Name – Durwankur Naik

Roll No – I4214

Div – 2

Sub – Information Storage & Retrieval

Practical No - 7

```
import java.net.*;
import java.io.*;
public class Crawler{
 public static void main(String[] args) throws Exception{
 String urls[] = new String[1000];
 String url = "https://www.cricbuzz.com/live-cricket-scores/20307/aus-vs-ind-
3rd-odiindia-tour-of-australia-2018-19";
 int i=0,j=0,tmp=0,total=0, MAX = 1000;
 int start=0, end=0;
 String webpage = Web.getWeb(url);
 end = webpage.indexOf("<body");
 for(i=total;i<MAX; i++, total++){
 start = webpage.indexOf("http://", end);
 if(start == -1){
 start = 0;
 end = 0;
 try{
 webpage = Web.getWeb(urls[j++]);
 }catch(Exception e){
 System.out.println("******************");
 System.out.println(urls[j-1]);
 System.out.println("Exception caught \n"+e);
 }
 /*logic to fetch urls out of body of webpage only */
 end = webpage.indexOf("<body");
 if(end == -1)
 end = start = 0;
 continue;
 }
 }
 end = webpage.indexOf("\"", start);
 tmp = webpage.indexOf("'", start);
 if(tmp < end && tmp != -1){
 end = tmp;
 }
 url = webpage.substring(start, end);
 urls[i] = url;
```

```
 System.out.println(urls[i]);
 }
 System.out.println("Total URLS Fetched are " + total);
 }
}
/*This class contains a static function which will fetch the webpage
 of the given url and return as a string */
class Web{
 public static String getWeb(String address)throws Exception{
 String webpage = "";
 String inputLine = "";
 URL url = new URL(address);
 BufferedReader in = new BufferedReader(
 new InputStreamReader(url.openStream()));
 while ((inputLine = in.readLine()) != null)
 webpage += inputLine;
 in.close();
 return webpage; }}
```

Output –

```
conn built
http://www.mit.edu
http://mit.edu/site/?ref=mithomepage
http://web.mit.edu/aboutmit
http://web.mit.edu/institute-events/visitor/
http://whereis.mit.edu
http://web.mit.edu/officesdir/
http://mitstory.mit.edu/
http://web.mit.edu/admissions/
http://web.mit.edu/admissions/graduate/
http://web.mit.edu/sfs/.
http://web.mit.edu/education
http://web.mit.edu/education/
http://ocw.mit.edu/
http://odl.mit.edu/mitx/
http://web.mit.edu/research/
http://www.ll.mit.edu/
http://libraries.mit.edu/
http://web.mit.edu/community/
http://resources.mit.edu/
http://web.mit.edu/faculty/
http://web.mit.edu/staff/
http://alum.mit.edu/
http://web.mit.edu/life
http://arts.mit.edu/
http://web.mit.edu/athletics/www/
http://connect.mit.edu/
http://web.mit.edu/initiatives/
http://mitei.mit.edu
http://ki.mit.edu
http://diversity.mit.edu/
http://global.mit.edu/
http://web.mit.edu/impact/
http://web.mit.edu/industry/
http://web.mit.edu/mitpsc/
http://web.mit.edu/commencement/2014/
```