Bim207 Homework-2

Homework Due Date: January 3, 2021

<mark>Due date will not be extended!</mark>

- You will create a maven project
- Your program extracts names from the given web page and prints them to the console.
- The program will take a URL as an argument.
- You will use jsoup library to get text of the body section from a web page.
  https://jsoup.org/
- The text produced by jsoup is passed to the namefinder of OpenNLP.
- The details of this process is given in Name Finder
  https://opennlp.apache.org/docs/1.9.3/manual/opennlp.html
  - You also may need Sentence Detector and Tokenizer.
  - Implementation details and examples are given in the documentation
  - Implementations require trained models and you can access them from the SourceForge Models section in following link:
    https://opennlp.apache.org/models.html
- Trained models should be under src/main/resources folder.

Important !

Make sure the following commands are running

mvn clean package

java -jar target\bim207hw2.jar *{URL}*

Example:

mvn clean package

java -jar target/bim207hw2.jar https://opennlp.apache.org/books-tutorials-and-talks.html

Note: Name finder model may not find all the names from the given text!

Output:

Berlin Buzzwords
Boris Galitsky
Drew Farris
Peter Thygesen
Tom Morton