**CSE591 Adaptive Web Assignment 2 Report**

**Technologies Used:**

1) Django for Backend Server

2) HTML, CSS, JavaScript, Ajax(for visualizing) for Front End technologies

3) Jsoup for Web Scrapping

4) Apache Lucene for Indexing

**Installation Instructions:**

**(Django installation using virtualenv with requirements.txt)**

**System Requirements:**

MacOs/Linux with a good configuration (Min: 2 GB RAM)

**Steps:**

1. The instructions are written such that we can run the Django project in the virtual environment.
2. pip install virtualenv
3. Create a directory for the project (can give any name you wish)

cd my\_project\_folder

1. Start the virtual environment by executing the command below

$ virtualenv my\_project

1. Activate the virtual environment by executing

source my\_project/bin/activate

1. Let’s install Django and other required libraries, there is a requirements.txt which has the requirement, so once you activate you can just install all required software using

pip install -r requirements.txt

1. Copy the adaptiveweb folder from the zip inside my\_project folder
2. Go inside the adaptiveweb folder using:

cd adaptiveweb

1. Run the Django server by running:

python manage.py runserver

1. Server would start running at 127.0.0.1:8000 and it will be displayed in the logs.
2. You can navigate through the site.
3. Credentials for the website are:

aaa:1234 bbb:1234 ccc:1234 admin:qwerty1234

1. Once the evaluation of the project is done, execute “deactivate” which will shut down the virtualenv and delete the whole folder.
2. Make sure JAVA\_HOME is set for the system as Apache Lucene is executed using Java.

An exectable jar named “lu1.jar” already exists in the adaptive web folder. So you don’t have to recompile, only make sure JAVA\_HOME is set so that python can run the java code.

**Architecture:**

1. The website is hosted using a Django server, and the indexing is done in the backend using Apache Lucene which has a java code base.
2. Each query is passed from the website, **stop words and stemming is done** by Django and then passed to the jar to the program, which outputs the indexing of the particular query and it is passed back to the Django server and to the website.
3. The first time the query is run, and the folder Website doesn’t have any files, then web Scrapping function is called which **scrapes all the pages from the below 2 links** :

<https://en.wikibooks.org/wiki/Java_Programming>

<https://docs.oracle.com/javase/tutorial/>

and put it into the folder Website.

1. Before adding it to the folder all the files are cleaned of their HTML tags and other unrequired characters so that we can have better accuracy in indexing.
2. All the files are indexed using the query which comes from the Django server which is cleaned of the stop words and other unwanted characters using segmentation and basic indexing technique in Apache Lucene.
3. Once the indexing is done, we return the file names, where Django Server selects the content of the file and returns to the client where it is shown.
4. **Another feature to search for a given query is included such that user can input any data and get a recommendation for it other than the given queries**
5. **UI has been integrated with accordions so better look and feel, screenshots are included at the bottom for quick view. So for viewing each recommendation, you have CLICK THE ACCORDION to expand it.**

**Evaluation:**

● Content Collection: Both pages have been Scrapped

● Content Indexing: Stop words and stemming has been done.

● Web app: Architecture has been presented.

● Originality: Search bar has been provided to search user content with the other recommendation, UI look, and feel has been modified for better user usability

