**МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ**

**Ордена Трудового Красного Знамени**

**Федеральное государственное бюджетное образовательное учреждение высшего образования**

**«Московский технический университет связи и информатики»**

Кафедра «Математическая кибернетика и информационные технологии»

**Отчет по лабораторной работе №7**

по дисциплине «Информационные технологии и программирование»

на тему:

Веб-сканер

Выполнил: студент группы БФИ2101

Иванов Глеб

Проверил преподаватель:

Мосева Марина Сергеевна

Москва, 2022

**Оглавление**

[Поставленная задача 3](#_Toc122873379)

[Скриншоты выполненной работы 4](#_Toc122873380)

[Код 5](#_Toc122873381)

[Вывод 8](#_Toc122873382)

[Список источников 9](#_Toc122873383)

**Поставленная задача**

1. Реализовать сканер для поиска кол-ва сайтов на странице с определённым значением глубины поиска

**Скриншоты выполненной работы**

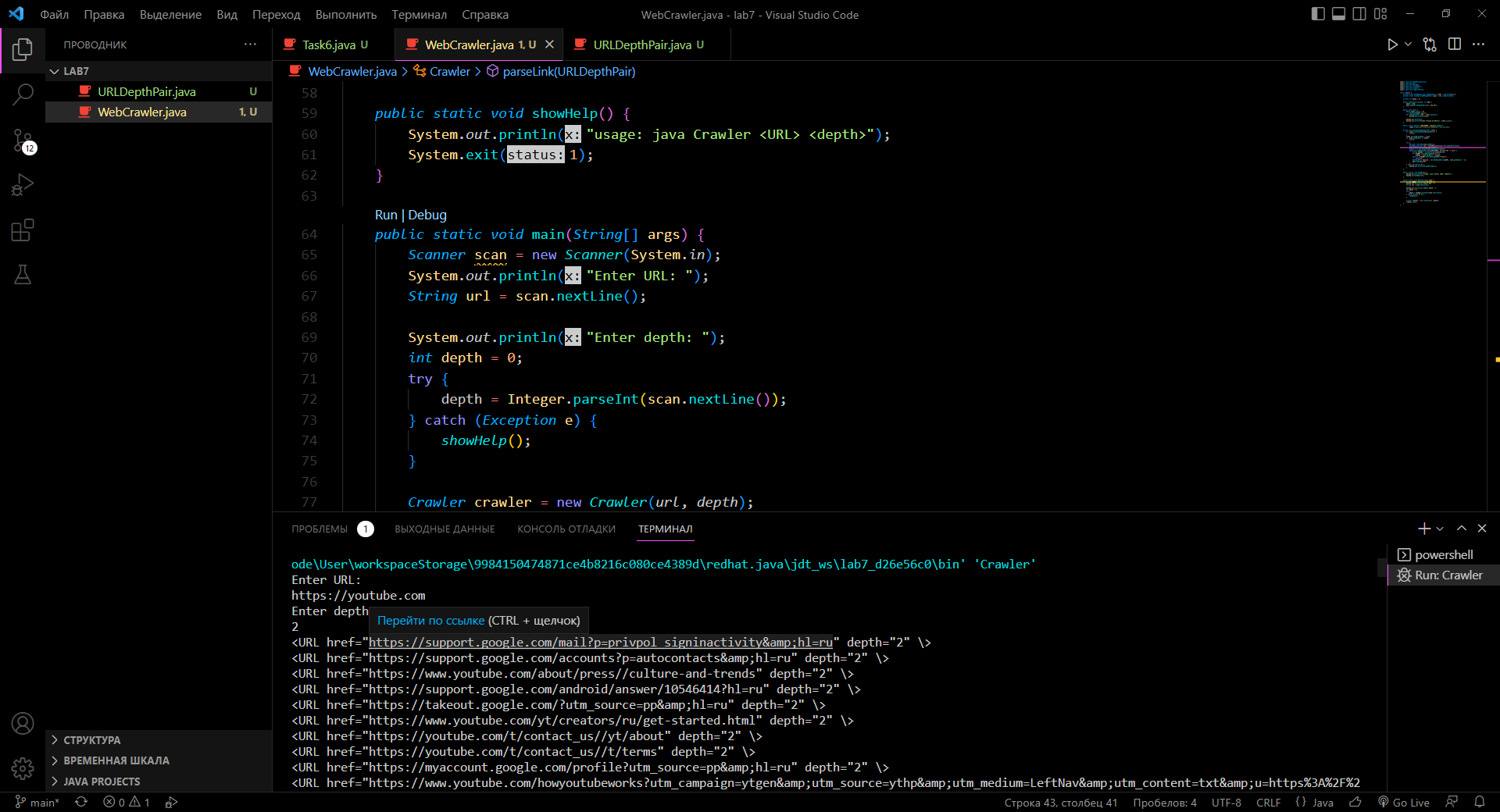
****

Рисунок 1 – Поиск всех сайтов на веб-странице <https://youtube.com>

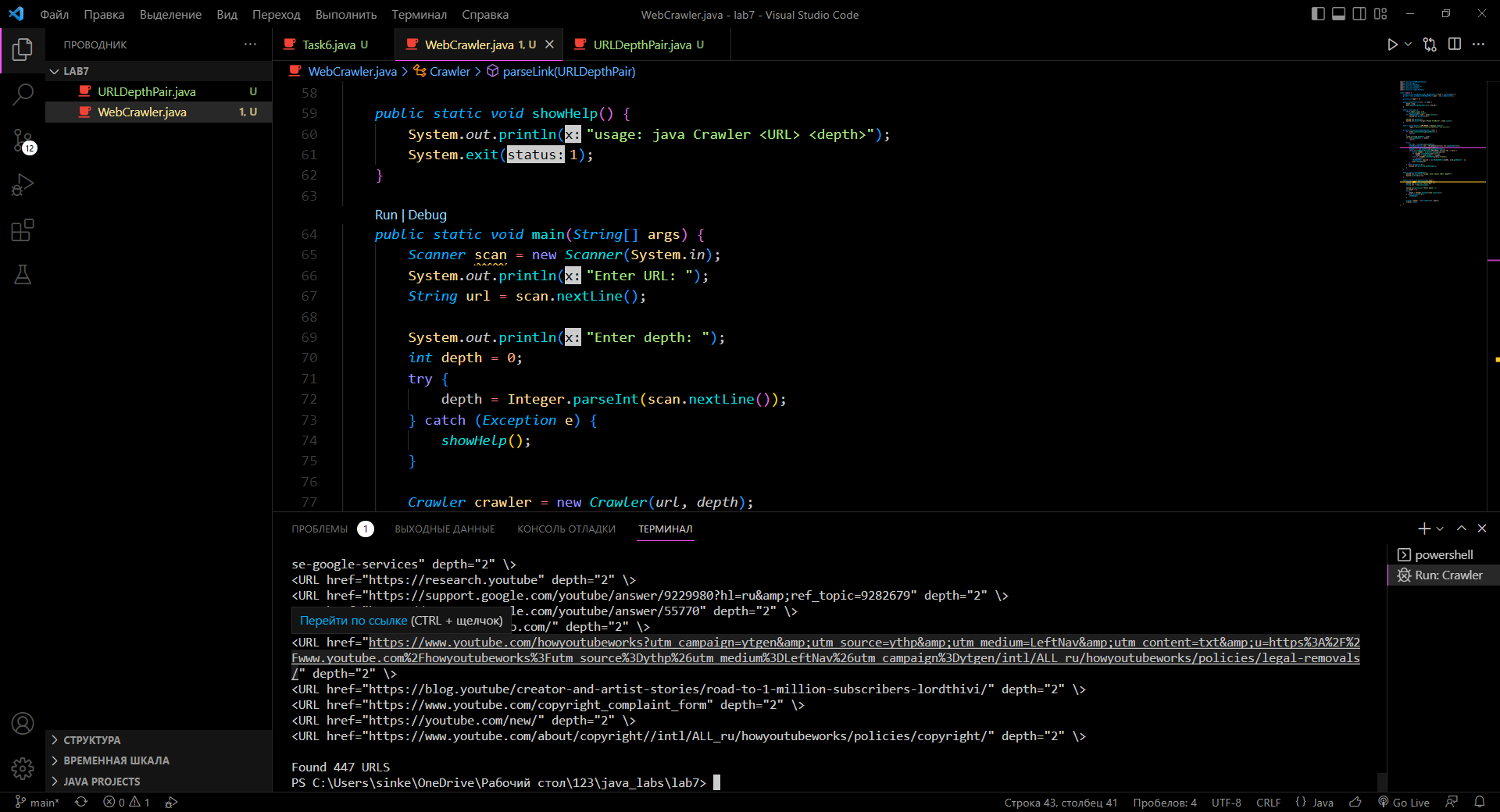


Рисунок 2 – Результат поиска

**Код**

*public* *class* URLDepthPair {

*private* *String* url;

*private* *int* depth;

*public* URLDepthPair(*String* URL, *int* dep) {

        url = URL;

        depth = dep;

    }

*public* *String* getURL() {

        return url;

    }

*public* *int* getDepth() {

        return depth;

    }

*public* *String* toString() {

