



ФЕДЕРАЛЬНАЯ СЛУЖБА  
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ  
ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):

2022610174

Дата регистрации: 10.01.2022

Номер и дата поступления заявки:

2021681789 24.12.2021

Дата публикации и номер бюллетеня:

10.01.2022 Бюл. № 1

Контактные реквизиты:

+7 (499) 263-69-84, patent.bmstu@mail.ru

Автор(ы):

Нелюб Владимир Александрович (RU),  
Бородулин Алексей Сергеевич (RU),  
Гантимуров Андрей Павлович (RU),  
Соколов Григорий Владимирович (RU),  
Калинников Александр Николаевич (RU)

Правообладатель(и):

Федеральное государственное бюджетное  
образовательное учреждение высшего  
образования «Московский государственный  
технический университет имени Н.Э. Баумана  
(Национальный исследовательский  
университет)» (МГТУ им. Н.Э. Баумана) (RU)

Название программы для ЭВМ:

Модуль автоматического парсинга данных и формирования датасетов

**Реферат:**

Программа предназначена для получения данных из различных источников и формирования их в датасет неструктурированных данных. В программе используется алгоритм, который позволяет обратиться по указанному пути к директории с общим доступом, просканировать её на предмет наличия данных табличной формы в форматах csv, txt, xlsx, sheet и выполнить загрузку в память для дальнейшей обработки. Из полученных данных формируется датасет, как структурированных, так и неструктурированных данных. Программа может быть применяться для подготовки входных данных, используемых для обучения модели искусственного интеллекта. Тип ЭВМ: IBM PC-совмест. ПК; ОС: Linux.

Язык программирования: Python

Объем программы для ЭВМ: 3 КБ