok's text exchanges with Ms. Page demonstrated animosity toward Mr. Trump. In one, Ms. Page asks: Trump is "n
'The report by the inspector general, Michael E. Horowitz, that preceded Mr. Strzok's firing not only criticized his c
'Mr. Horowitz said in his report that he was "deeply troubled" by the text messages. Hundreds exchanged over months we
'On Twitter, Mr. Strzok said he was "deeply saddened by this decision," adding, "It has been an honor to serve my coun
'Mr. Strzok became emblematic of Mr. Trump's unfounded assertions that a so-called deep state of bureaucrats opposed t
'After Mr. Horowitz uncovered the texts, Mr. Mueller, who had by then taken over the investigation, removed Mr. Strzok
'The inspector general's report also took issue with the reaction by Mr. Strzok and other F.B.I. officials to the disc
'At the time, Mr. Strzok was in the early stages of investigating whether any Trump associates had conspired with Russ
'"Under these circumstances, we did not have confidence that Strzok's decision to prioritize the Russia investigation
'The delays were merely the "result of bureaucratic snafus," Mr. Strzok's lawyer wrote last month in USA Today. ' + \
'But the justifications for the delay were "unpersuasive" and had "far-reaching consequences," the inspector general s
'In addition, the inspector general said that Mr. Strzok had forwarded a proposed search warrant to his personal email
'In a heated congressional hearing last month, Mr. Strzok expressed "significant regret" for the texts and rebutted th
'Mr. Strzok's dismissal was not unexpected. He is the second senior F.B.I. agent to be fired as a result of the inspec
'The firing was politically motivated, Mr. McCabe has said, as an effort to discredit him as a witness in the special
'Both men were fired before they were eligible for their pension and health benefits. '
'Mr. Strzok, 48, a graduate of Georgetown University, served as an officer in the Army before he joined the F.B.I. He
'He handled many important espionage cases including one involving a former C.I.A. officer suspected of working for Ch
```

Out[154]: 'Mr. Strzok, 48, a graduate of Georgetown University, served as an officer in the Army before he joined the F.B.I. He held several key positions in the F.B.I., eventually becoming a top deputy in the counterintelligence division. He handled many important espionage cases including one involving a former C.I.A. officer suspected of working for China and a group of Russian spies who had been working undercover in the United States.'

```
B [155]:   import spacy
           from spacy import displacy
           import en_core_web_md

           nlp = spacy.load('en_core_web_sm')
           # nlp = en_core_web_md.load()
           # ny_bb = url_to_string('https://www.nytimes.com/2018/08/13/us/politics/peter-strzok-fired-fbi.html?hp&action=click&pg