        return "<URL href=\"" + url + "\" depth=\"" + depth + "\" \\>";

    }

}

Код файла URLDepthPair.java

import *java*.*net*.*HttpURLConnection*;

import *java*.*net*.*URL*;

import *java*.*util*.*HashMap*;

import *java*.*util*.*LinkedList*;

import *java*.*util*.*Scanner*;

import *java*.*util*.*regex*.*Pattern*;

*class* Crawler {

*private* *final* *HashMap*<*String*, *URLDepthPair*> links = new *HashMap*<>();

*private* *final* *LinkedList*<*URLDepthPair*> pool = new *LinkedList*<>();

*private* *int* depth = 0;

*public* Crawler(*String* url, *int* dep) {

        depth = dep;

        pool.add(new *URLDepthPair*(*url*,0));

    }

*public* *void* run() {

        while (pool.size() > 0)

*parseLink*(pool.*pop*());

        for (*URLDepthPair* link : links.values())

            System.*out*.println(link);

        System.*out*.println();

        System.*out*.printf("Found %d URLS\n", links.size());

    }

*public* *static* *Pattern* LINK\_REGEX = Pattern.compile(

            "<a\\s+(?:[^>]\*?\\s+)?href=([\"'])(.\*?)\\1");

*private* *void* parseLink(*URLDepthPair* link) {

        if (links.containsKey(link.getURL())) {

            return;

        }

        links.put(link.getURL(), link);

        if (link.getDepth() >= depth)

            return;

        try {

*URL* url = new *URL*(link.*getURL*());

*HttpURLConnection* con = (HttpURLConnection) url.openConnection();

            con.setRequestMethod("GET");

*Scanner* s = new *Scanner*(con.*getInputStream*());

