

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



Лабораторная работа №5
По курсу «Методы машинного обучения»

«Предобработка текста»

ИСПОЛНИТЕЛЬ:

Чичикин Тимофей Дмитриевич
Группа ИУ5-25М

ПРОВЕРИЛ:

Гапанюк Ю.Е.

Цель работы:

Изучение методов предобработки текста.

Задание:

Для произвольного предложения или текста решите следующие задачи:

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

Описание задания:

Для выполнения лабораторной работы возьмём фразу: «Предобработка данных в XML файле».

Выполнение работы:

1. Токенизация. NLTK
2. Частеречная разметка. Natasha
3. Лемматизация. Natasha
4. Выделение именованных сущностей. Natasha
5. Разбор предложения. Natasha

Вывод:

Была проделана работа по изучению методов предобработки текста, все задачи были выполнены.