**Министерство науки высшего образования Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего образования**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»**

**(Университет ИТМО)**

Факультет информационных технологий и программирования

Образовательная программа Программирование и интернет-технологии

Направление подготовки (специальность) 09.03.02 - Информационные системы и технологии

О Т Ч Е Т

о производственной практике

Тема задания: Утилита DOCX и ODT -> ASCIIDoc

Обучающийся Афти Александр Александрович, М33031

Руководитель практики от профильной организации: Зубок Д.А., доцент ФИТИП, Университет ИТМО

Руководитель практики от университета: Зубок Д.А., доцент ФИТИП, Университет ИТМО

Практика пройдена с оценкой \_\_\_\_

Дата 01.07.2022

Санкт-Петербург

2022

**Введение:**

По заданию требуется утилита, которая автоматизирует перевод документации из docx(odt) в adoc. Основные требования:

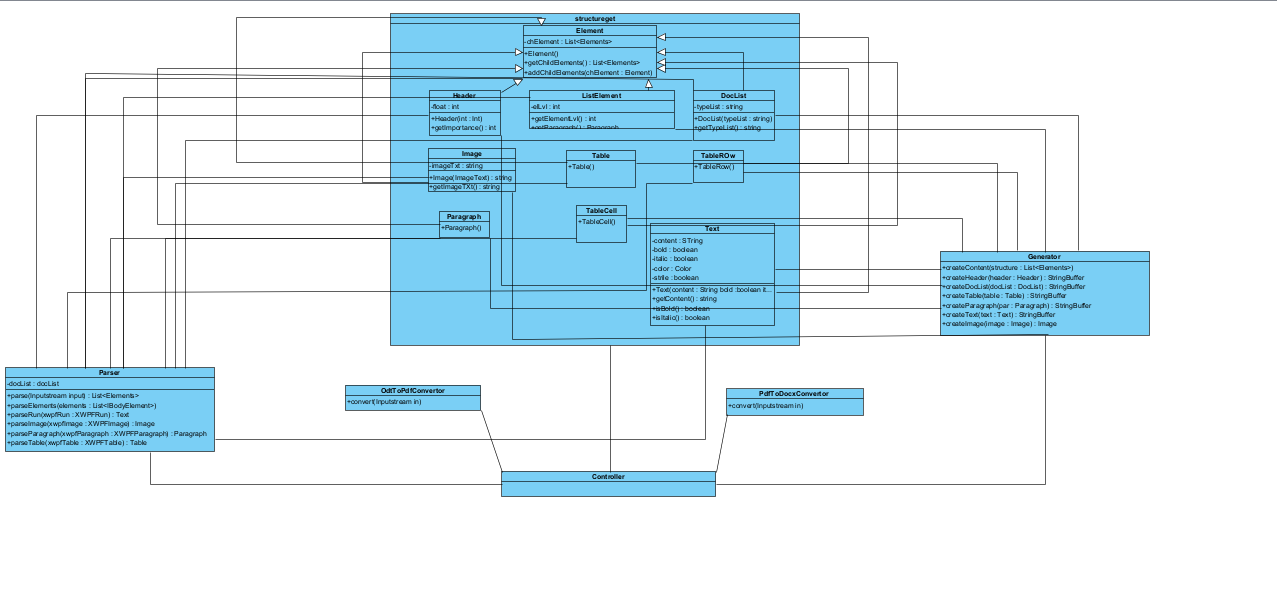
* Форматирование текста по уровням заголовков должно быть соблюдено.
* Списки должны остаться списками.
* Таблицы должны мигрироваться.
* Иллюстрации можно не мигрировать, но на их месте тогда должен быть определенный текст. В нашем случае , мы выводим запись о картинке

**Анализ требований и проектирование:**

В первую очередь , был подобран способ преобразования формата .odt документа в .word . Для этого мы использовали библиотеки Xdocreport, для преобразования формата .odt в формат .pdf , далее используя библиотеку Asponce , мы преоброзовали файл формата .pdf в нужный нам формат .docx. К сожалению , преобразования происходят не совсем корректно.

Далее перед нами стоит задача распарсить файл формата .docx и преоброзовать в формат .adoc(AsciiDoc) .Для документов в формате docxподходящей является библиотека под названием Apache POI (XWPF классы). Эта библиотека содержит все необходимые инструменты для чтения файлов docx. C помощью ее мы спокойно можем распарсить файл формата .docx. Далее по полученным данным , нам необходимо собрать файл в формате .adoc. Используя базовый синтаксис формата .adoc.

