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

Лабораторная работа №3
по дисциплине
«Методы машинного обучения»

Выполнил:
студент группыИУ5И-23М

Лю Цзычжан

Москва — 2025 г.

Цель лабораторной работы: изучение методов предобработки текстов.

Требования к отчету:

Отчет по лабораторной работе должен содержать:

1. титульный лист;
2. описание задания;
3. текст программы;
4. экранные формы с примерами выполнения программы.

Задание:

Для произвольного предложения или текста решите следующие задачи:

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

```
import spacy

# Загружаем модель spacy для английского языка
nlp = spacy.load("en_core_web_sm")

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

# Пропускаем текст через пайплайн spacy
doc = nlp(text)

```

```
# Токенизация
print("Tokens:")
for token in doc:
    print(f"{token.text}", end=" | ")
print("\n")

# Частеречная разметка
print("POS tagging:")
for token in doc:
    print(f"{token.text}: {token.pos_} ({token.tag_})")
print("\n")

# Лемматизация
print("Lemmas:")
for token in doc:
    print(f"{token.text} -> {token.lemma_}")
print("\n")

# Именованные сущности
print("Named Entities:")
for ent in doc.ents:
    print(f"{ent.text} ({ent.label_})")
print("\n")

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

```

```
(venv) PS E:\BMSTU\一下\jqxx\MMO\np5> python -u "e:\BMSTU\一下\jqxx\MMO\np5\lab5.py"
Tokens:
Apple | is | looking | at | buying | U.K. | startup | for | $ | 1 | billion | . |

POS tagging:
Apple: PROPN (NNP)
is: AUX (VBZ)
looking: VERB (VBG)
at: ADP (IN)
buying: VERB (VBG)
U.K.: PROPN (NNP)
startup: VERB (VBD)
for: ADP (IN)
$: SYM ($)
1: NUM (CD)
billion: NUM (CD)
.: PUNCT (.)


Lemmas:
Apple -> Apple
is -> be
looking -> look
at -> at
buying -> buy
U.K. -> U.K.
startup -> startup
for -> for
$ -> $
1 -> 1
billion -> billion
. -> .
```

```
. -> .


Named Entities:
Apple (ORG)
U.K. (GPE)
$1 billion (MONEY)

Lemmas:
Apple -> Apple
is -> be
looking -> look
at -> at
buying -> buy
U.K. -> U.K.
startup -> startup
for -> for
$ -> $
1 -> 1
billion -> billion
. -> .


Named Entities:
Apple (ORG)
U.K. (GPE)
$1 billion (MONEY)
Apple -> Apple
is -> be
looking -> look
at -> at
buying -> buy
U.K. -> U.K.
startup -> startup
for -> for
$ -> $
1 -> 1
billion -> billion
. -> .
```

```
Named Entities:
Apple (ORG)
U.K. (GPE)
$1 billion (MONEY)
looking -> look
at -> at
buying -> buy
U.K. -> U.K.
startup -> startup
for -> for
$ -> $
1 -> 1
billion -> billion
. -> .


Named Entities:
Apple (ORG)
U.K. (GPE)
$1 billion (MONEY)
for -> for
$ -> $
1 -> 1
billion -> billion
. -> .


Named Entities:
Apple (ORG)
U.K. (GPE)
$1 billion (MONEY)
billion -> billion
. -> .


Named Entities:
Apple (ORG)
U.K. (GPE)
$1 billion (MONEY)
```

```
Named Entities:
Apple (ORG)
U.K. (GPE)
$1 billion (MONEY)

Named Entities:
Apple (ORG)
U.K. (GPE)
$1 billion (MONEY)


Dependency Parsing:
Apple (ORG)
U.K. (GPE)
$1 billion (MONEY)


Dependency Parsing:
$1 billion (MONEY)


Dependency Parsing:


Dependency Parsing:
Apple --> nsubj --> looking
Apple --> nsubj --> looking
is --> aux --> looking
is --> aux --> looking
looking --> ROOT --> looking
looking --> ROOT --> looking
at --> prep --> looking
at --> prep --> looking
buying --> pcomp --> at
U.K. --> nsubj --> startup
startup --> ccomp --> buying
```

```
startup --> ccomp --> buying
for --> prep --> startup
$ --> quantmod --> billion
1 --> compound --> billion
billion --> pobj --> for
. --> punct --> looking
```