

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

Аннотация к дипломной работе

**ИЗВЛЕЧЕНИЕ ДАННЫХ ИЗ СКАНОВ ДОКУМЕНТОВ МЕТОДАМИ
ГЛУБОКОГО ОБУЧЕНИЯ**

Чурко Тимур Олегович

Научный руководитель:
кандидат физ.-мат. наук,
доцент
Голубева Лариса
Леонидовна

В дипломной работе 56 страниц, 26 рисунков, 1 таблица, 9 источников, 1 приложение.

ОПТИЧЕСКОЕ РАСПОЗНАВАНИЕ СИМВОЛОВ, ИСКУССТВЕННАЯ НЕЙРОННАЯ СЕТЬ, АЛГОРИТМ, ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА, ПОСТОБРАБОТКА, СЛОЙ.

Объектом исследования является алгоритм извлечения данных из сканов документов.

Целью работы является проектирование и разработка алгоритма извлечения данных из сканов документов, который будет способен работать как на персональных компьютерах, так и на мобильных устройствах.

Для достижения поставленной цели использовались следующие методы и инструменты: метод глубокого обучения, оптическое распознавание символов, язык программирования Python 3.11

В дипломной работе получены следующие результаты:

- 1) Изучены алгоритмы предварительной обработки изображений, оптического распознавания символов и постобработки данных
- 2) Разработан алгоритм извлечения данных из сканов документов
- 3) Проверена работа разработанного алгоритма и проверена оптимальность выбранных параметров

Дипломная работа является практической. Ее результаты могут быть использованы для заполнения web-форм и мобильных приложениях.

Дипломная работа выполнена, поставленные задачи полностью решены, есть возможность дальнейшего развития исследования.

Дипломная работа выполнена автором самостоятельно.

Thesis project is presented in the form of an explanatory note of 56 pages, 26 figures, 1 table, 9 references, 1 application.

OPTICAL CHARACTER RECOGNITION, ARTIFICIAL NEURAL NETWORK, ALGORITHM, PREPROCESSING, POSTPROCESSING, LAYER.

The research object of this thesis project is to design and development of an algorithm for extracting data from document scans, which will be able to work both on personal computers and on mobile devices.

The purpose of this work is to design and development of an algorithm for extracting data from document scans, which will be able to work both on personal computers and on mobile devices

The following methods and tools were used to achieve the goal: deep learning method, optical character recognition, programming language Python 3.11

The main results of the thesis project are as follows:

- 1) Studied algorithms for image pre-processing, optical character recognition and data post-processing
- 2) An algorithm for extracting data from document scans has been developed
- 3) The work of the developed algorithm was checked and the optimality of the selected parameters was checked.

This thesis project is a practical one. Its results can be used to fill web forms and mobile applications.

The thesis project is complete, all tasks have been successfully done, there is a possibility for further research and development.

The thesis project was done solely by the author.