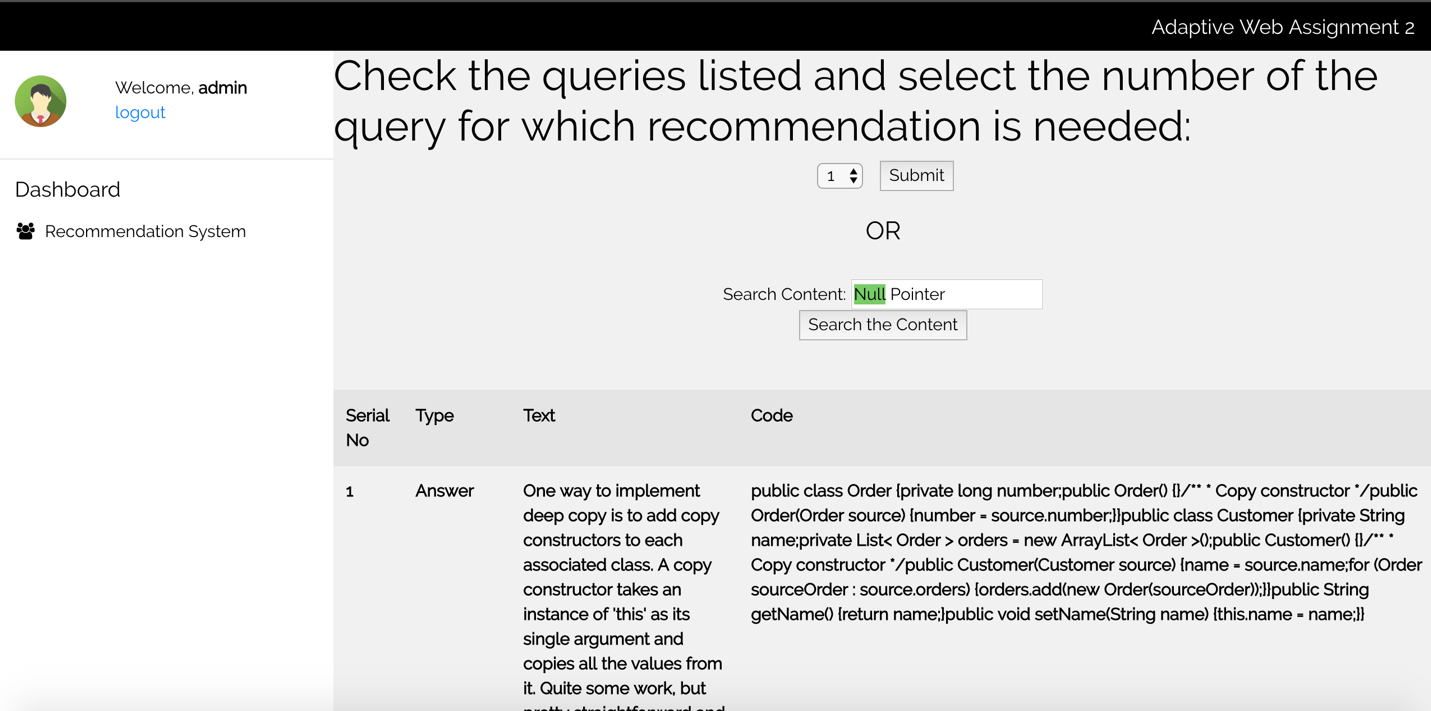
**Contents of the Directory:**

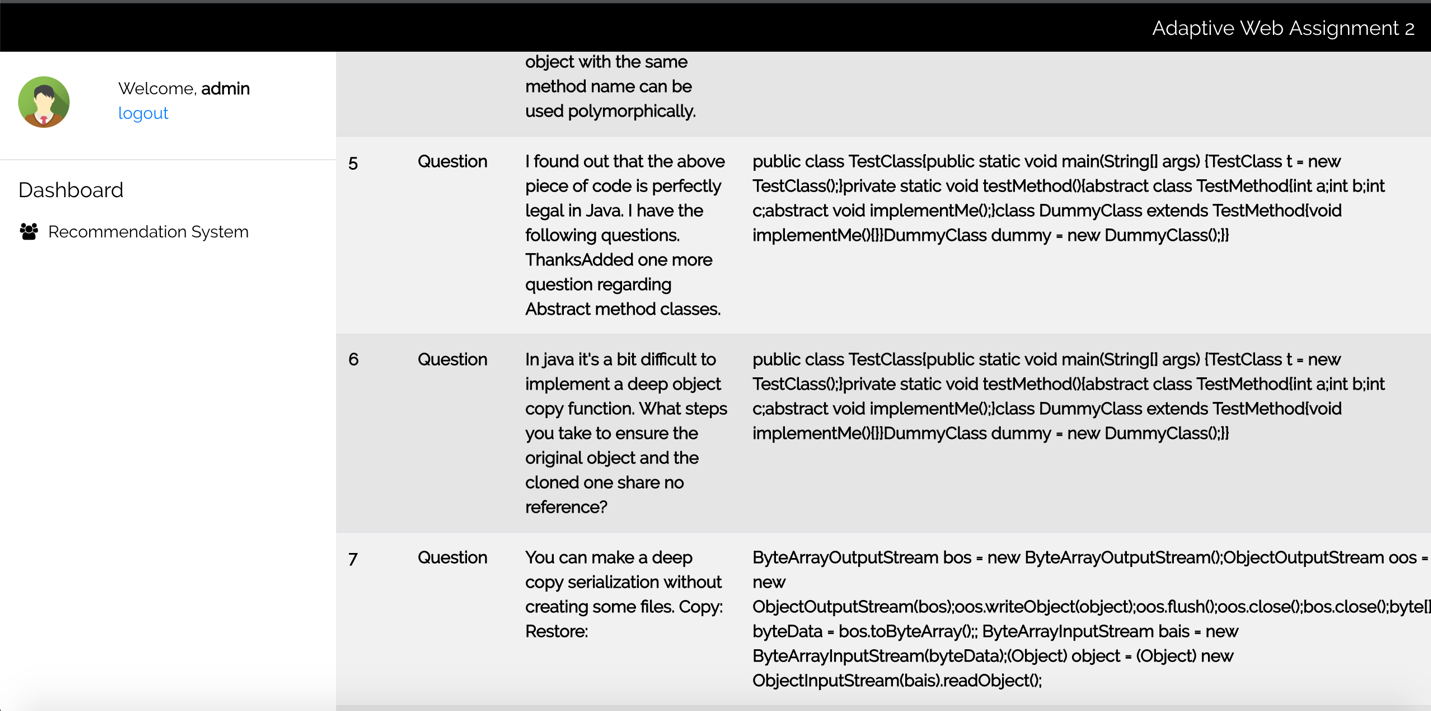
1) AdaptiveWeb folder: Contants Django code and runnable jar for Lucene as well.

