

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

Аннотация к магистерской диссертации
ОБРАБОТКА ДОКУМЕНТОВ И ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

Ключник Карина Владимировна

Научный руководитель:
кандидат физ.-мат. наук,
доцент А. Э. Малевич

2024

В магистерской диссертации 39 страниц, 23 иллюстрации, 19 источников.

Ключевые слова: РАСПОЗНАВАНИЕ ТЕКСТА, РУССКИЙ РУКОПИСНЫЙ ТЕКСТ, ГЕНЕРАЦИЯ НОВЫХ ДАННЫХ, ИЗБАВЛЕНИЕ ОТ ШУМОВ.

Объектом исследования диссертации является распознавание рукописного текста на русском языке.

Целью работы является выявление проблем, связанных с распознаванием рукописного русского текста, и предложение способов их решения.

Для достижения поставленной цели были использованы: язык программирования Python и фреймворк PyTorch.

В диссертации получены следующие результаты:

1. Выявлены две основные проблемы, связанные с распознаванием русского рукописного текста: зашумленность данных и недостаточное количество данных для обучения.

2. Предложены варианты решения выявленных проблем несколькими способами.

Магистерская работа является завершенной, поставленные задачи решены в полной мере. Обоснованность и достоверность полученных результатов обусловлена практическими экспериментами.

Магистерская диссертация выполнена автором самостоятельно.

The master thesis is presented in the form of an explanatory note of 39 pages, 23 figures, 19 references.

Keywords: TEXT RECOGNITION, RUSSIAN HANDWRITTEN TEXT, GENERATION OF NEW DATA, GETTING RID OF NOISE.

The object of the dissertation research is the recognition of handwritten text in Russian.

The purpose of the work is to identify problems related to the recognition of handwritten Russian text and to propose ways to solve them.

To achieve this goal, the Python programming language and the PyTorch framework were used.

The following results were obtained in the dissertation:

1. Two main problems related to the recognition of Russian handwritten text have been identified: noisy data and insufficient data for training.
2. Solutions to the identified problems are proposed in several ways.

The master's thesis is completed, the tasks have been fully solved. The validity and reliability of the results obtained is due to practical experiments.

The master thesis was done solely by the author.