

РАЗРАБОТКА ПРОГРАММНОГО МОДУЛЯ МНОГОЧИСЛЕННОЙ КОНВЕРТАЦИИ ФАЙЛОВ ORJ В ФОРМАТ CSV

Бич Н. Н.¹, Кынкурогов А. А.², Скируха А. А.²

¹Гродненский государственный медицинский университет

²Гродненский государственный университет имени Янки Купалы
Гродно, Беларусь

Актуальность. Актуальность поставленной в статье задачи, суть которой содержится в ее названии, обусловлена отсутствием соответствующего программного продукта как в отечественном мире разработок, так и в зарубежном. Наличие такой разработки поможет одновременно проанализировать содержимое многочисленного количества файлов.

Цель. Разработать программный модуль для массового преобразования файлов формата orj в формат csv.

Материалы и методы исследования. Основательное исследование наночастиц, как правило, проводится посредством конфокального микроскопа. Аппарат работает, непрерывно сканируя с определенным диапазоном регистрации волн некоторые гистологические образцы. В данном процессе для каждой точки образца происходит считывание суперпозиции пиков спектра. После чего формируется файл orj.

Файл orj представляет собой проект, созданный с помощью программного обеспечения Origin для визуализации данных [1]. Этот файл обычно содержит один или несколько наборов данных и графиков, сгенерированных из результатов сканирования. С ним неудобно работать в связи со сложностью организации анализа его формата, но csv-файл, напротив, имеет простое представление, доступен для исследования не только путем программирования.

Файлы csv (файлы данных с разделителями-запятыми) – это файлы особого типа, которые можно создавать и редактировать в Excel [1]. В csv-файлах данные хранятся не в столбцах, а разделены запятыми. Текст и числа, сохраненные в csv-файле, можно легко переносить из одной программы в другую [1]. Основываясь на простоте их структуры, рационально импортировать формат orj в csv.

Несмотря на то, что оригинальная программа Origin не имеет возможности анализа и обработки файлов orj, в ней содержится функция преобразования данного формата в csv-файл. Недостатком данной опции является выполнение конвертации для каждого отдельного файла, что делает данную процедуру медленной и неэффективной. Доработать функции Origin посредством встроенного языка программирования не предоставляется возможным, так как он узконаправлен. Поэтому принято решение разработать самостоятельный программный модуль на языке высокого уровня.

Выбор среди инструментов программирования, таких как Python и R, был сделан в пользу второго.

R – язык программирования, созданный специально для статистического анализа данных [2]. Он оригинален, имеет свой уникальный синтаксис, функции и принципы работы [2]. У языка R четкая сфера применения – статистические вычисления, анализ данных и машинное обучение [2]. Кроме того, R – не только язык для анализа данных, но и целая рабочая среда, в которую уже встроены готовые методы статистического анализа и инструменты для визуализации [2]. Однако основной его плюс заключается в том, что он работает почти под любой операционной системой, включая Windows7, что немаловажно, так как многие государственные учреждения, не располагая дополнительными средствами бюджетного и внебюджетного финансирования, работают с малоэффективным аппаратным, а в следствие чего и программным обеспечением.

Результаты.

Схема работы приложения представлена на рисунке 1.

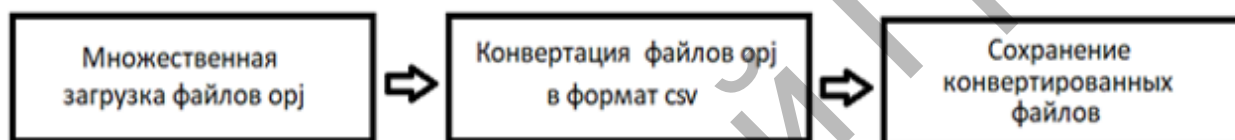


Рисунок 1 – Схема работы программного модуля

Исходя из схемы видно, что разработанное приложение однозадачно, но его открытый код позволит в дальнейшем дописать дополнительные библиотеки для анализа полученных данных, рассматривая их не только как выборки небольших объемов, но и в формате BigData.

Выводы. Разработанный программный продукт, несмотря на свою функциональную простоту, не имеет аналогов. Он может быть использован как базовый модуль в дальнейшем спроектированной и реализованной сложной информационной системы, позволяющей анализировать данные, обучаться, делать прогнозы.

Литература

1. URL: <https://clck.ru/3GmVPa> (дата обращения: 28.02.2025).
2. URL: <https://clck.ru/3GmVfT> (дата обращения: 28.02.2025).
3. URL: <https://clck.ru/3GmVsK> (дата обращения: 28.02.2025).