**Модель классов нашего приложения:**

Рисунок 1. Модель классов

Рассмотрим рисунок 1. В соответствии с тем, что работу программы можно поделить на два этапа, были выделены два пакета – “parser” и “generator”. Так как для обрабатываемых данных необходима структура внутри программы, по которым несложно восстановить структуру файла, то был выделен третий пакет – “structure” в котором были созданы необходимые классы в соответствии с элементами, которые встречаются в документации. В пакетах “parser” и “generator” были созданы интерфейсы – таким образом дальнейшее возможное расширение функциональности для других форматов файлов не будет затрагивать уже написанный код.

**Результаты разработки:**

В результате разработки было создано консольное приложение, которое конвертирует файл.

Результат корректности утилиты виден на рисунках 2, 3:

Рис1 . Модель классов обработчика

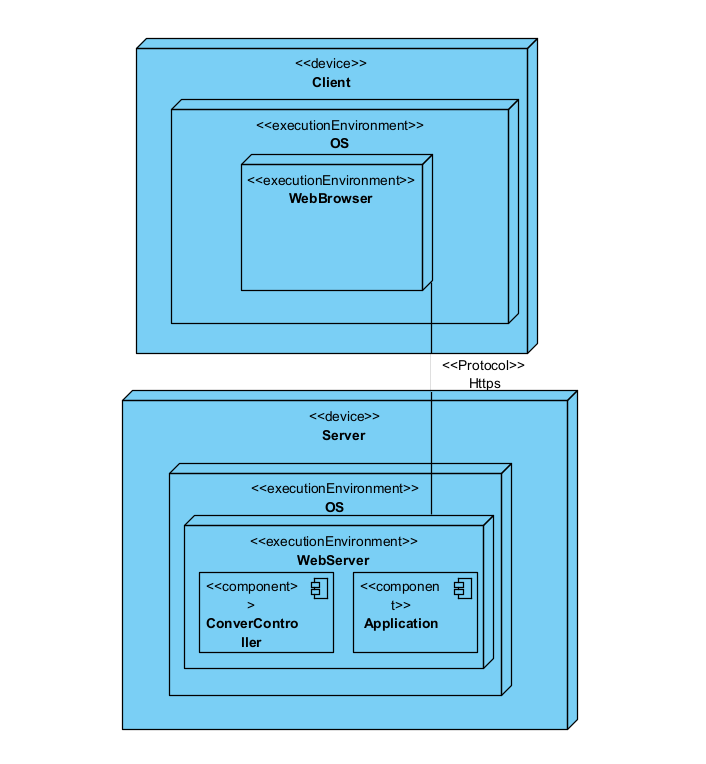
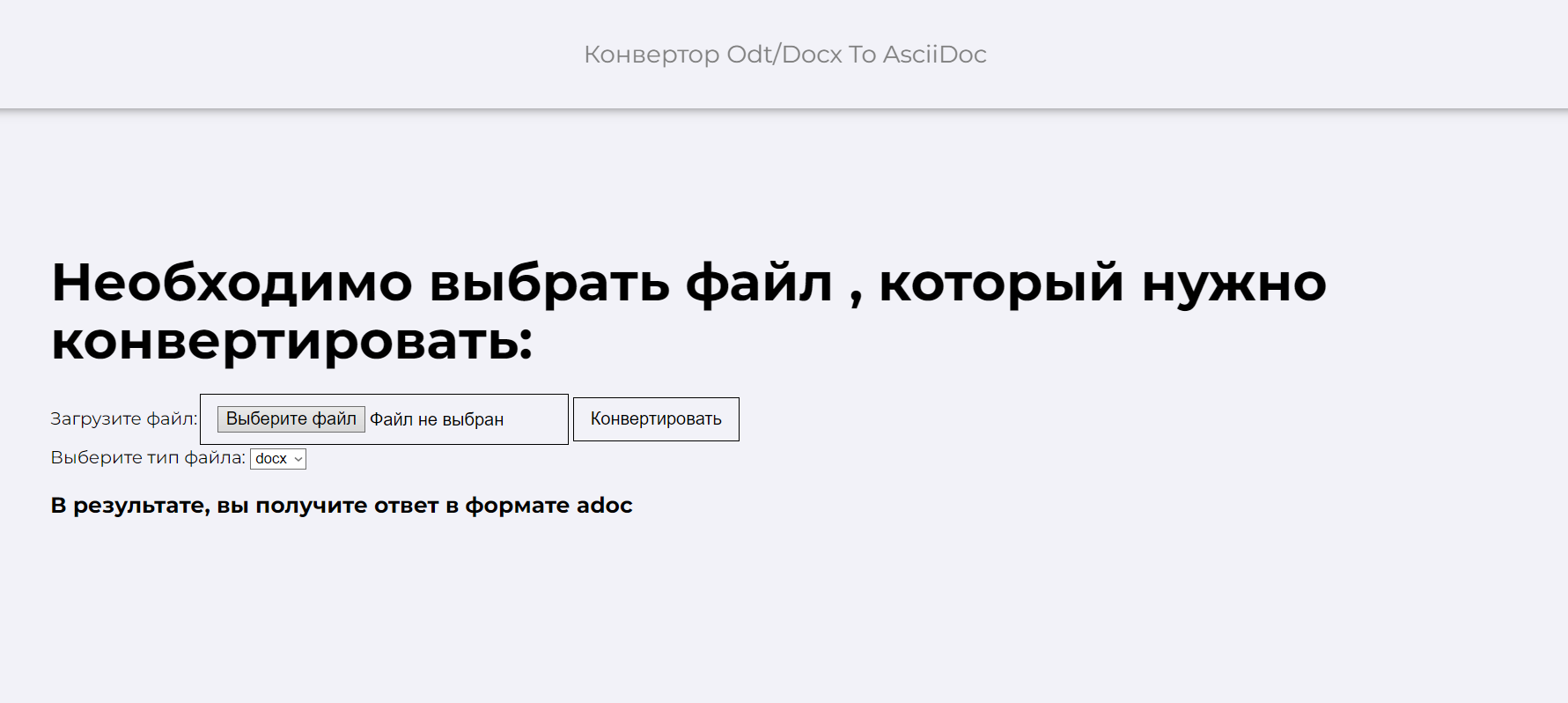
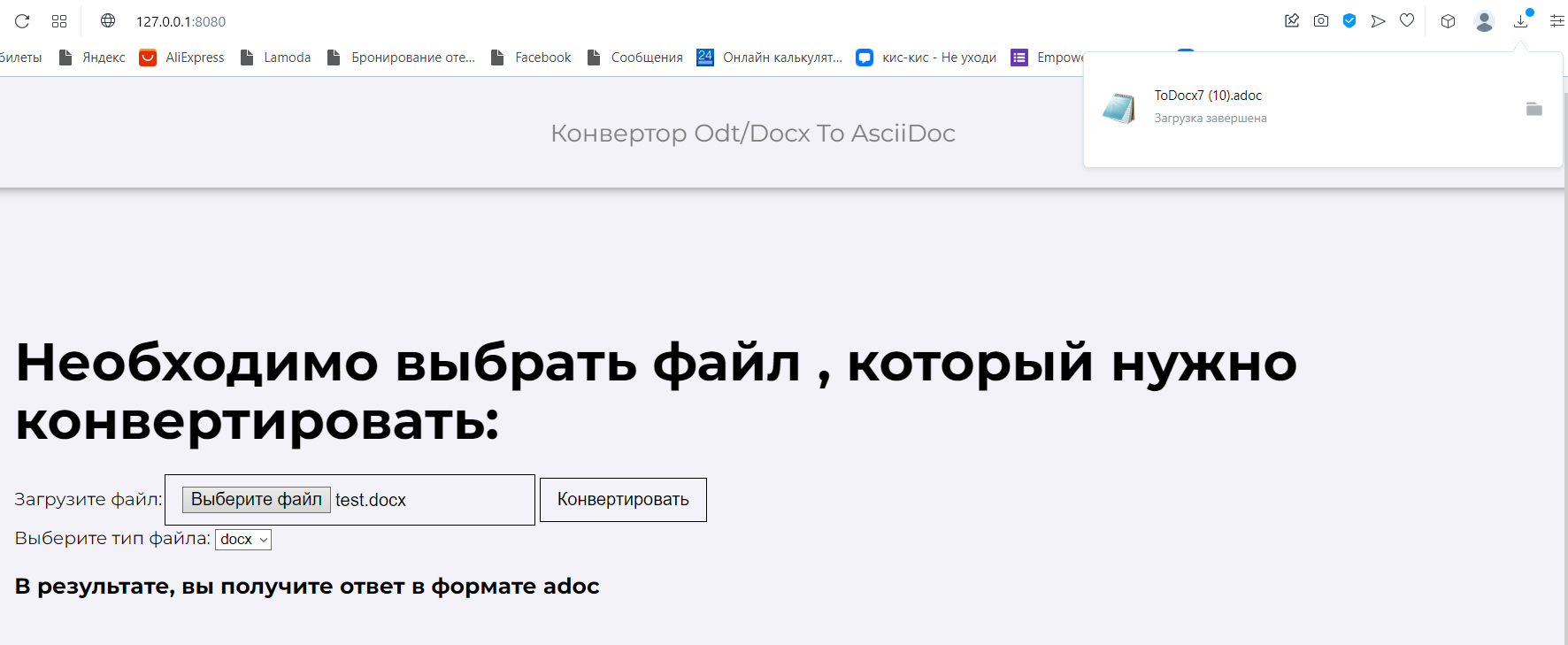
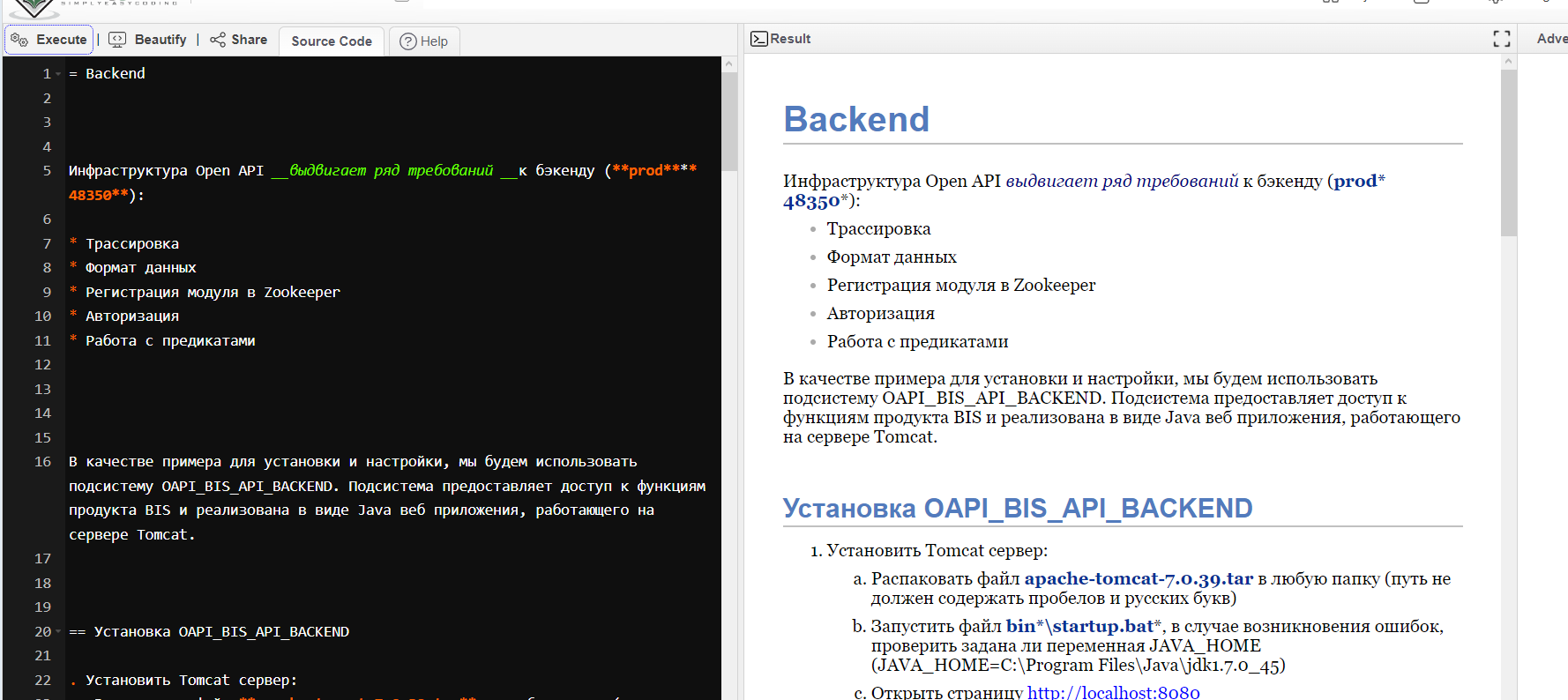


