**Feature 2 :-**

**Please follow the steps below for executing Feature 2-**

1. Clone the code from the branch named ‘feature2’ in the repository surajsrivathsa/fiction.

2. Build maven project from the file pom.xml

3. In the ‘FRConstants.java’ file, change the name of the **REP\_FN** variable to the path to the folder where the HTML content of the epub files will be extracted.

4. Replace the paths for the following folders in the ‘config.properties’ file :

a. folder.in.epub– The path to the folder where the source epubs files are present.

b. folder.out.content – The path to the folder where the HTML content of the epub files will be extracted.

c. folder.out.chunks- The path to the folder where the chunks would be stored.

d. file.feature– The location of the file containing the features extracted for all the books.

5. Replace the path for the logger file in the ‘log4j.properties’ file:

a. log4j.appender.file.File - The path where the log file will be stored.

6. Place your desired source epub files in the folder.in.epub folder

7. Run the code by right clicking on the ‘FictionRetrievalDriver.java’ file in the Project Explorer in Eclipse and select ‘Run As’ -> ‘Java Application’.

Once the program is successfully executed, the extracted feature file will be present in the specified location. The columns for the new features : 'main character presence' and 'dialog interaction ratio' are columns F20 and F21 respectively.

**Feature 1 :-**

**Assumptions:**

1. User has python (version 3.6 and above) installed, can set up virtual environments on their own and install libraries from requirements.txt in the environment.

2. User has java (1.8 and above) installed along with IDE and git integration.

3. User already has the feature extracted file for features F1-F21.

4. User has a “Final\_Booklist.csv” file, that has the list of all the book pg-ids and their corresponding book languages.

5. User has created a virtual conda environment “myenv” using the “requirements.txt” file present in Feature1 folder.

**Method:**

1. Setup local repository from a remote GitHub branch feature1(link provided above)

2. Create a conda or any other python virtual environment and install libraries as provided in requirements.txt file.

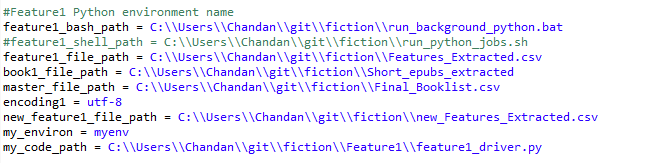
Also, download the spacy models into your local environment by executing the below commands. You can download other models too if you wish according to the website from spacy

python -m spacy download en\_core\_web\_sm

python -m spacy download de\_core\_news\_sm

3. Import and build the “pom.xml” file from the cloned project in the branch feature1.

4. Do the necessary changes for the config.properties file.



**Feature1\_bash\_path** consist of the path where the batch file is present

**Feature1\_shell\_path** consist of the path where the shell script is present

**Feature1\_file\_path** consist of the path to the already extracted feature file. (**This is the same as file.feature in the above section – part 4(d))**

**Book1\_file\_path** consist of the path to the extracted HTML books.

**Master\_file\_path** consist of the path to the “Final\_Booklist.csv” file, that has the list of all the book pg-ids and their corresponding book languages.

**New\_feature1\_file\_path** consist of the path to the new feature file that will be created after merging features of feature\_extracted csv to the extracted features of feature1 file. my\_environ consist of the virtual conda environment. my\_code\_path consist of the path of “feature1\_driver” python file.

4. Once the config.properties file is updated, run the java program.

Please note, even after installing all the python packages from the “requirement.txt” file, there could be a situation where an error is thrown for a package not being present. Please run pip install <module-name> in “myenv” environment.

**Feature 3:**

Please refer to the **Feature3 Run Document.docx** document for full