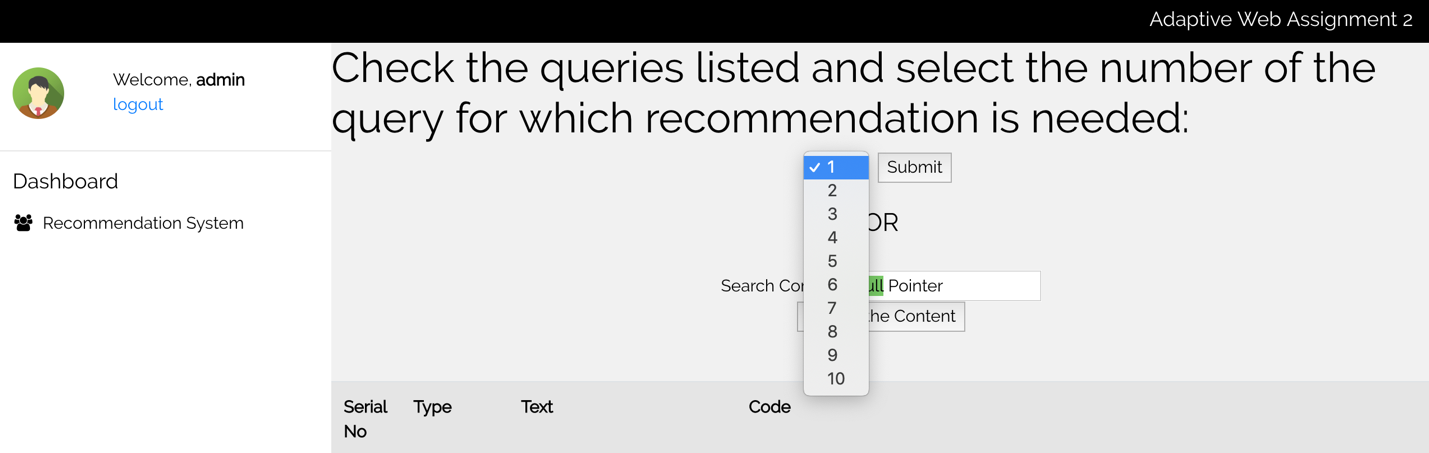
2) Lucene: JAVA Lucene project source code is included.

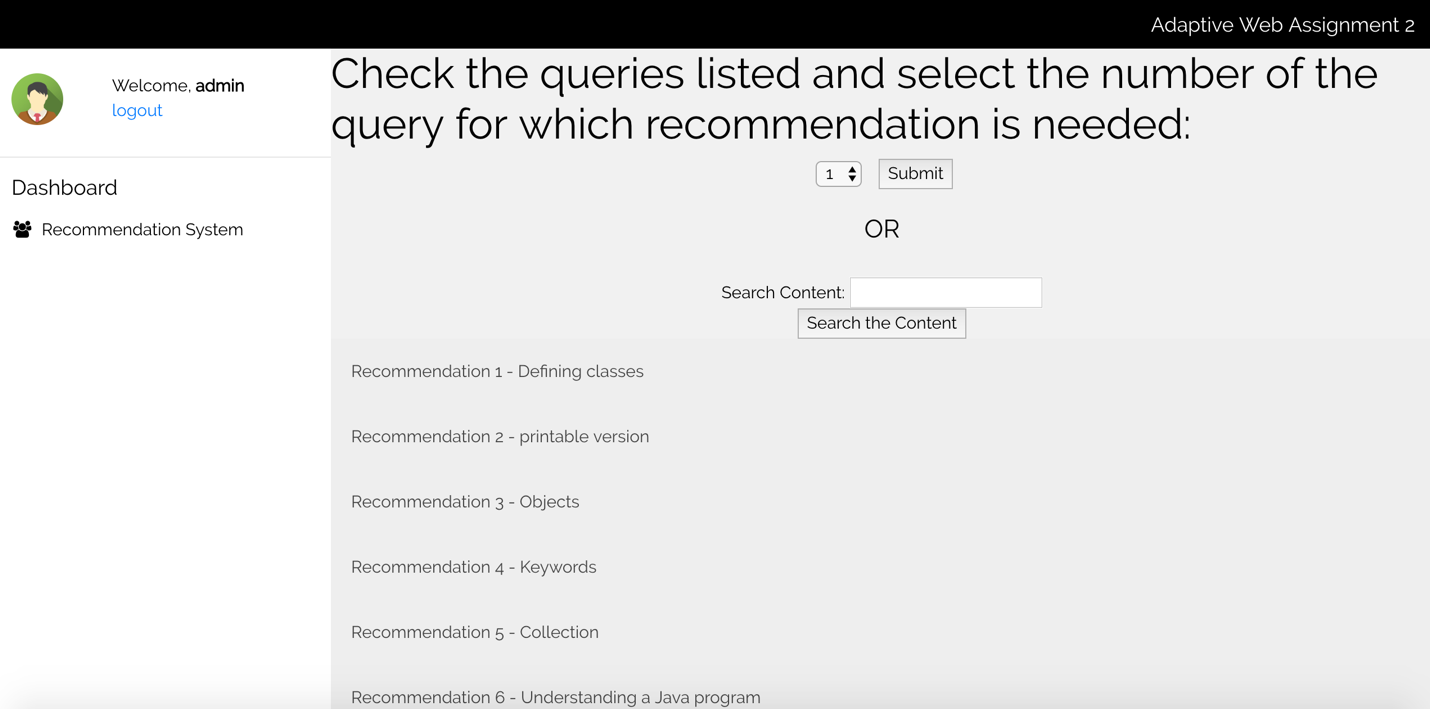
3) requirements.txt for installing the requirements in the virtualenv

