Robert White, Karan Singh, Kevin Wu and Russell Rowell

Project Description

Web Extractor

Nowadays, Internet and websites are everywhere, we can use them for pretty much everything, watching cool videos, chat with friends, or shopping online.

The project is an implementation of a web information extractor. With a given shopping websites, the project (Web extractor) will extract some critical information and store them into a database.

We will use Walmart as our retailer example and extract information like product title, product price, product description and product id. After the extraction, we can process the data and use them for later assignment.

Table of Tasks

1. Create Program for capturing product information
   1. Application Class
      * The application class will consist of the main function for executing the program and print the information extracted from the retailer’s website.
   2. Retailer Class
      * This class will need be used to set and get all product information for the application class.
   3. Product Class
      * The Product class will be used to capture the title, price, description and various other information that will be stored in the database.
   4. Extractor Class
      * Will be used for obtaining all of the information we want about a product from the retailers website.
2. Setup Communication to database
   1. We will need to create a connection method for connecting the program to the MySQL database. Example: MySQL connector has library for connecting to a database.
3. Create database
   1. Create a database using MySQL with ten tables for the first retailer that will be implemented for the extractor.
4. Create test cases for validating database.
   1. Test cases will need to be created for all ten tables in the database.
   2. QA will need to validate the user is able to write and read from the database.
5. Create test cases for validating program.
   1. Test cases will need to be created for all functions within each class.
   2. QA will need to validate all functions are working properly.

Schedule and Allocation Chart

Week 1

* Product Description
  + Rob: UML Diagram
  + Russ: Table of tasks and Schedule
  + Kevin and Karan: Description of project

Week 2

* Application Development
  + Rob: Create Application class.
    - The application class will need a main function for executing the application. We will be using this class to also print all information stored in the database to the IDE console.
  + Russ: Create Retailer class.
    - The retailer class will contain all information for each retailer that we will extract are product information from.
  + Kevin: Create Product class.
  + Karan: Create Extractor class.

Week 3

* Database
  + Russ:
  + Karan:
* Communication
  + Rob:
  + Kevin:

Week 4

* QA
  + Rob: Create test plan for application.
  + Russ: Regression for application.
  + Kevin: Create test plan for database.
  + Karan: regression for database.

Week 5

* Product Release and Presentation