

# Act 3.6.- Web Scraping con URLConnection

## Enunciado 1

## Código

```
1 public class Actividad_3_6 {  
2  
3     public static void main(String[] args) {  
4         try {  
5             URL url = new URL("http://www.example.com/");  
6             try (InputStream is = url.openStream(); var br = new BufferedReader(new InputStreamReader(is))) {  
7                 String linea;  
8                 while ((linea = br.readLine()) != null) {  
9                     System.out.println(linea);  
10                }  
11            } catch (IOException e) {  
12                System.out.printf("ERROR: %s\n", e.getMessage());  
13            }  
14        }  
15    }  
16 }
```

## Resultado

```
Programacion-de-Servicios-y-Procesos - C:\Users\Sergio\Desktop\Estudios\Programacion-de-Servicios-y-Procesos X  Actividad_3_6 (run) X
run:
<!doctype html>
<html>
<head>
  <title>Example Domain</title>

  <meta charset="utf-8" />
  <meta http-equiv="Content-type" content="text/html; charset=utf-8" />
  <meta name="viewport" content="width=device-width, initial-scale=1" />
  <style type="text/css">
    body {
      background-color: #f0f0f2;
      margin: 0;
      padding: 0;
      font-family: -apple-system, system-ui, BlinkMacSystemFont, "Segoe UI", "Open Sans", "Helvetica Neue", Helvetica, Arial, sans-serif;
    }
    div {
      width: 600px;
      margin: 5em auto;
      padding: 2em;
      background-color: #fdfdff;
      border-radius: 0.5em;
      box-shadow: 2px 3px 7px 2px rgba(0,0,0,0.02);
    }
    a:link, a:visited {
      color: #38488f;
      text-decoration: none;
    }
    @media (max-width: 700px) {
      div {
        margin: 0 auto;
        width: auto;
      }
    }
  </style>
</head>

<body>
<div>
  <h1>Example Domain</h1>
  <p>This domain is for use in illustrative examples in documents. You may use this
  domain in literature without prior coordination or asking for permission.</p>
  <p><a href="https://www.iana.org/domains/example">More information...</a></p>
</div>
</body>
</html>
BUILD SUCCESSFUL (total time: 0 seconds)
```

## Enunciado 2

## Código

```
2
3 import java.io.BufferedReader;
4 import java.io.IOException;
5 import java.io.InputStream;
6 import java.io.InputStreamReader;
7 import java.net.URL;
8
9 public class Actividad_3_7 {
10
11     public static void main(String[] args) {
12         try {
13             URL url = new URL("http://www.example.com/");
14             try (InputStream is = url.openStream(); var br = new BufferedReader(new InputStreamReader(is))) {
15                 String linea;
16                 while ((linea = br.readLine()) != null) {
17                     if (linea.contains("https")) {
18                         System.out.println(":" + linea.trim().subSequence(3, linea.length() - 8));
19                     }
20                 }
21             }
22         } catch (IOException e) {
23             System.out.printf("ERROR: %s\n", args: e.getMessage());
24         }
25     }
26 }
27
```

## Resultado

```
run:
<a href="https://www.iana.org/domains/example">More information...</a>
BUILD SUCCESSFUL (total time: 1 second)
```