

This documents talks about the run instructions to extract feature 1.

Feature1 Git Branch: <https://github.com/surajsrivathsa/fiction/tree/feature1>

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.

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 Python environment name
feature1_bash_path = C:\\Users\\Chandan\\git\\fiction\\run_background_python.bat
feature1_file_path = C:\\Users\\Chandan\\git\\fiction\\Features_Extracted.csv
book1_file_path = C:\\Users\\Chandan\\git\\fiction\\Short_epubs_extracted
master_file_path = C:\\Users\\Chandan\\git\\fiction\\Final_Booklist.csv
encoding1 = utf-8
new_feature1_file_path = C:\\Users\\Chandan\\git\\fiction\\new_Features_Extracted.csv
my_environ = myenv
my_code_path = C:\\Users\\Chandan\\git\\fiction\\Feature1\\feature1_driver.py
```

Feature1_bash_path consist of the path where the batch file is present

Feature1_file_path consist of the path to the already extracted feature file.

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.