Contribution ID: **139** Type: **Sectional reports**

Применение методов машинного обучения для задачи распознавания русских дореволюционных печатных текстов

Tuesday 6 July 2021 17:20 (15 minutes)

Настоящая работа посвящена вопросам применения технологий оптического распознавания символов и методов машинного обучения для распознавания печатных русскоязычных текстов XIX века. Анализируются особенности данной задачи по сравнению с общей задачей оптического распознавания символов. Проводится обзор существующих методов и программ для решения рассматриваемой проблемы. Предлагается свой адаптивный подход к построению программной системы распознавания подобных текстов на основе открытой платформы Tesseract. Приводятся предварительные результаты исследования эффективности предложенного подхода и сравнения с имеющимися решениями. Работа выполнена при финансовой поддержке РФФИ (грант № 20-07-01053 А).

Summary

Authors: Mr ФЕДОРОВ, Владислав (ф-т ВМК МГУ им. М.В. Ломоносова); ERSHOV, Nikolay (Moscow State University)

Presenter: Mr ФЕДОРОВ, Владислав (ф-т ВМК МГУ им. М.В. Ломоносова)

Session Classification: Distributed computing, HPC and ML for solving applied tasks

Track Classification: 9. Big data Analytics and Machine learning