           article = nlp(ny_bb)
           print('\n!!!', article.ents, '!!!\n')
           displacy.render(article, jupyter=True, style='ent')
```

!!! (Peter Strzok, Texts, Peter Strzok, Trump, Peter Strzok, Trump, T.J. Kirkpatrick, The New York Times, Adam Goldman, Michael S. Schmidt, Aug. 13, 2018 WASHINGTON, — Peter Strzok, F.B.I., Trump, Hillary Clinton, Russia, Strzok, Monday, Trump, 2016, F.B.I., Lisa Page, Russia, Strzok, 20 years, the early months, Strzok, F.B.I., Trump, Strzok, last summer, Robert S. Mueller III, Strzok, Twitter, Monday, Trump, June, Strzok, F.B.I., Hillary Clinton's, 2016, Strzok, Office of Professional Responsibility, Strzok, 60 days, Strzok, House, July, Strzok, F.B.I., David Bowdich, the Office of Professional Responsibility, Strzok, Strzok, Strzok, Trump, Bowdich, F.B.I., Christopher A. Wray, 9, Pan With Lifeguards Scarce, World War I Aitan Goelman, Strzok, Strzok, Wray, Congress, Goelman, Americans, Goelman, Strzok, Page, Trump, one, Page, Strzok, Michael E. Horowitz, Strzok, Strzok, Clinton, just weeks, 2016, Horowitz, Hundreds, over months, Trump, Clinton, Strzok, FBI, Strzok, Trump, Trump, Strzok, Strzok, Strzok, Russia, Trump, Horowitz, Mueller, Strzok, F.B.I., Page, Mueller, May, Strzok, Clinton, Midyear Exam, late September 2016, Anthony D. Weiner, Clinton, Strzok, Trump, Russia, nearly a month, Weiner, Horowitz, Strzok, Clinton, Strzok, Russia, Midyear, Strzok, last month, USA, James B. Comey, F.B.I., Congress, Strzok, last month, Strzok, Russia, Strzok, second, March, Andrew G. McCabe, McCabe) !!!

F.B.I. Agent Peter Strzok **PERSON** , Who Criticized Trump in Texts **GPE** , Is Fired Give this article 1.6K Peter Strzok **PERSON** , a top F.B.I. counterintelligence agent who was taken off the special counsel investigation after his disparaging texts about President Trump **PERSON** were uncovered, was fired. Peter Strzok **PERSON** , a top F.B.I. counterintelligence agent who was taken off the special counsel investigation after his disparaging texts about President Trump **PERSON** were uncovered, was fired.Credit... T.J. Kirkpatrick **PERSON** for The New York Times **ORG** By Adam Goldman **PERSON** and Michael S. Schmidt **PERSON** Aug. 13, 2018 WASHINGTON **DATE** — Peter Strzok **PERSON** , the F.B.I. **ORG** senior counterintelligence agent who disparaged President Trump **PERSON** in inflammatory text messages and helped oversee the Hillary Clinton **PERSON** email and Russia **GPE** investigations, has been fired for violating bureau policies, Mr. Strzok **PERSON** 's lawyer said Monday **DATE** . Mr. Trump **PERSON** and his allies seized on the texts — exchanged during the 2016 **DATE** campaign with a former F.B.I. **ORG** lawyer, Lisa Page **PERSON** — in assailing the Russia **GPE** investigation as an illegitimate "witch hunt." Mr. Strzok **PERSON** , who rose over 20 years **DATE** at the F.B.I. to become one of its most experienced counterintelligence agents, was a key figure in the early months **DATE** of the inquiry. Along with writing the texts, Mr. Strzok **PERSON** was accused of sending a highly sensitive search warrant to his personal email account. The F.B.I. **ORG** had been under immense political pressure by Mr. Trump **PERSON** to dismiss Mr. Strzok **PERSON** , who was removed last summer **DATE** from the staff of the special counsel, Robert S. Mueller III **PERSON** . The president has repeatedly denounced Mr. Strzok **PERSON** in posts on Twitter **PERSON** , and on Monday **DATE** expressed satisfaction that he had been sacked. Mr. Trump **PERSON** 's victory traces back to June **DATE** , when Mr. Strzok **PERSON** 's conduct was laid out in a wide-ranging inspector general's report on how the F.B.I. **ORG** handled the investigation of Hillary Clinton's **PERSON** emails in the run-up to the 2016 **DATE** election. The report was critical of Mr. Strzok **PERSON** 's conduct in sending the texts, and the bureau's Office of Professional Responsibility **ORG** said that Mr. Strzok **PERSON** should be suspended for 60 days **DATE** and demoted. Mr. Strzok **PERSON** had testified before the House **ORG** in July **DATE** about how he had not allowed his political views to interfere with the investigations he was overseeing. But Mr. Strzok **PERSON** 's lawyer said the deputy director of the F.B.I. **ORG** , David Bowdich **PERSON** , had overruled the Office of Professional Responsibility **ORG** and fired Mr. Strzok **PERSON** . A spokeswoman for the F.B.I. did not respond to a message seeking comment about why Mr. Strzok **PERSON** was dismissed rather than demoted. Firing Mr. Strzok **PERSON** , however, removes a favorite target of Mr. Trump **PERSON** from the ranks of the F.B.I. and gives Mr. Bowdich **PERSON** and the F.B.I. **ORG** director, Christopher A. Wray **PERSON** , a chance to move beyond the president's ire. Editors' Picks 13 Potluck Desserts You Can Make in a 9 **CARDINAL** -by-13 Pan With Lifeguards Scarce **PRODUCT** , It's Beach Season That Needs Rescuing 'The Facemaker,' a Grisly but Inspiring Story of Surgery During World War I Aitan Goelman **EVENT** , Mr. Strzok **PERSON** 's lawyer, denounced his client's dismissal. "The decision to fire Special Agent Strzok **PERSON** is not only a departure from typical bureau

