**Program name:**

Crawl Issue Reports

**Program requirements:**

I plan to study all issue reports filed against Apache Camel project. However, there is no database available. Could you build a crawler to download a Jira issue report, parse it, and store it into a .csv file? After you finish, please share with me the .csv file. Specially, I want to get all properties of the issue CAMEL-10597

**Requirements analyse:**

Need to crawl information from html source code and extract content of all properties. Locate the information by html tag. By drawing help from Jsoup, we can analyze html source and get elements required. The program can crawl issue reports from other camel pages as well. The program will leave the cell blank if any property is empty. As requirements and example shown, properties included in report are:

Type, Priority, Affects Version/s, Component/s, Labels, Patch Info, Estimated Complexity, Status, Resolution, Fix Version/s, Assignee, Reporter, Votes, Watchers, Created Date, Created Epoch, Updated Date, Updated Epoch, Resolved Date, Resolved Epoch, Description, Comments.

A .csv file will be created automatically and title line will be initialed. A new line will be appended in .csv file if the program crawl a new report. Each property is written in one single cell.

**Key functions designed:**

1. static String textFormatter(Element rawElement) {}

Parameter is content extracted by Jsoup from html source. Need to judge if it is null then format the text. Return the formatted String to write into report.

1. static long getTime(String create) {}

This function identifies date on html page and transfer the timestamp to the number of seconds from 1970-01-01.

**Program test:**

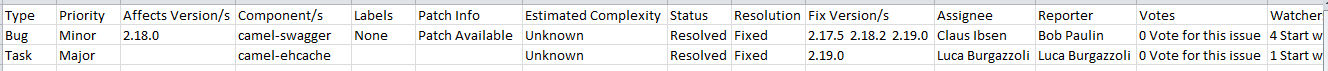
Crawl and parse the following issue report:

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

<https://issues.apache.org/jira/browse/CAMEL-10598>

Run the program and a new report.csv file is created. Issue report from the first link is written with all properties.

Run the program again by changing link to the second one. Issue report is written with all properties successfully in next line.



If the file is open when the program is started, a warning message will pop and program stops.

