

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Н.Э. Баумана

Факультет «Информатика и системы управления»
Кафедра «Систем обработки информации и управления»

ОТЧЕТ

Лабораторная работа № 8
по дисциплине «Методы машинного обучения в АСОИУ»

ИСПОЛНИТЕЛЬ:

группа ИУ5-22М

Воронцова А.В.

ФИО

подпись

" " 2024 г.

ПРЕПОДАВАТЕЛЬ:

Балашов А. М.

ФИО

подпись

" " 2024 г.

Москва - 2024

Задание:

Для произвольного предложения или текста решите следующие задачи:

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

```
import spacy

nlp = spacy.load("en_core_web_sm")
text = "Apple is looking at buying U.K. startup for $1 billion"

# Обработка текста
doc = nlp(text)

# Токенизация
print("Токены:")
tokens = [token.text for token in doc]
print(tokens)

# Частеречная разметка (Part-of-Speech Tagging)
print("\nЧастеречная разметка:")
for token in doc:
    print(f"{token.text} ({token.pos_})")

# Лемматизация
print("\nЛеммы:")
lemmas = [token.lemma_ for token in doc]
print(lemmas)

# Распознавание именованных сущностей (Named Entity Recognition)
print("\nИменованные сущности:")
entities = [(ent.text, ent.label_) for ent in doc.ents]
print(entities)

# Синтаксический разбор (Dependency Parsing)
print("\nСинтаксический разбор:")
for token in doc:
    print(f"{token.text} -> {token.dep_} -> {token.head.text}")
```

Токены:
['Apple', 'is', 'looking', 'at', 'buying', 'U.K.', 'startup', 'for', '\$', '1', 'billion']

Частеречная разметка:
Apple (PROPN)
is (AUX)
looking (VERB)

at (ADP)
buying (VERB)
U.K. (PROPN)
startup (NOUN)
for (ADP)
\$ (SYM)
1 (NUM)
billion (NUM)

Леммы:

['Apple', 'be', 'look', 'at', 'buy', 'U.K.', 'startup', 'for', '\$',
'1', 'billion']

Именованные сущности:

[('Apple', 'ORG'), ('U.K.', 'GPE'), ('\$1 billion', 'MONEY')]

Синтаксический разбор:

Apple -> nsubj -> looking
is -> aux -> looking
looking -> ROOT -> looking
at -> prep -> looking
buying -> pcomp -> at
U.K. -> dobj -> buying
startup -> dep -> looking
for -> prep -> startup
\$ -> quantmod -> billion
1 -> compound -> billion
billion -> pobj -> for