1. **Task:** Crawl Issue Reports

Build a crawler to download a Jira issue report, parse it, and store it into .csv file.

1. **Inputs:** The URL of the issue report. In this case, the URL to Apache Camel project.

<https://issues.apache.org/jira/browse/CAMEL-10597>

**Outputs:** A.csv file with extracted information.

1. **Methods:**

|  |  |
| --- | --- |
| **Method** | **Description** |
| public static String[] getIssueProperties(…) | The method that collect property types.  Inputs:  URL of the web page, web page as a Document object  Output:  Returns an array[type:String] that contains names of property types |
| public static String[] getIssuePropertiesValue(..) | The method that collect values of properties.  Inputs:  URL of the web page, web page as a Document object  Output:  Returns an array[type:String] that contains values of properties. |
| void writeDataAtOnce (..) | The method that writes extracted information to .csv file.  Inputs:  File path to .csv file [type;String], array[type:String] that contains property types/property values  Output: - |
| joinArrayGeneric(T[]... arrays) | joinArrayGeneric(T[]... arrays).  Inputs:  Multiple arrays  Output:  Returns one single array |

1. **Libraries**

Jsoup : JSoup is an open source project which provides a powerful API for data extraction. It can be used to parse HTML from URLs, files, and Strings. It can also manipulate HTML elements or attributes.

opencsv: a library for writing, reading, serializing, deserializing, and/or parsing .csv files