
Programming Task : BMTC Web Scraper

The task is for the job description : [Programmer](#)

From the [BMTC website](#) scrape the same to find the following data:

1. Fares
2. BMTC routes
3. Schedules or Bus Timings
4. Bus Stops

Your tasks will include the following:

1. Write a web scraper script.
2. Create an appropriate datastructure to store the scraped data in JSON or XML formats.
3. Store the Data in a relational database. The relational database must be normalized to at least 2 NF. Use any appropriate DBMS. Along with the parsing script, you should mail the ER diagram for your schema.
4. The script written must be able to identify changes done to the website whenever the website is updated and files need to be created accordingly.

For example : You've scrapped the website on 6th of January the JSON created can be named after the data
bmtc_scrap_jan_06_2018_1.json. But if BMTC decides to update

their website on the same day at a later point in time, your script must be able to identify the same and store the new data in a new json file `bmtc_scrap_jan_06_2018_2.json`.

5. Write a separate object oriented program which reads from the database.
6. Represent the above data in an interface web-based program such that the information is best represented in the form of charts, tables based on the design given.

You may use any object oriented framework such as Django, Rails, Symfony, etc. You may use any of the following languages: Python, Perl, C, C++, Java, C#. You may use a combination of these languages. You have to use git as the version control system, with commits at regular intervals, with appropriate comments. You have to write unit test cases wherever appropriate.

You will submit:

1. The JSON files.
2. The web scraper script.
3. The SQL file with all data and ER diagram.
4. Program to fetch data.
5. Web program.

You will be evaluated on:

1. The ER diagram and JSON schema (20%)
2. Script 1: Fetching data, parsing it, and entering it to a RDBMS. (20%)

3. Script 2: Web Scraper program. (20%)
4. Script 3: Script to represent data (Object oriented) and UI. (20%)
5. Standard development lifecycle (version control, documentation, etc.) (10%)
6. Unit tests. (10%)

The general rules for the exercise are below:

1. You have until 10.00 AM 22nd January 2018 to submit the exercise. Mail us the github link for your submission. Your repository must contain all the necessary parts of the submission.
2. Add a README file in the root directory of your project. This should describe the directory structure, build instructions, and the list of dependencies. Your project must build under a GNU/Linux environment.
3. If your project does not build using the dependencies you have mentioned, you will not be eligible for further rounds of the interview.
4. Plagiarism of any sort is strongly discouraged at FoV. We will disqualify any entries found to have plagiarized code.

Please share the Github URL project for your project in the email. Make sure that the last commit is before 10.00 AM 22nd January 2018. Please share the GitHub URL to work@fieldsofview.in. In the subject of the email, please add “[Application for post of Programmer]” followed by your name and affiliation. Submissions not following this format will not be evaluated.