****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 группы ИУ5-22М

Джин Шуо

Москва — 2024 г.

**1. Цель лабораторной работы**

изучение методов предобработки текстов.

**2. Задание**

Для произвольного предложения или текста решите следующие задачи:

Токенизация.

Частеречная разметка.

Лемматизация.

Выделение (распознавание) именованных сущностей.

Разбор предложения.

**3. текст программы**

import spacy

nlp = spacy.load("en\_core\_web\_sm")

text = "Apple is looking at buying U.K. startup for $1 billion."

doc = nlp(text)

print("Participle：")

for token in doc:

print(token.text)

print("\nPart of speech tag：")

for token in doc:

print(f"{token.text}: {token.pos\_}")

print("\nLemmatization：")

for token in doc:

print(f"{token.text}: {token.lemma\_}")

print("\nNamed entity recognition：")

for ent in doc.ents:

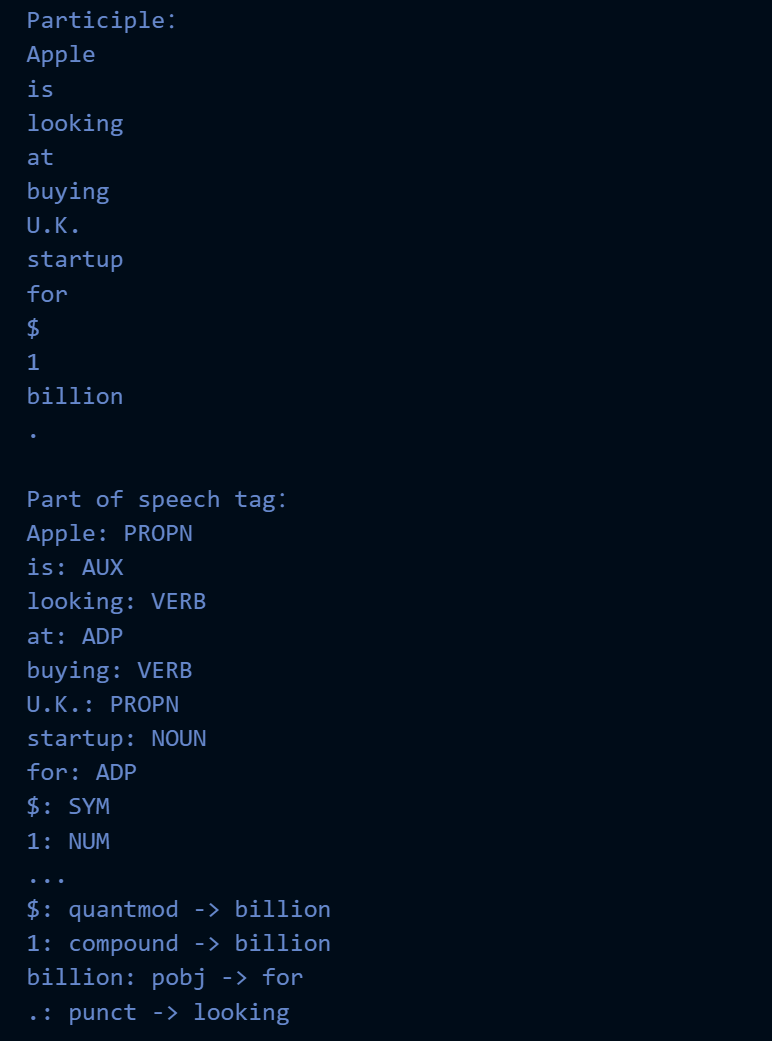
print(f"{ent.text}: {ent.label\_}")

print("\nSyntactic analysis：")

for token in doc:

print(f"{token.text}: {token.dep\_} -> {token.head.text}")

1. **экранные формы с примерами выполнения программы.**



**Список литературы**

[1] https://github.com/ugapanyuk/ml\_course\_2021/wiki/LAB\_MMO\_\_TEXT\_PRE