practice, but also contradicts Director Wray **PERSON** 's testimony to Congress **ORG** and his assurances that the F.B.I. intended to follow its regular process in this and all personnel matters," Mr. Goelman **PERSON** said. "This decision should be deeply troubling to all Americans **NORP** ," Mr. Goelman **PERSON** added. "A lengthy investigation and multiple rounds of congressional testimony failed to produce a shred of evidence that Special Agent Strzok's personal views ever affected his work." Mr. Strzok **PERSON** 's text exchanges with Ms. Page **PERSON** demonstrated animosity toward Mr. Trump **PERSON** . In one **CARDINAL** , Ms. Page **PERSON** asks: Trump is "not ever going to become president, right? Right?!" Mr. Strzok **PERSON** responds: "No. No he won't. We'll stop it." The inspector general, who uncovered the messages, found no evidence that the pair imposed their political views on their investigative decisions but cited that exchange as "not only indicative of a biased state of mind but, even more seriously, implies a willingness to take official action to impact the presidential candidate's electoral prospects." The report by the inspector general, Michael E. Horowitz **PERSON** , that preceded Mr. Strzok **PERSON** 's firing not only criticized his conduct in sending the texts but also his use of personal email accounts to handle sensitive information. In addition, the inspector general criticized Mr. Strzok **PERSON** 's decision not to move swiftly to examine new emails related to the Clinton **PERSON** investigation just weeks **DATE** before the 2016 **DATE** election. Mr. Horowitz **PERSON** said in his report that he was "deeply troubled" by the text messages. Hundreds **CARDINAL** exchanged over months **DATE** were found in which the pair disparaged Mr. Trump **PERSON** and, to a lesser extent, Mrs. Clinton **PERSON** , exchanged work gossip and bantered. On Twitter, Mr. Strzok **PERSON** said he was "deeply saddened by this decision," adding, "It has been an honor to serve my country and work with the fine men and women of the FBI **ORG** ." Mr. Strzok **PERSON** became emblematic of Mr. Trump **PERSON** 's unfounded assertions that a so-called deep state of bureaucrats opposed to him was undermining his presidency. Mr. Trump **PERSON** contended that Mr. Strzok **PERSON** targeted the president and accused Mr. Strzok **PERSON** of being "treasonous" and a "disgrace." Mr. Strzok **PERSON** told lawmakers that he never leaked information about the Russia **GPE** inquiry, which could have upended the election and hurt Mr. Trump **PERSON** 's chances of becoming president. After Mr. Horowitz **PERSON** uncovered the texts, Mr. Mueller **PERSON** , who had by then taken over the investigation, removed Mr. Strzok **PERSON** from his team. He was reassigned to the F.B.I. **ORG** 's human resources division. Ms. Page **PERSON** , who had left Mr. Mueller **PERSON** 's team before the discovery of the text messages, quit the F.B.I. in May **DATE** . The inspector general's report also took issue with the reaction by Mr. Strzok **PERSON** and other F.B.I. officials to the discovery of possible new evidence in the Clinton **PERSON** investigation, known internally as Midyear Exam **ORG** , in late September 2016 **DATE** on a laptop that belonged to the disgraced politician Anthony D. Weiner **PERSON** , the husband of a top Clinton **PERSON** aide. At the time, Mr. Strzok **PERSON** was in the early stages of investigating whether any Trump **ORG** associates had conspired with Russia **GPE** 's interference in the presidential election, and nearly a month **DATE** passed before agents and analysts began to act on the emails found on Mr. Weiner **PERSON** 's laptop. Mr. Horowitz **PERSON** could not rule out that Mr. Strzok **PERSON** had slow-walked the examination of the new emails to help Mrs. Clinton **PERSON** 's presidential bid. "Under these circumstances, we did not have confidence that Strzok **ORG** 's decision to prioritize the Russia **GPE** investigation over following up on the Midyear **DATE** -related investigative lead discovered on the Weiner laptop was free from bias," he wrote. The delays were merely the "result of bureaucratic snafus," Mr. Strzok **PERSON** 's lawyer wrote last month **DATE** in USA **GPE** Today. But the justifications for the delay were "unpersuasive" and had "far-reaching consequences," the inspector general said. James B. Comey **PERSON** , the former F.B.I. **ORG** director, has told investigators that if he had known about the emails earlier, it might have influenced his decision to alert Congress **ORG** to their existence days before the election. In addition, the inspector general said that Mr. Strzok **PERSON** had forwarded a proposed search warrant to his personal email account. The inspector general said the email, which included a draft of the search warrant affidavit, contained information that appeared to be under seal. In a heated congressional hearing last month **DATE** , Mr. Strzok **PERSON** expressed "significant regret" for the texts and rebutted the president's attacks on the Russia **GPE** inquiry. "This investigation is not politically motivated; it is not a witch hunt; it is not a hoax," he said. Mr. Strzok **PERSON** 's dismissal was not unexpected. He is the second **ORDINAL** senior F.B.I. agent to be fired as a result of the inspector general's investigation. In March **DATE** , Andrew G. McCabe **PERSON** , the former deputy director, was fired after the inspector general repeatedly faulted him for misleading investigators. The firing was politically motivated, Mr. McCabe **PERSON** has said, as an effort to discredit him as a witness in the special

counsel investigation. Both men were fired before they were eligible for their pension and health benefits.

2.3. проверить deeppavlov

B [204]:

B [205]:

B [206]:

B [207]:

B [209]:

B [202]:
```
import tensorflow as tf

