4.0 hours

Project 6

Build a Content Scraper

Why can't I submit this project yet?

You must first complete all prior Techdegree content before you can submit this project. Additionally, you cannot submit more than one project at a time.

Submit for Peer Review

Your current activity is [HTML Tables](#).

- [Instructions](#)
- [How you'll be graded](#)

Imagine you work for a price comparison website. You've been given the task to create a Node.js command line application that goes to an ecommerce site to get the latest prices and save them to a spreadsheet (CSV format). You will write the application as if the spreadsheet will be used by another application to populate a database. The application you build will run once every day. You should use npm modules to assist you in the project. You have to research and use npm packages that will help you scrape a website and create a CSV file.

Before you start

To prepare for this project you'll need to make sure you complete and understand these steps.

[1 step](#)

- **Make sure you have a GitHub account and know how to create a new repository and upload files to it. Check out the [Git Basics Course](#) or the [Github command line cheat sheet](#).**

Project Instructions

To complete this project, follow the instructions below. If you get stuck, ask a question in the community.

[7 steps](#)

- **Create a `scraper.js` file that will contain your command line application. Your project should also include a `package.json` file that includes your project's dependencies. The `npm install` command should install your dependencies.**

- **Program your scraper to check for a folder called 'data'. If the folder doesn't exist, the scraper should create one. If the folder does exist, the scraper should do nothing.**

- **Choose and use two third-party npm packages. One package should be used to scrape content from the site. The other package should create the CSV file. Be sure to research the best package to use**

(see the project resources for a link to the video about how to choose a good npm package) Both packages should meet the following requirements:

- At least 1,000 downloads
- Has been updated in the last six months

- **Program your scraper so that it visits the website http://shirts4mike.com and uses http://shirts4mike.com/shirts.php as single entry point to scrape information for 8 tee-shirts from the site, without using any hard-coded urls like http://www.shirts4mike.com/shirt.php?id=101. If you're unsure of how to get started, try googling 'node scraper' to get a feel for what a scraper is and what it does.**

- **Scraping and Saving Data:**

  - The scraper should get the price, title, url and image url from the product page and save this information into a CSV file.
  - The information should be stored in an CSV file that is named for the date it was created, e.g. `2016-11-21.csv`.
  - Assume that the the column headers in the CSV need to be in a certain order to be correctly entered into a database. They should be in this order: Title, Price, ImageURL, URL, and Time
  - The CSV file should be saved inside the 'data' folder.

- **If your program is run twice, it should overwrite the data in the CSV file with the updated information.**

- **If http://shirts4mike.com is down, an error message describing the issue should appear in the console.**

  - The error should be human-friendly, such as "There's been a 404 error. Cannot connect to the to [http://shirts4mike.com.](http://shirts4mike.com)"
  - To test and make sure the error message displays as expected, you can disable the wifi on your computer or device.

Extra Credit

To get an "exceeds" rating, you can expand on the project in the following ways:

2 steps

- **Edit your `package.json` file so that your program runs when the `npm start` command is run.**

- **When an error occurs, log it to a file named `scraper-error.log`. It should append to the bottom of the file with a time stamp and error e.g. `[Tue Feb 16 2016 10:02:12 GMT-0800 (PST)] <error message>`**

NOTE:

  - To get an "Exceeds Expectations" grade for this project, you'll need to complete **each** of the items in this section. See the rubric in the "**How You'll Be Graded**" tab above for details on how you'll be graded.

- If you're shooting for the "Exceeds Expectations" grade, it is recommended that you mention so in your submission notes.
- Passing grades are final. If you try for the "Exceeds Expectations" grade, but miss an item and receive a "Meets Expectations" grade, you won't get a second chance. Exceptions can be made for items that have been misgraded in review.

Project Resources

[External Link](#)

[Web scraper defined](#)

[External link](#)

[CSV file defined](#)

[External Link](#)

[JSDoc](#)

Need Help?

Have questions about this project? Start a discussion with the community and Treehouse staff.

[Get Help](#)