

# Реализация библиотеки для поточковой обработки .xlsx файлов

Свитков Сергей

группа 344

научный руководитель Ю.В. Литвинов

консультант М.В. Заведеев

СПбГУ

кафедра системного программирования

2 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

# Итоги обзора

- ▶ Stream-API POI всё равно имеет проблемы с памятью
- ▶ Библиотеки, реализованные комьюнити, эти проблемы не решают
- ▶ Для записи в файл стоит создать свою реализацию
- ▶ Для чтения можно использовать парсер POI
- ▶ Полученную реализацию сравнить с существующими

# Постановка задачи

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