            while (s.findWithinHorizon(LINK\_REGEX, 0) != null) {

*String* newURL = s.match().group(2);

                if (newURL.startsWith("/"))

                    newURL = link.getURL() + newURL;

                else if (!newURL.startsWith("http"))

                    continue;

*URLDepthPair* newLink = new *URLDepthPair*(*newURL*,link.*getDepth*()+1);

                pool.add(newLink);

            }

        } catch (*Exception* e) {

            System.*out*.println(e.getMessage());

        }

    }

*public* *static* *void* showHelp() {

        System.*out*.println("usage: java Crawler <URL> <depth>");

        System.exit(1);

    }

*public* *static* *void* main(*String*[] args) {

*Scanner* scan = new *Scanner*(System.*in*);

        System.*out*.println("Enter URL: ");

*String* url = scan.nextLine();

        System.*out*.println("Enter depth: ");

*int* depth = 0;

        try {

            depth = Integer.parseInt(scan.nextLine());

        } catch (*Exception* e) {

*showHelp*();

        }

*Crawler* crawler = new *Crawler*(*url*, *depth*);

        crawler.run();

    }

}

Код файла WebCrawel.java

**Вывод**

В ходе данной лабораторной работы я реализовал сканер для поиска сайтов на определенной веб-странице.

**Список источников**

1. Камаев В.А., Костерин В.В. Технологии программирования. М.: Высшая школа, 2006.

2. Жоголев Е.А.Технология программирования. – М.: Научный мир, 2004.