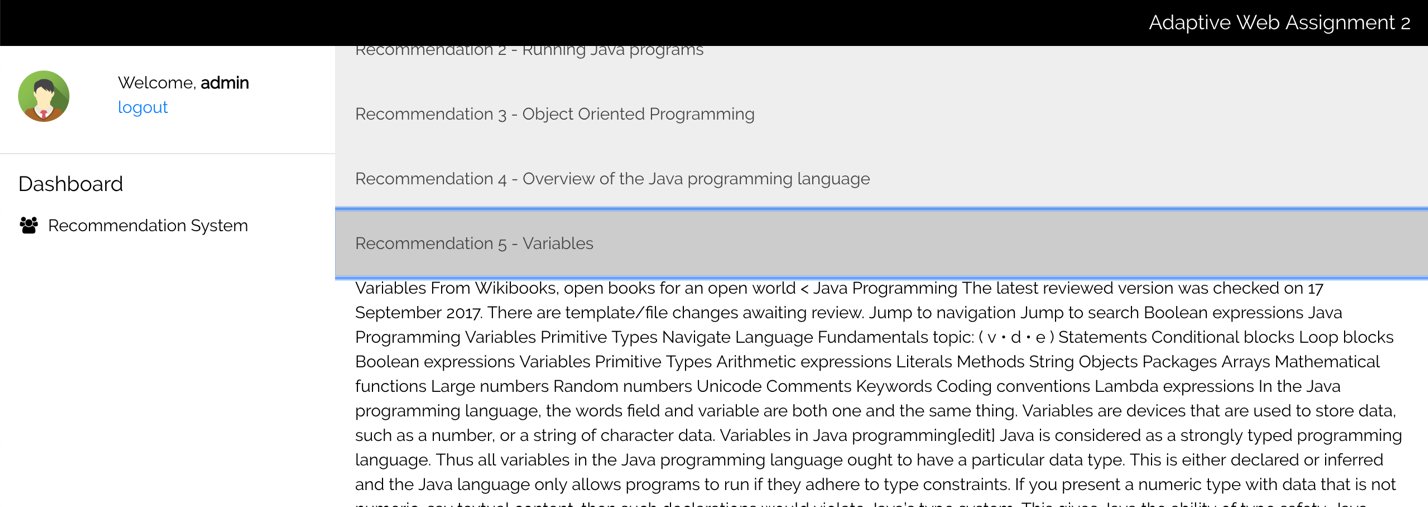
**Screenshots:**

****

****

****

****

****

**References:**

1. <https://developer.mozilla.org/en-US/docs/Web/HTML/Element/input/date>
2. <https://docs.python-guide.org/dev/virtualenvs/>
3. <https://tutorial-extensions.djangogirls.org/en/homework_create_more_models/>
4. <https://tutorial.djangogirls.org/en/>
5. <https://github.com/d3/d3/wiki/Gallery>
6. <https://southernweb.com/2018/01/track-user-movements-site/>
7. <https://stackoverflow.com/questions/7790725/javascript-track-mouse-position>
8. <https://docs.djangoproject.com/en/2.1/intro/tutorial01/>
9. <https://stackoverflow.com/questions/21126837/jquery-restrict-the-difference-between-two-datepickers>