**Федеральное государственное общеобразовательное учреждение высшего образования**

**Московский государственный технический университет**

**им. Н. Э. Баумана**

Факультет "Информатика и системы управления"

Кафедра "Информационные системы и телекоммуникации"

**Отчет по лабораторной работе №7**

**"Поисковый робот"**

Студент: Задубин А.А.

Группа: ИУ3-42Б

Преподаватель: Руденкова Ю.С.

**Задание:**

В этой лабораторной работе, требовалось написать простейший поисковый робот. Робот должен был автоматически загружать веб страницы из сети Интернет, искать в них новые ссылки, и повторять эту операцию для каждой найденной ссылки.

Он просто просматривает новые URL (указывающее на расположение других веб страниц) на каждой странице, сохраняет эти ссылки и печатает их в конце работы программы. Программа должна была иметь два аргумента в командной строке:

1. Строка, содержащая URL страницы с которой начинается поиск

2. Положительное целое число – максимальная глубина поиска

Если программа получает некорректные аргументы, она немедленно прекращает работу и печатает сообщение, содержащее правила вызова, например:

Правила вызова: java Crawler <URL> <глубина поиска>

Программа должна была сохранять URL в виде строкиString вместе с глубиной (которая, в начале, равна 0). Для этого мы должны был создать класс, содержащий пару значений [URL, глубина].

Программа подключается к 80 порту заданного в URL сайта, используя класс Socket и запрашивает указанную веб страницу.

Программа должна была разбирать полученный в ответ текст (если он получен), строка за строкой отыскивая фрагменты строк такого формата:

<a href="[какой либо URL начинающийся с http://]">

Найденные URL сохраняются вместе с новым значением глубины ссылки в списке LinkedList объектов (URL, глубина). Новое значение глубины должно было быть на единицу больше глубины URL обрабатываемой страницы. Затем программа закрывает соединение с хостом.

Перед завершением программа выводит все найденные URL в пределах заданной глубины поиска.

Crawler.java

**public** **class** **Crawler** {

**public** **static** **int** lim = **0**;

**public** **static** LinkedList<URLDepthPair> waitingList = **new** LinkedList<URLDepthPair>();

**public** **static** LinkedList<URLDepthPair> visitedList = **new** LinkedList<URLDepthPair>();

**public** **static** **void** **urlSearch**(URLDepthPair linkAndDepth){}

**public** **static** **boolean** **isContainsList**(LinkedList<URLDepthPair> obj, String str) {}

**public** **static** String **patternSearch**(String line) {}

**public** **static** **void** **main**(String[] args) **throws** Exception, IOException {}

}

URLDepthPair.java

**public** **class** **URLDepthPair** {

**private** String url;

**private** **int** depth;

**public** **URLDepthPair**(String url, **int** depth) {}

**@Override**

**public** String **toString**() {}

**public** String **hostNameString**() **throws** MalformedURLException { }

**public** String **pathNameString**() **throws** MalformedURLException { }

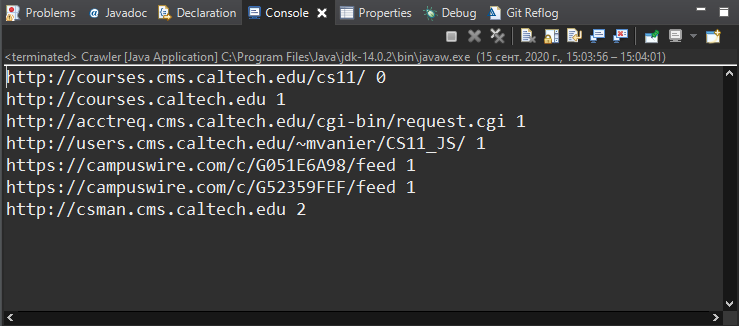
**public** **int** **getDepth**() { }

**public** String **bestContains**() { }

}

**Вывод в консоль:**

***Аргументы командной строки:***http://courses.cms.caltech.edu/cs11/ 10

****

**Вывод**

В ходе лабораторной работы был написан простейший поисковый робот, который ищет ссылки на веб-странице до заданной глубины поиска и выводит их на экран.