Московский государственный технический университет

им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Системы обработки информации и управления»



**Лабораторная работа № 7**

**По курсу «методы машинного обучения в АСОИУ»**

**«Предобработка текста»**

**Выполнила:**

студент ИУ5-24М

Сорокин А.Э.

**Проверил:**

Балашов А.М.

Подпись:

23.05.2024

Москва, 2024

**Описание задания**

Для произвольного предложения или текста решите следующие задачи:

* Токенизация.
* Частеречная разметка.
* Лемматизация.
* Выделение (распознавание) именованных сущностей.
* Разбор предложения.

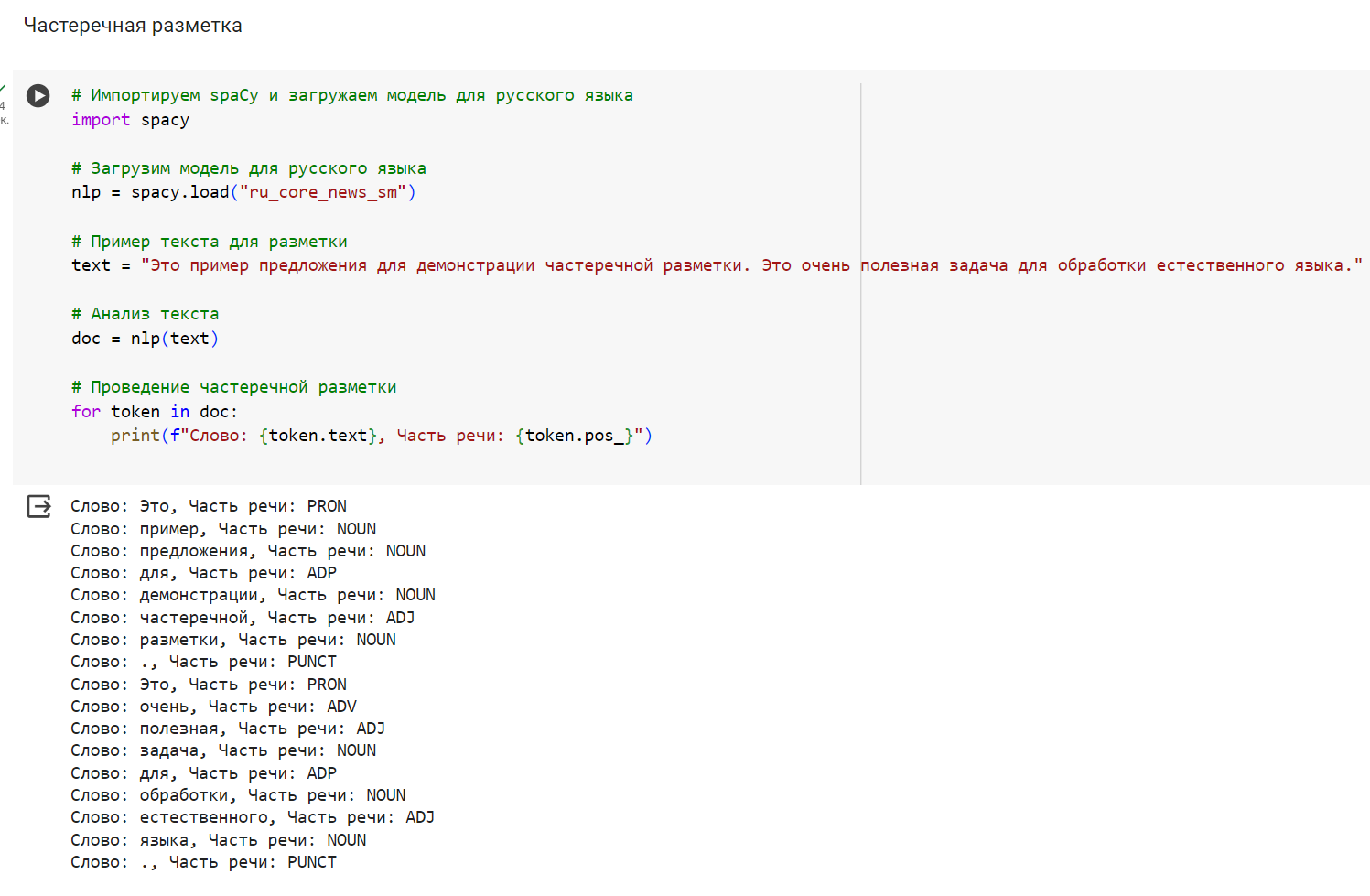
**Описание задач**

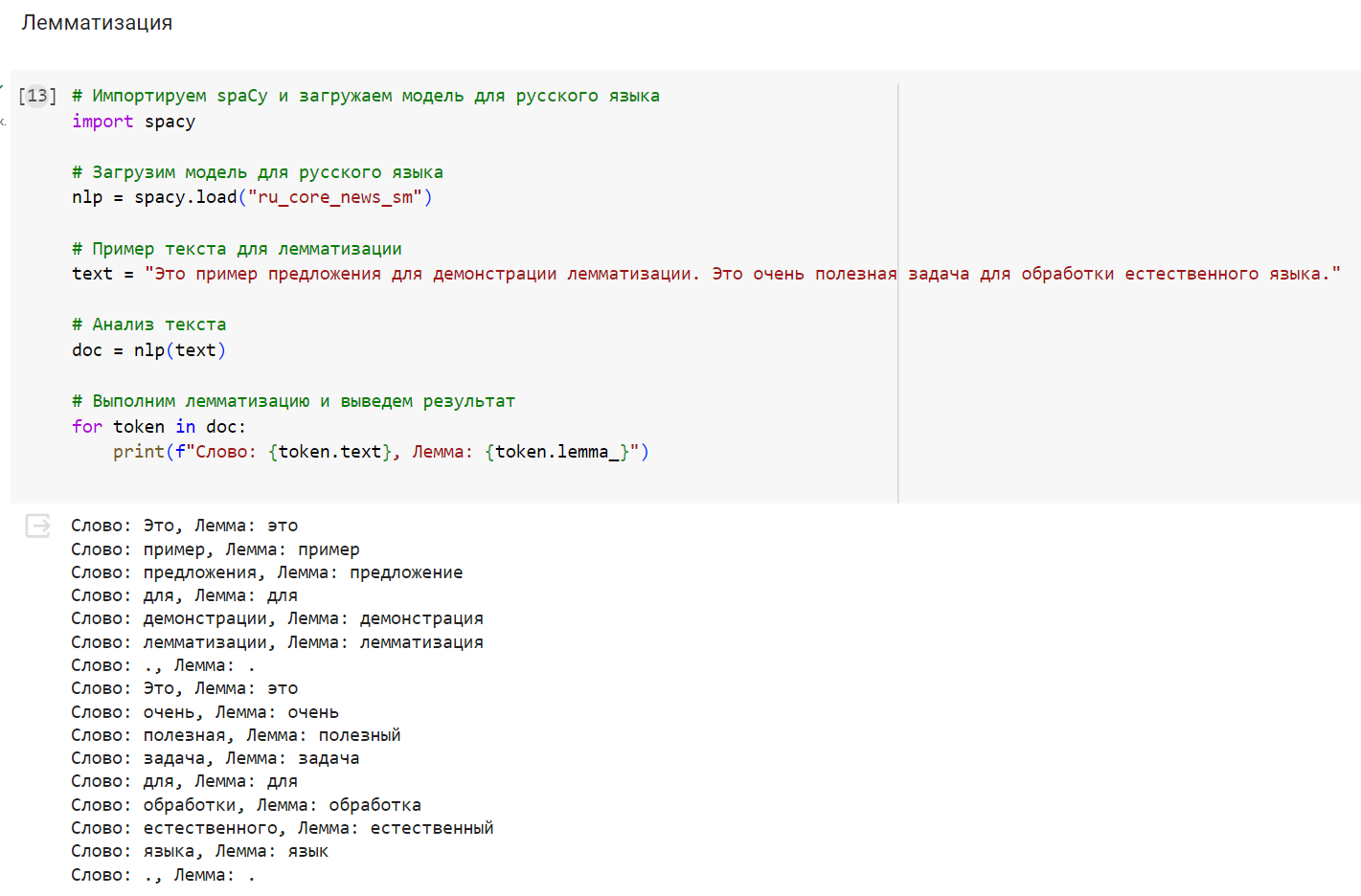
Частеречная разметка текста (также известная как POS-tagging или Part-of-Speech tagging) — это процесс определения грамматических категорий (частей речи) слов в тексте. В качестве таких категорий могут выступать существительные, глаголы, прилагательные и так далее.

