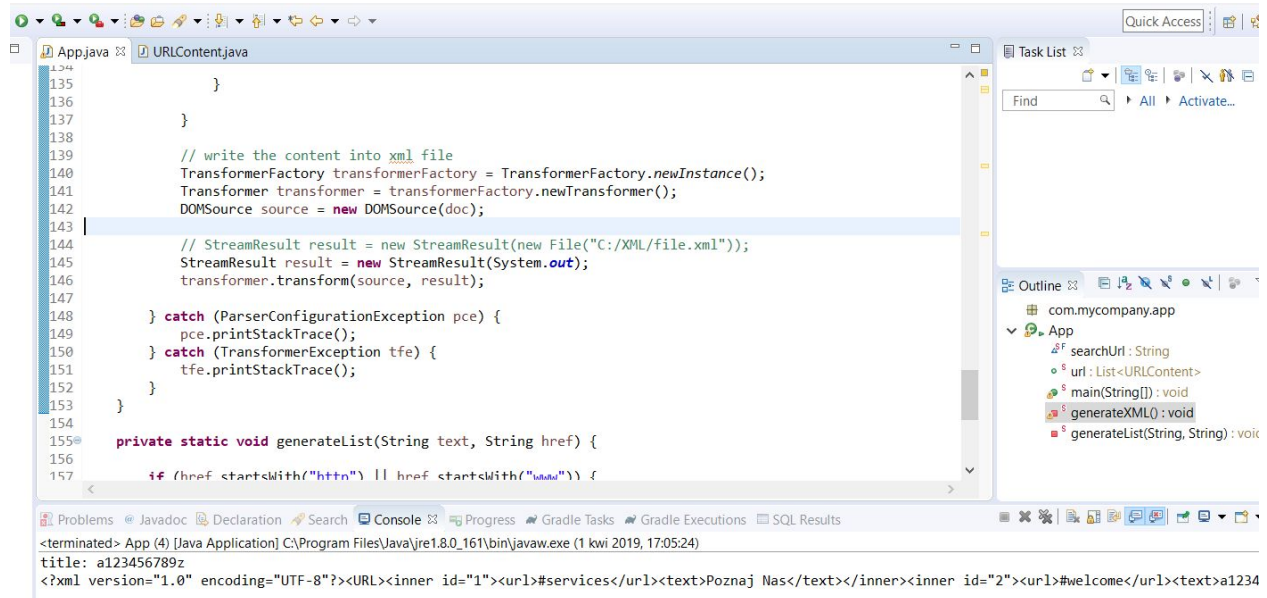


GitHub: <https://github.com/TheBestSoftInTheWorld/web-scraping>

After run project we get data from website defined **searchUrl** field. The data are save to XML file.



The screenshot shows an IDE with the following components:

- Main Editor:** Displays `URLContent.java` with Java code for parsing HTML and generating XML. The code includes comments and uses `TransformerFactory`, `DOMSource`, and `StreamResult`.
- Task List:** Located on the right, it shows a search bar and a list of tasks.
- Outline:** Also on the right, it shows the project structure with a package `com.mycompany.app` and a class `App` containing fields `searchUrl` and `url`, and methods `main`, `generateXML`, and `generateList`.
- Console:** At the bottom, it shows the output of the application, including the XML file content.

```
134 }
135
136 }
137
138 // write the content into xml file
139 TransformerFactory transformerFactory = TransformerFactory.newInstance();
140 Transformer transformer = transformerFactory.newTransformer();
141 DOMSource source = new DOMSource(doc);
142
143 // StreamResult result = new StreamResult(new File("C:/XML/file.xml"));
144 StreamResult result = new StreamResult(System.out);
145 transformer.transform(source, result);
146
147 } catch (ParserConfigurationException pce) {
148     pce.printStackTrace();
149 } catch (TransformerException tfe) {
150     tfe.printStackTrace();
151 }
152
153 }
154
155 private static void generateList(String text, String href) {
156     if (href.startsWith("http") || href.startsWith("https")) {
157
```

Console Output:

```
<terminated> App (4) [Java Application] C:\Program Files\Java\jre1.8.0_161\bin\javaw.exe (1 kwi 2019, 17:05:24)
title: a123456789z
<?xml version="1.0" encoding="UTF-8"?><URL><inner id="1"><url>#services</url><text>Poznaj Nas</text></inner><inner id="2"><url>#welcome</url><text>a1234
```