

Реализация библиотеки для поточковой обработки .xlsx файлов

Свитков Сергей

группа 344

научный руководитель Ю.В. Литвинов

консультант М.В. Заведеев

СПбГУ

кафедра системного программирования

3 мая 2018 г.

Введение

- ▶ Веб-приложения в сфере биллинга, телекоммуникаций
- ▶ Различные отчеты, статистика
- ▶ Формат .xlsx
- ▶ Поточковая обработка для экономии памяти

Существующие решения

- ▶ Apache POI
 - ▶ До версии 3.8 — только работа in-memory
 - ▶ Начиная с 3.8 — Stream-API для записи, Event-API для чтения
 - ▶ Часть операций всё равно только in-memory
 - ▶ Отсутствие полных и подробных примеров работы с Stream-API
- ▶ SJXLSX
 - ▶ Не поддерживается с 2015 года
- ▶ Excel Streaming Reader
 - ▶ Разработка коммьюнити
 - ▶ Обертка над POI

Постановка задачи

- ▶ Реализовать библиотеку для потоковой обработки `xlsx` файлов
- ▶ Опубликовать библиотеку в Maven
- ▶ Провести апробацию
- ▶ Сравнить полученную реализацию с существующими

Реализация

- ▶ Идея: использовать временные файлы на диске для экономии RAM

Результаты

- ▶ Реализована библиотека для потоковой обработки `xlsx` файлов (ещё нет)
- ▶ Реализация опубликована в репозитории Maven (ещё нет)
- ▶ Проведена апробация библиотеки (ещё нет)
- ▶ Проведено сравнение реализации с существующими решениями (ещё нет)