tf.__version__
```

Out[202]: '2.8.2'

B [203]:
```
import tensorflow
# !python -m venv env
# #.\env\Scripts\activate.bat
# !pip install deeppavlov
# !python -m deeppavlov install squad_bert
#!python -m deeppavlov install ner_ontonotes
import fasttext
import deeppavlov
from deeppavlov import configs, build_model

deeppavlov_ner = build_model(configs.ner.ner_rus, download=True)
rus_document = "Нью-Йорк, США, 30 апреля 2020, 01:01 — REGNUM В администрации президента США Дональда Трампа планируют
deeppavlov_ner([rus_document])
```

```
2022-06-09 14:58:54.518 INFO in 'deeppavlov.download'['download'] at line 138: Skipped http://files.deeppavlov.ai/emb
eddings/lenta_lower_100.bin (http://files.deeppavlov.ai/embeddings/lenta_lower_100.bin) download because of matching
hashes
2022-06-09 14:58:55.479 INFO in 'deeppavlov.download'['download'] at line 138: Skipped http://files.deeppavlov.ai/dee
ppavlov_data/ner_rus_v3_cpu_compatible.tar.gz (http://files.deeppavlov.ai/deeppavlov_data/ner_rus_v3_cpu_compatible.t
ar.gz) download because of matching hashes
2022-06-09 14:58:55.754 INFO in 'deeppavlov.core.data.simple_vocab'['simple_vocab'] at line 115: [loading vocabulary
from /root/.deeppavlov/models/ner_rus/word.dict]
2022-06-09 14:58:55.829 INFO in 'deeppavlov.core.data.simple_vocab'['simple_vocab'] at line 115: [loading vocabulary
from /root/.deeppavlov/models/ner_rus/tag.dict]
2022-06-09 14:58:55.838 INFO in 'deeppavlov.core.data.simple_vocab'['simple_vocab'] at line 115: [loading vocabulary
from /root/.deeppavlov/models/ner_rus/char.dict]
2022-06-09 14:58:55.841 INFO in 'deeppavlov.models.embedders.fasttext_embedder'['fasttext_embedder'] at line 53: [loa
ding fastText embeddings from `/root/.deeppavlov/downloads/embeddings/lenta_lower_100.bin`]
Warning : `load_model` does not return WordVectorModel or SupervisedModel any more, but a `FastText` object which is
very similar.

---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
<ipython-input-203-edc252894614> in <module>()
```

2.3 написать свой нер попробовать разные подходы

- передаём в сетку токен и его соседей
- передаём в сетку только токен
- свой вариант

B [210]:
```
Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://us-python.pkg.dev/colab-wheels/public
/simple/ (https://us-python.pkg.dev/colab-wheels/public/simple/)
Collecting razdel
  Downloading razdel-0.5.0-py3-none-any.whl (21 kB)
Installing collected packages: razdel
Successfully installed razdel-0.5.0
```

B [213]:

B [217]:

B [218]:

B [219]:
```
records = corus.load_ne5(path)
words_docs = []
```

```python
for ix, rec in enumerate(records):
    words = []
    for token in tokenize(rec.text):

        result = 'None'

        for item in rec.spans:
            if (token.start >= item.start) and (token.stop <= item.stop) and (item.type == 'PER'):
                result = 'PER'
                break
            if (token.start >= item.start) and (token.stop <= item.stop) and (item.type == 'ORG'):
                result = 'ORG'
                break
            if (token.start >= item.start) and (token.stop <= item.stop) and (item.type == 'MEDIA'):
                result = 'MEDIA'
                break
            if (token.start >= item.start) and (token.stop <= item.stop) and (item.type == 'LOC'):
                result = 'LOC'
                break
            if (token.start >= item.start) and (token.stop <= item.stop) and (item.type == 'GEOPOLIT'):
                result = 'GEOPOLIT'
                break