Рис2. Deployment диаграмма

**Результаты работы программы:**

Рис3. Программный интерфейс

Рис4. Результат работы

Рис5. Результат работы AsciiDoc

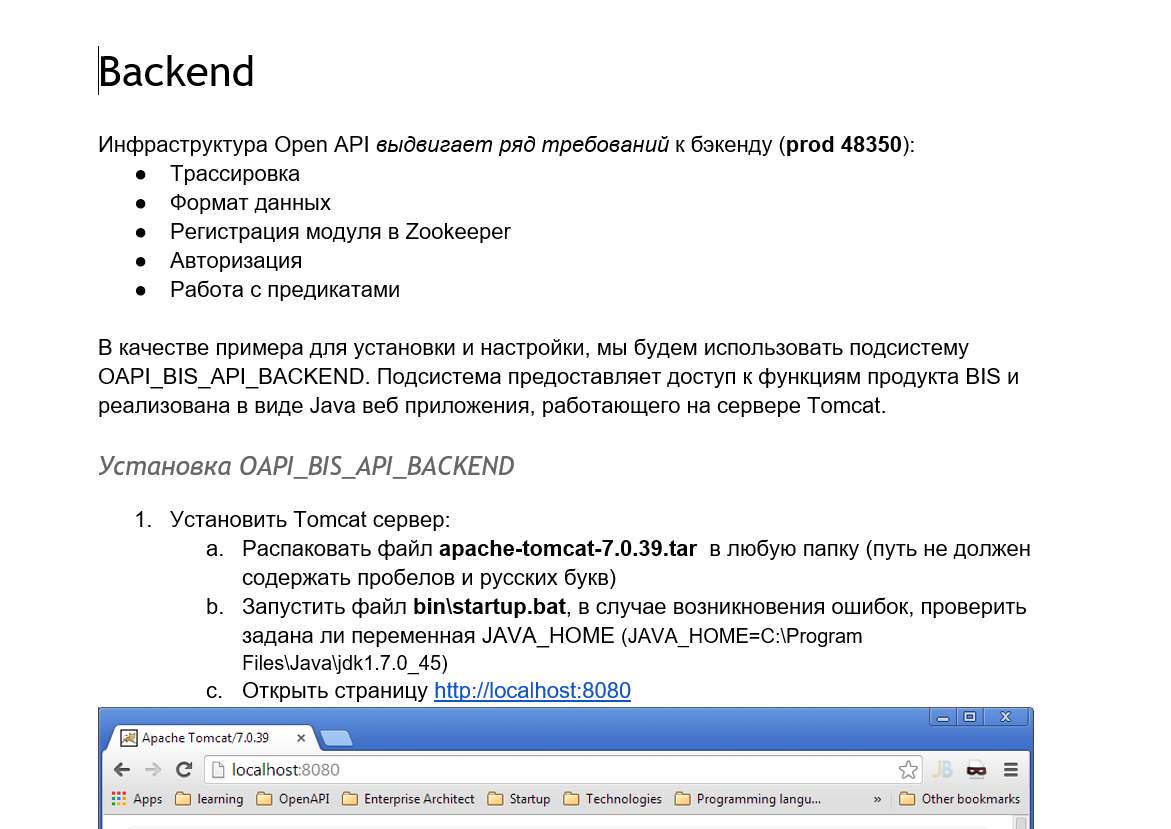


Рис6. Исходный docx документ

Как мы можем наблюдать все заголовки , таблицы перенесены успешно. Значит наша утилита работает успешно.

**Вывод:**

В ходе практической работы была разработана утилита , которая преобразует документы формата .docx(.odt) в формат .adoc. Утилита посмотроена на работе нескольких библиотек , а также на ручной сборке .adoc используя синтаксиси .adoc.