Московский государственный технический университет им. Н.Э. Баумана

Кафедра «Системы обработки информации и управления»

****

Лабораторная работа №8

по дисциплине

«Методы машинного обучения»

на тему

## **«Предобработка текста»**

Выполнил:

студент Люй Чжэ

группы ИУ5И-23М

Москва — 2024 г.

**1. Цель лабораторной работы**

изучение методов предобработки текстов.

**2. Задание**

Для произвольного предложения или текста решите следующие задачи:

1.Токенизация.

2.Частеречная разметка.

3.Лемматизация.

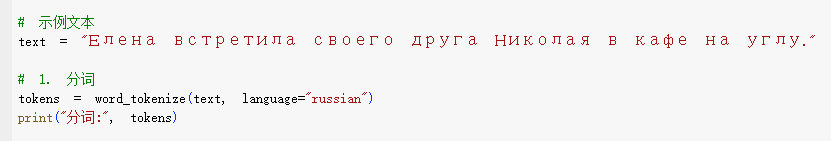
4.Выделение (распознавание) именованных сущностей.

5.Разбор предложения.

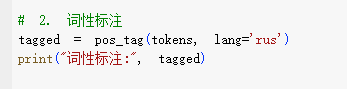
**3. Ход выполнения работы**

Я выбрал библиотеку nltk и spacy.

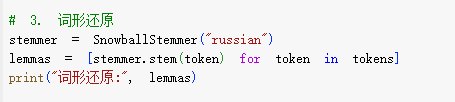
**Шаг 1: Токенизация:** Разбивает текст на отдельные слова (токены).



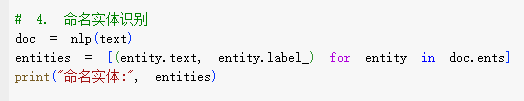
**Шаг 2: Частеречная разметка:** Присваивает каждому слову соответствующую часть речи.



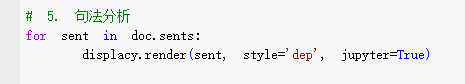
**Шаг 3: Преобразует слова к их базовой форме (лемме).**



**Шаг 4: Выделение именованных сущностей:** Определяет и категоризирует именованные сущности.



**Шаг 5: Разбор предложения:** Показывает структуру предложения, включая зависимости между словами.



**Шаг 6: Визуализация процессов**

