

Санкт-Петербургский политехнический университет Петра Великого  
Институт компьютерных наук и технологий  
“Высшая школа программной инженерии”

Лабораторная работа

«Скачать всё»

по дисциплине

«Технологии разработки качественного программного обеспечения»

Выполнил студент:  
гр. в3530904/80321  
Дьяченко А. Б.

Преподаватель:  
Смирнов Н. Г.

Санкт-Петербург  
2021

### **Цель работы:**

Целью работы является разработка приложения, которое позволяет скачивать все изображения с сайта

### **Постановка задач работы:**

Приложение принимает на вход URL-страницы и скачивает все изображения на странице а также на всех дочерних страницах, на которые можно перейти по гиперссылкам.

Скачивает только те изображения, размер которых превышает заданный пользователем.

Параметры приложения (URL, каталог для сохранения, размер изображения, глубина скачивания) должны задаваться в property-файле

### **Описание архитектуры приложения:**

Приложение состоит из одного класса и четырёх функций. Выполнение приложения начинается с метода `init()`. Здесь приложение считывает основные параметры, необходимые для его работы: URL-адрес начальной страницы, минимальный размер скачиваемого изображения в килобайтах и адрес локального каталога, в который необходимо загружать файлы, а так же выделяется базовый адрес страницы и создаётся объект `HashSet`, который будет выполнять функцию каталога посещённых страниц .

После этого приложение выполняет функцию `getHTML( URL )`, где создаётся другой (временный) объект `HashSet`, куда будут добавляться все ссылки с текущей страницы, чтобы избежать повторений.

Далее, чтобы исключить скачивание картинок с другого сайта, приложение проверяет вхождение базового адреса в адрес текущей страницы и если всё в порядке, то происходит соединение с сайтом. Скачиваются все изображения по тегу `"img"` с помощью функции `getImg()`, где выделяются ссылки из атрибута `src` каждого тега, относительные ссылки дополняются базовым URL и вызывается функция `download()`, где перед скачиванием, GET запросом проверяется размер файла и если он достаточно велик, изображение скачивается и записывается в файл.

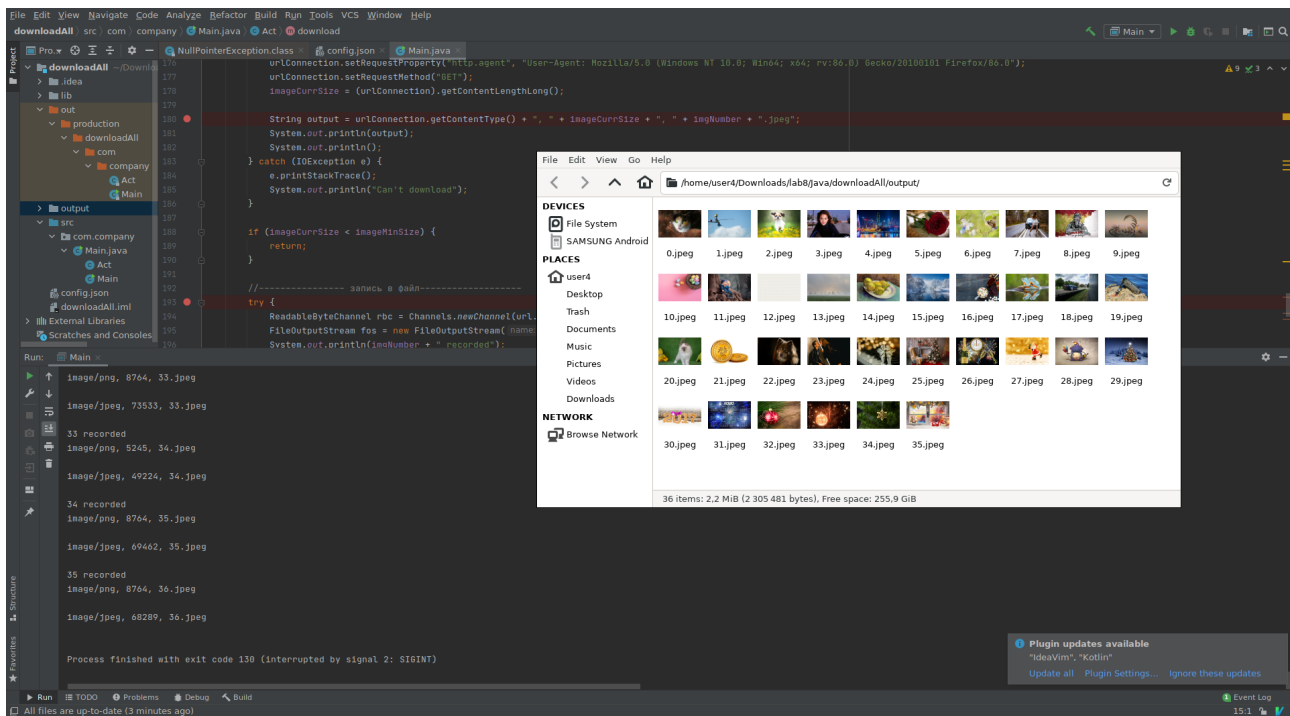
Потом собираются все ссылки по тегу `"a"` и если приложение ещё не перешло на необходимую глубину, каждая ссылка по атрибуту `"href"` добавляется во временный `HashSet`, в том числе формируются абсолютные ссылки из относительных.

Затем, если программа была запущена впервые, все полученные ссылки копируются в каталог, а в последующие разы из временного `HashSet`а` будут

удаляться уже известные сайты и оставшиеся будут заноситься в общий каталог.

После, по сформированному списку ссылок из временного каталога, приложение “спускается глубже” снова вызывая функцию getHTML( URL )

## Скриншот приложения:



## Перечисление использованных API и библиотек:

Приложение использует библиотеку JSOUP для парсинга HTML и Json-simple для удобного извлечения значений из конфигурационного файла

## Вывод:

В результате работы было реализовано приложение, способное скачивать изображения с сайтов в зависимости от их размера. Приложение определяет размер до скачивания, не переходит на сторонние сайты, учитывает относительные ссылки и ведёт учёт посещённых страниц.