

**Федеральное агентство связи**  
**Ордена Трудового Красного Знамени**  
**Федеральное государственное бюджетное образовательное учреждение**  
**высшего образования**  
**«Московский технический университет связи и информатики»**

Кафедра Математической Кибернетики и Информационных Технологий



**Отчет по лабораторной работе**  
по предмету «Кроссплатформенное программирование»  
на тему:  
«Веб–сканер»

Выполнил: студент группы

БВТ1802

Дворянинов Павел Владимирович

Руководитель:

Ксения Андреевна Полянцева

Москва 2020

## Цель работы

Изучить работу web-сканера.

## Задание

Написать программу «Web-сканер» для поиска всех ссылок, которые находятся на web-странице, в пределах указанной глубины поиска. URL-адрес и глубина поиска задаются в аргументах программы.

## Выполнение

Class Crawler.java

```
import java.io.*;
import java.net.*;
import java.util.LinkedList;

public class Crawler {
    final static int AnyDepth = 0;
    private LinkedList<URLDepthPair> Processed_se = new LinkedList<URLDepthPair>();
    private LinkedList<URLDepthPair> NotProcessed_se =
        new LinkedList<URLDepthPair>();

    private int Depth_se;
    private String StartHost_se;

    /** Нем "/" для поддержки https */
    private String Prefix_se = "http";

    public Crawler(String host, int depth) {
        StartHost_se = host;
        Depth_se = depth;
        NotProcessed_se.add(new URLDepthPair(StartHost_se, Depth_se));
    }

    public void Scan() throws IOException {
        while (NotProcessed_se.size() > 0) {
            Process(NotProcessed_se.removeFirst());
        }
    }

    public void Process(URLDepthPair pair) throws IOException{
        /** Установка соединения и перенаправление */
        URL url = new URL(pair.getURL());
        URLConnection url_connection = url.openConnection();
        String redirect = url_connection.getHeaderField("Местоположение");

        if (redirect != null) {
            url_connection = new URL(redirect).openConnection();
        }
        Processed_se.add(pair);

        if (pair.getDepth() == 0) {
            return;
        }
    }
}
```

```

    /** Чтение ссылки */
    BufferedReader reader = new BufferedReader(new
        InputStreamReader(url_connection.getInputStream()));
    String input;
    while ((input = reader.readLine()) != null) {
        while (input.contains("a href=\"" + Prefix_se)) {
            input = input.substring(input.indexOf("a href=\"" + Prefix_se) + 8);
            String link = input.substring(0, input.indexOf('\'));
            if(link.contains(" ")) {
                link = link.replace(" ", "%20");
            }
            /** Не обрабатывает посещение одинаковых ссылок */
            if (NotProcessed_se.contains(new URLDepthPair(link, AnyDepth)) ||
                Processed_se.contains(new URLDepthPair(link, AnyDepth))) {
                continue;
            }
            NotProcessed_se.add(new URLDepthPair(link, pair.getDepth() - 1));
        }
    }
    reader.close();
}

public void getSites() {
    /** Вывод ссылок */
    for (URLDepthPair elem : Processed_se){
        System.out.println(elem.getURL());
    }
    System.out.println("Посещённые ссылки: " + Processed_se.size());
}
}

```

## Class URLDepthPair.java

```

import java.util.Objects;

public class URLDepthPair {
    private String Url_se;
    private int Depth_se;

    public URLDepthPair(String host, int depth) {
        Url_se = host;
        Depth_se = depth;
    }

    public String getURL() {
        return Url_se;
    }

    public int getDepth() {
        return Depth_se;
    }

    @Override
    public boolean equals(Object obj) {
        if (obj instanceof URLDepthPair) {
            URLDepthPair o = (URLDepthPair)obj;
            return this.Url_se.equals(o.getURL());
        }
        return false;
    }
}

```

```

@Override
public int hashCode() {
    return Objects.hash();
}
}

```

## Class ScannerApp.java

```

import java.io.IOException;

public class ScannerApp {
    public static void main(String args[]) throws IOException {
        /** Аргументы программы: [0] - ссылка, [1] - глубина поиска */
        Crawler search_url = new Crawler(args[0], Integer.parseInt(args[1]));
        search_url.Scan();
        System.out.println("Глубина поиска: " + Integer.parseInt(args[1]));
        search_url.getSites();
    }
}

```

## Результат работы

Ссылка: <https://www.kia.ru/>, глубина поиска: 1

```

Глубина поиска: 1
https://www.kia.ru/
https://www.kia.ru/press/news/113/
https://youtu.be/tdJ7VdWGMuY
https://www.kia.ru/press/news/103/
https://www.kia.ru/buy/special/
https://www.kia.ru/used\_cars/
https://www.kia.ru/buy/fleet/
https://www.kia.ru/request/order\_to/
https://www.kia.ru/service/special/
https://www.kia.ru/buy/accessories/
https://www.kia.ru/personal/loyalty/
https://www.kia.ru/about/info/
https://www.kia.ru/kiamotorsrus/magazine/
https://kia.ru/kiamotorsrus/vacancies/
https://www.youtube.com/user/KIAMotorsRussia
https://www.instagram.com/kiamotorsrussia/
https://vk.com/kiamotorsrus
https://www.facebook.com/kiamotorsrus
https://ok.ru/group/57054967365668
https://zen.yandex.ru/kia
Посещённые ссылки: 20

```

Ссылка: <https://www.kia.ru/>, глубина поиска: 2

```
https://www.facebook.com/KIA.Motors.Thailand/  
https://www.facebook.com/kiaoman/  
https://www.facebook.com/cartimes/  
https://www.facebook.com/supervictor16/  
https://www.facebook.com/KIACrnaGora/  
https://www.facebook.com/dbazar.ru/  
https://www.facebook.com/jordan.kia/  
https://www.facebook.com/kiaqatar/  
https://www.facebook.com/KiaMotorsNZ/  
https://www.facebook.com/MoscowAutoSalon/  
https://www.facebook.com/kiafun/  
https://www.facebook.com/kiabuzz/  
https://www.facebook.com/kia.aljabr/  
https://www.facebook.com/kiaofficial.kz/  
https://www.facebook.com/privacy/explanation  
https://www.facebook.com/ad_campaign/landing.p  
https://www.facebook.com/help/cookies?ref_type  
https://www.facebook.com/kiamotorsrus/?hc_ref=  
.ARDmLkePyDdDkoCFXyxaqA60Qn8lXfEGsqo77IOy4gna  
-6Hiq9hkxQpb6fU83EANQdKwb50c6Whrmp04Wbg2S9uyg  
-2A8U0nThpLV3-tMKR2HwEorY-CQFabTqZw&_tn  
https://tv.ok.ru?from=feed  
https://yandex.ru  
Посещённые ссылки: 112
```

## Вывод

Изучили работу web–сканера. Написали программу «Web–сканер» для поиска ссылок на web–странице.