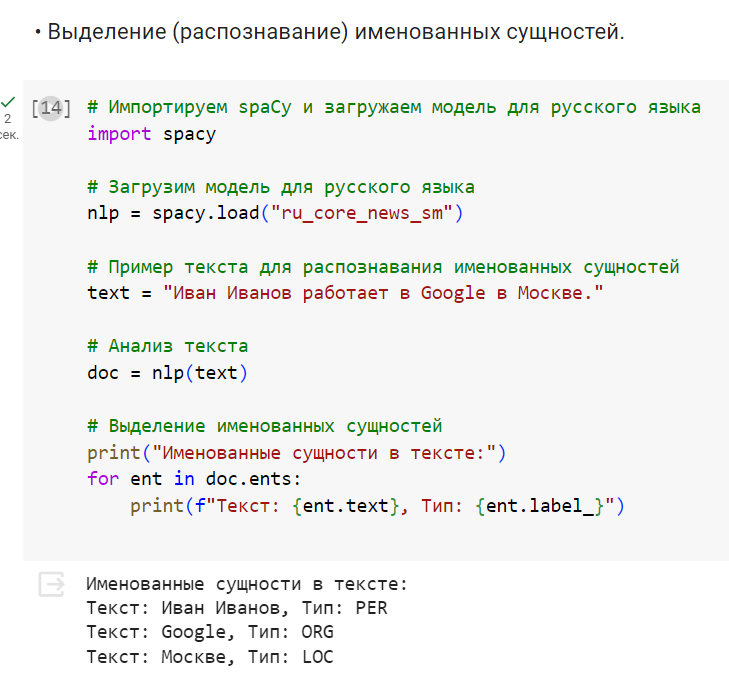
Выделение (распознавание) именованных сущностей (NER, Named Entity Recognition) — это процесс распознавания и классификации слов или фраз в тексте в определенные категории, такие как имена людей, названия мест, организации и так далее.

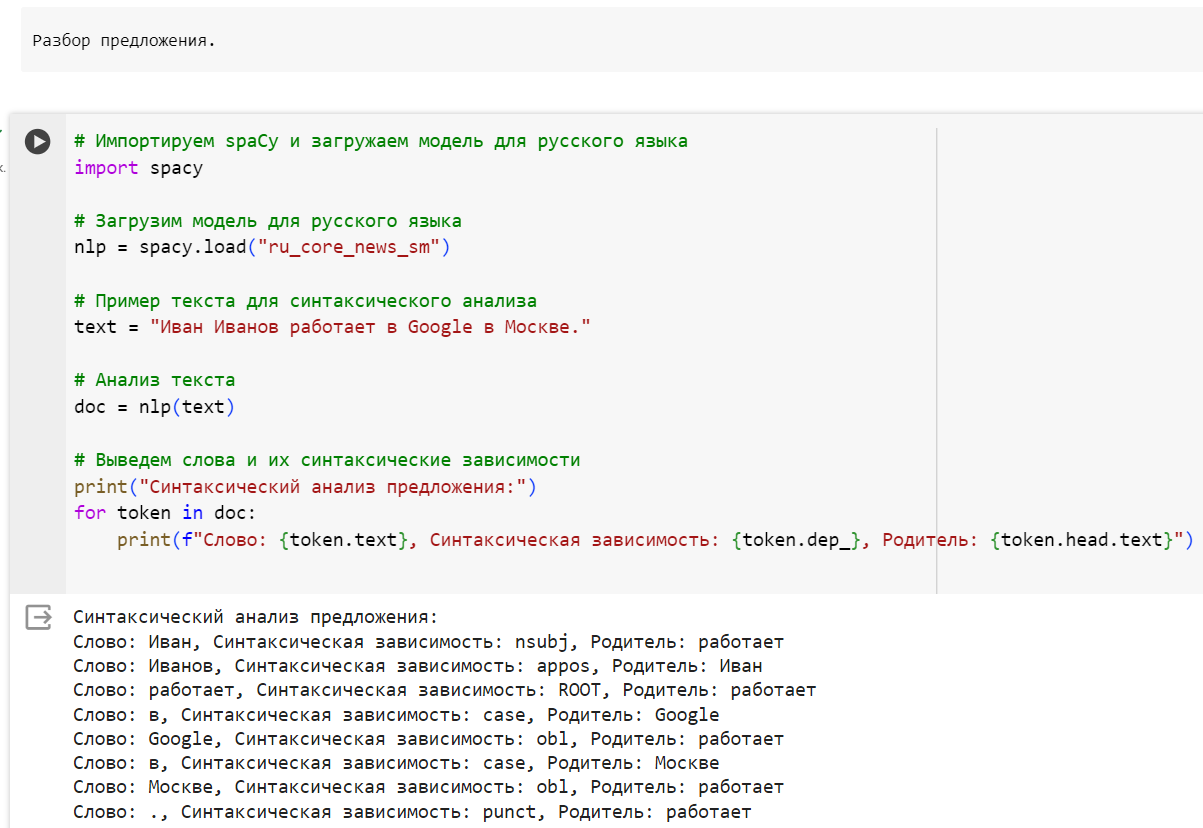
**Текст программы и экранные формы с примерами выполнения программы**

****

****

****

****

****

**Вывод:**

В рамках данной лабораторной работы была проведена токенезация, частеречная разметка, лемматизация, выделение (распознавание) именованных сущностей.