        words.append([token.text, result])
```

B [273]:
```python
# !pip install pandas
!pip install pandas.io.formats.string
```

Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://us-python.pkg.dev/colab-wheels/public/simple/ (https://us-python.pkg.dev/colab-wheels/public/simple/)
ERROR: Could not find a version that satisfies the requirement pandas.io.formats.string (from versions: none)
ERROR: No matching distribution found for pandas.io.formats.string

B [267]:
```python
import pandas as pd
```

B [268]:
```python
type(words_docs)
# list(words_docs)
```

Out[268]:
```
[['4', 'None'],
 ['октября', 'None'],
 ['назначены', 'None'],
 ['очередные', 'None'],
 ['выборы', 'None'],
 ['Верховного', 'ORG'],
 ['Совета', 'ORG'],
 ['Аджарской', 'LOC'],
 ['АР', 'LOC'],
 ['По', 'None']]
```

B [270]:

B [257]:

Out[257]:
```
None        219214
PER          21200
ORG          13651
LOC           4568
GEOPOLIT      4356
MEDIA         2482
Name: tag, dtype: int64
```

B [258]:

Out[258]: (265471, 2)

B [259]:

Out[259]: pandas.core.frame.DataFrame

B [260]:

B [277]:

Out[277]:
```
0            4
1      октября
2    назначены
3    очередные
4       выборы
Name: word, dtype: object
```

```
B [278]:
```

```
Out[278]: 0    None
          1    None
          2    None
          3    None
          4    None
          Name: tag, dtype: object
```

```python
B [279]:  import tensorflow as tf
          from tensorflow.keras import Sequential
          from tensorflow.keras.layers import Dense, Embedding, GlobalAveragePooling1D, GlobalMaxPooling1D, Conv1D, GRU, LSTM, D
          from tensorflow.keras.layers.experimental.preprocessing import TextVectorization
```

```python
B [280]:  from sklearn import model_selection, preprocessing, linear_model

          train_x, valid_x, train_y, valid_y = model_selection.train_test_split(df_words['word'], df_words['tag'])

          # LabelEncode целевую переменную
          encoder = preprocessing.LabelEncoder()
          train_y = encoder.fit_transform(train_y)
```

```python
B [281]:  train_data = tf.data.Dataset.from_tensor_slices((train_x, train_y))
          valid_data = tf.data.Dataset.from_tensor_slices((valid_x, valid_y))

          train_data = train_data.batch(16)
```

```python
B [282]:  AUTOTUNE = tf.data.AUTOTUNE

          train_data = train_data.cache().prefetch(buffer_size=AUTOTUNE)
```

```python
B [283]:  def custom_standardization(input_data):
              return input_data

          vocab_size = 30000
          seq_len = 10

          vectorize_layer = TextVectorization(
              standardize=custom_standardization,
              max_tokens=vocab_size,
              output_mode='int',
              #ngrams=(1, 3),
              output_sequence_length=seq_len)

          # Make a text-only dataset (no labels) and call adapt to build the vocabulary.
          text_data = train_data.map(lambda x, y: x)
```

```
B [284]:
```

```
Out[284]: 29858
```

```python
B [285]:  embedding_dim = 64

          class modelNER(tf.keras.Model):
              def __init__(self):
                  super(modelNER, self).__init__()
                  self.emb = Embedding(vocab_size, embedding_dim)
                  self.gPool = GlobalMaxPooling1D()
                  self.fc1 = Dense(300, activation='relu')
                  self.fc2 = Dense(50, activation='relu')
                  self.fc3 = Dense(6, activation='softmax')

              def call(self, x):
                  x = vectorize_layer(x)
                  x = self.emb(x)
                  pool_x = self.gPool(x)

                  fc_x = self.fc1(pool_x)
                  fc_x = self.fc2(fc_x)

                  concat_x = tf.concat([pool_x, fc_x], axis=1)
                  prob = self.fc3(concat_x)
```

```
B [286]:
```

```python
B [287]:  mmodel.compile(optimizer='adam',
                  loss=tf.keras.losses.SparseCategoricalCrossentropy(),
```

```
B [288]:
```

```
Epoch 1/5
12444/12444 [==============================] - 327s 26ms/step - loss: 0.2968 - accuracy: 0.9134 - val_loss: 0.2296 -
val_accuracy: 0.9372
Epoch 2/5
12444/12444 [==============================] - 324s 26ms/step - loss: 0.1265 - accuracy: 0.9621 - val_loss: 0.2405 -
val_accuracy: 0.9399
Epoch 3/5
12444/12444 [==============================] - 315s 25ms/step - loss: 0.1097 - accuracy: 0.9653 - val_loss: 0.2290 -
val_accuracy: 0.9405
Epoch 4/5
12444/12444 [==============================] - 317s 26ms/step - loss: 0.1040 - accuracy: 0.9662 - val_loss: 0.2454 -
val_accuracy: 0.9402
Epoch 5/5
```
Out[288]: <keras.callbacks.History at 0x7f9814fa89d0>

B [ ]: