

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Факультет прикладной математики и информатики
Кафедра математического моделирования и анализа данных

Аннотация к дипломной работе

**«Обнаружение вирусов в PDF-файлах с использованием методов
машинного обучения»**

Круглик Карина Сергеевна

Научный руководитель — кандидат физ.-мат. наук, доцент

М.С. Абрамович

Минск, 2020

РЕФЕРАТ

Дипломная работа, 61 с., 25 рис., 25 табл., 16 источников, 5 прил.

PDF-ФАЙЛ, ОБНАРУЖЕНИЕ ВИРУСОВ, СТРУКТУРНЫЙ ПУТЬ, КЛАСТЕРИЗАЦИЯ, МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ, НЕЙРОННЫЕ СЕТИ, ПОДБОР ГИПЕРПАРАМЕТРОВ

Объект исследования — файлы формата PDF с вирусами и без.

Цель работы — разработка и реализация алгоритмов обнаружения вирусов в файлах формата PDF и исследование эффективности классификации PDF-файлов с вирусами и без.

В ходе работы рассматриваются три алгоритма для классификации файлов с вирусами и без:

- 1) при помощи структурных путей;
- 2) при помощи кластеризации ключевых слов;
- 3) при помощи представления файла в виде изображения.

Результаты работы — разработанные алгоритмы классификации PDF-файлов с вирусами и без.

Область применения — анализ файлов формата PDF на наличие вирусов.

ABSTRACT

Diploma work, 61 p., 25 ill., 25 tab., 16 sources, 5 app.

PDF FILE, VIRUS DETECTION, STRUCTURAL PATH, CLUSTERING,
MACHINE LEARNING METHODS, NEURAL NETWORKS,
HYPERPARAMETERS TUNING

Object of research — PDF files with and without viruses.

Purpose of work — development and implementation algorithms for detecting viruses in PDF files, also research the effectiveness of classifying PDF files with and without viruses.

During the work we consider three algorithms for classifying files with and without viruses:

- 1) using the structural paths;
- 2) using the clustering keyword;
- 3) by presenting the file as an image.

Results of work — developed algorithms for classifying PDF files with and without viruses.

Area of application — analysis of PDF files for viruses.