## Московский государственный технический университет им. Н.Э. Баумана Кафедра «Системы обработки информации и управления»



## Лабораторная работа №5 по дисциплине «Методы машинного обучения»

Выполнил: студент группы ИУ5И-21М Мьоу 3о У **1. Цель лабораторной работы:** изучение методов предобработки текстов.

## Задание:

Для произвольного предложения или текста решите следующие задачи:

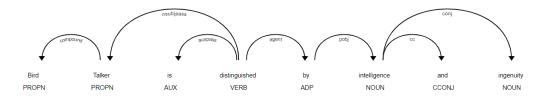
- 1. Токенизация.
- 2. Частеречная разметка.
- 3. Лемматизация.
- 4. Выделение (распознавание) именованных сущностей.
- 5. Разбор предложения.

## Текст программы и экранные формы

```
In [4]: text1 = 'Bird Talker is distinguished by intelligence and ingenuity'
         In [6]: nlp = spacy.load('en_core web sm')
                 spacy_text1 = nlp(text1)
                 spacy_text1
         Out[6]: Bird Talker is distinguished by intelligence and ingenuity
         In [7]: for t in spacy_text1:
                    print(t)
                Bird
                 Talker
                is
                distinguished
                 intelligence
                 and
                ingenuity
In [8]: for token in spacy text1:
              print('{} - {} '.format(token.text, token.pos_, token.dep_))
          Bird - PROPN - compound
          Talker - PROPN - nsubjpass
          is - AUX - auxpass
          distinguished - VERB - ROOT
          by - ADP - agent
          intelligence - NOUN - pobj
          and - CCONJ - cc
          ingenuity - NOUN - conj
 In [9]: for token in spacy_text1:
                print(token, token.lemma, token.lemma_)
          Bird 15457739884780427154 Bird
          Talker 17522498178439073006 Talker
          is 10382539506755952630 be
          distinguished 1025273512777152759 distinguish
          by 16764210730586636600 by
          intelligence 11044490816763727375 intelligence
          and 2283656566040971221 and
          ingenuity 6309617539516521249 ingenuity
In [10]: for ent in spacy text1.ents:
              print(ent.text, ent.label_)
          Bird Talker PERSON
```

In [3]: from spacy.lang.en import English

import spacy



**←**