

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
им. Н.Э. Баумана

Кафедра «Систем обработки информации и управления»

**Лабораторная работа №5**  
по курсу «Методы машинного обучения»

ИСПОЛНИТЕЛЬНИЦА: Абросимова Н.Г.  
ИУ5-24М

\_\_\_\_\_  
подпись

"\_\_" \_\_\_\_\_ 2021 г.

ПРЕПОДАВАТЕЛЬ: \_\_\_\_\_  
ФИО

\_\_\_\_\_  
подпись

"\_\_" \_\_\_\_\_ 2021 г.

Москва - 2021

---

## Задание

Для произвольного предложения или текста решите следующие задачи:

1. Токенизация.
2. Частеречная разметка.
3. Лемматизация.
4. Выделение (распознавание) именованных сущностей.
5. Разбор предложения.

## Текст программы и экранные формы

```
from spacy.lang.ru import Russian
import spacy
text1 = 'Птица Говорун отличается умом и сообразительностью'
```

### Токенизация

```
nlp = spacy.load('ru_core_news_sm')
spacy_text1 = nlp(text1)
spacy_text1
Птица Говорун отличается умом и сообразительностью
for t in spacy_text1:
    print(t)
```

Птица  
Говорун  
отличается  
умом  
и  
сообразительностью

### Частеречная разметка

```
for token in spacy_text1:
    print('{} - {} - {}'.format(token.text, token.pos_, token.dep_))
```

Птица - NOUN - nsubj  
Говорун - PROPN - appos  
отличается - VERB - ROOT  
умом - NOUN - obl  
и - CCONJ - cc  
сообразительностью - NOUN - conj

### Лемматизация

```
for token in spacy_text1:
    print(token, token.lemma, token.lemma_)
```

Птица 14689663505855917165 птица  
Говорун 12845814657201599450 говорун  
отличается 11943280915295780619 отличать  
умом 5115711925887050639 ум  
и 15015917632809974589 и  
сообразительностью 3950228048489898853 сообразительность

### Выделение (распознавание) именованных сущностей

```
for ent in spacy_text1.ents:
    print(ent.text, ent.label_)
```

Говорун PER

```
from spacy import displacy
displacy.render(spacy_text1, style='ent', jupyter=True)
Разбор предложения
```

```
from spacy import displacy
```

```
displacy.render(spacy_text1, style='dep', jupyter=True)
```

```
obj: displacy.render(spacy_text1, style='dep', jupyter=True)
```

