

**Ministerul Educației și Cercetării al Republicii Moldova**

**Universitatea Tehnică a Moldovei**

**Facultatea Calculatoare Informatică și Microelectronică**

**Departamentul Informatică și Ingineria Sistemelor**

**Raport practica de documentare**

**Tema: Metode de procesare a limbajului natural pe dispozitive cu resurse hardware limitate fără conexiune la internet.**

**Student :**

**st. gr. MAI-211M Ivashchenko Nikolai**

**Coordinator:**

**Istrati Daniela, conf. Univ., mag.**

**Chișinău 2023**

## 1. ОГЛАВЛЕНИЕ

1. ОГЛАВЛЕНИЕ .....	2
2. ВВЕДЕНИЕ .....	2
3. АНАЛИЗ И ОПРЕДЕЛЕНИЕ ИНТЕРЕСУЮЩЕЙ ОБЛАСТИ.....	4
3.1. Общая информация по использованным технологиям .....	4
3.1.1. Что такое Natural Language Processing? .....	4
3.1.3. Причина использования машинного обучения .....	5
3.1.3.1. Распознавание паттерна .....	6
3.1.3.2. Выявление признаков .....	7
3.2. Этапы оптического распознавания символов (OCR) .....	8
3.2.1. Предобработка.....	8
3.2.2. Распознавание .....	8
3.2.3. Постобработка .....	9
3.2.4. Особые случаи .....	9
4. ИССЛЕДОВАНИЕ ТРЕХ СУЩЕСТВУЮЩИХ СИСТЕМ .....	11
4.1. “OCR” .....	11
4.2. “Google-translate” .....	15
4.3. “Text scanner” .....	19
5. ОПРЕДЕЛЕНИЕ ПРЕДНАЗНАЧЕНИЯ ОБЪЕКТА .....	23
6. КОНЦЕПЦИЯ НОВОЙ СИСТЕМЫ .....	24
6.1. Идентификация технических методов и средств .....	30
6.1.1. Распознавание текста с помощью нейронной сети.....	30

6.1.2. Классификация текста с помощью нейронной сети и машинный перевод.	31
6.1.3. Синтез речи или Text-to-Speech (TTS).....	31
6.1.4. Как работает Text-to-Speech .....	33
6.2. Методы, используемые в системном анализе.....	36
7. ВЫВОДЫ .....	38
8. СПИСОК ЛИТЕРАТУРЫ.....	39



## 2. ВВЕДЕНИЕ

Практически каждый студент хотя бы раз в жизни сталкивался с задачей перевода напечатанного на бумаге текста в цифровой, например, в процессе написания дипломной работы, перепечатывая целые страницы из старой и очень редкой литературы, использованной в качестве источника, содержавшего в себе матчасть, которую совершенно невозможно найти в цифровом виде. Даже те документы, которые возможно увидеть на экране, могут быть представлены в виде изображений, созданных с помощью фотографии или сканера для бумаг, или pdf-файлов, откуда также невозможно скопировать текст и приходится его перепечатывать для дальнейшей обработки.

С течением времени, в результате технологического прогресса, в том числе и в сфере информационных технологий, проблемы, решение которых ранее казалось фантастикой, сегодня решаются за считанные секунды, и даже не требуют больших затрат ресурсов. Порой для решения таких задач, достаточно всего лишь карманного смартфона, которым сегодня пользуется каждый. Одна из главных технологий, упрощающих жизнь современного человека - машинное обучение.

## 8. СПИСОК ЛИТЕРАТУРЫ

1. Стивен Берд, Эдуард Лопер, и Юэн Кляйн – «Natural Language Processing with Python», 2009г.
2. Йоав Гольдберг - «Нейросетевые методы в обработке естественного языка», 2017г.
3. Ричард Риз - «Обработка естественного языка на Java», 2022г.
4. Ануй Гупта, Харшит Сурана, Соумья Вайяла, Бодхисаттва Мажумдер - «Practical Natural Language Processing», 2020г.
5. Алекс Томас - «Natural Language Processing with Spark NLP», 2020г.
6. Денис Ротман - «Transformers for Natural Language Processing», 2021г.
7. James H. Martin, Дэниел Журафски - «Speech and Language Processing», 2000г.