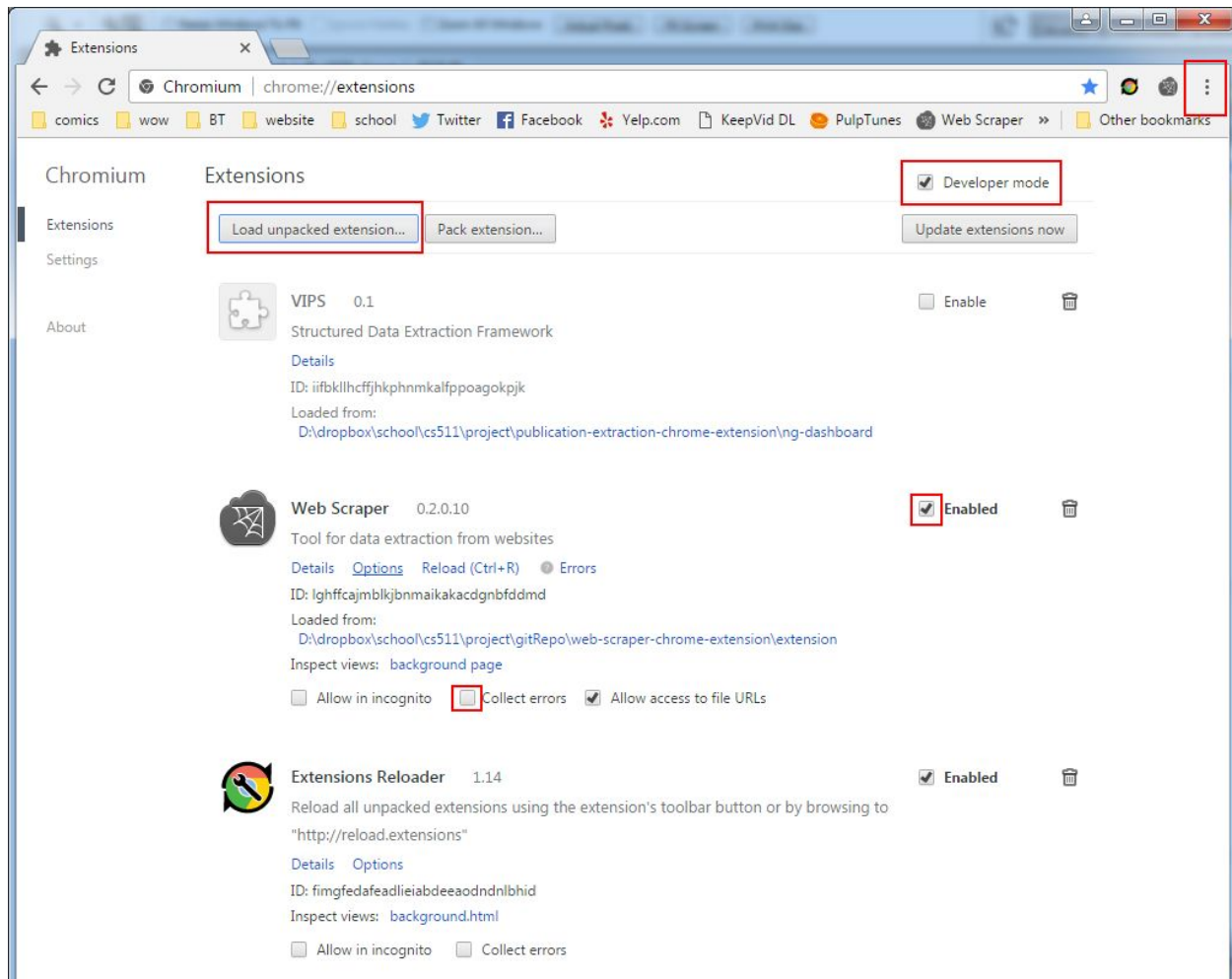
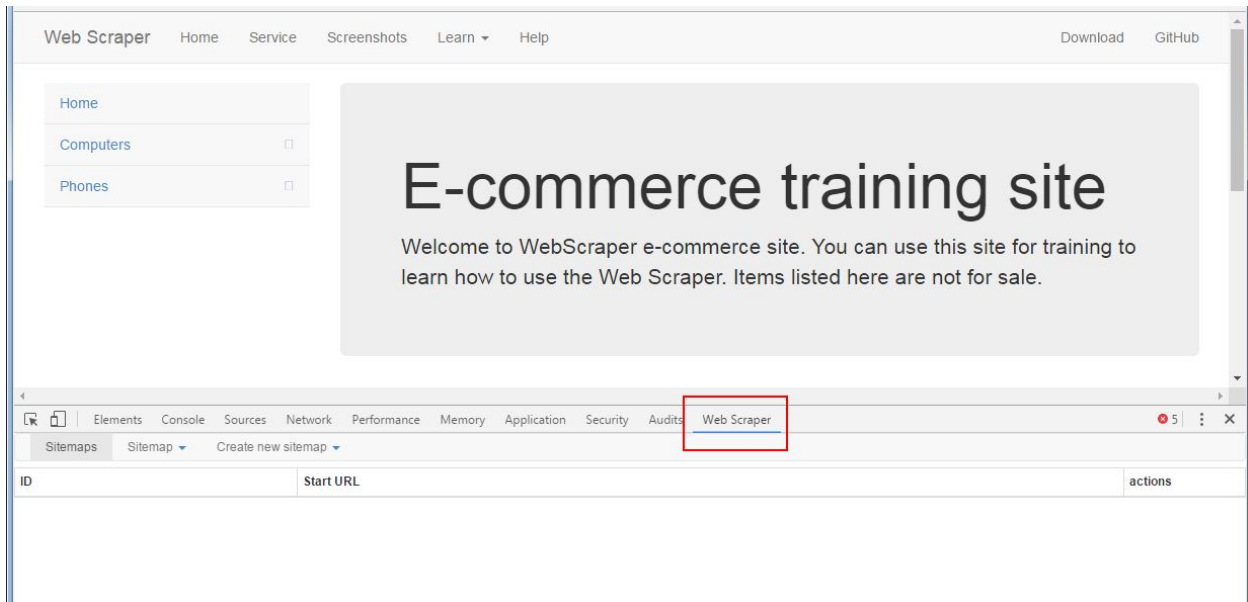


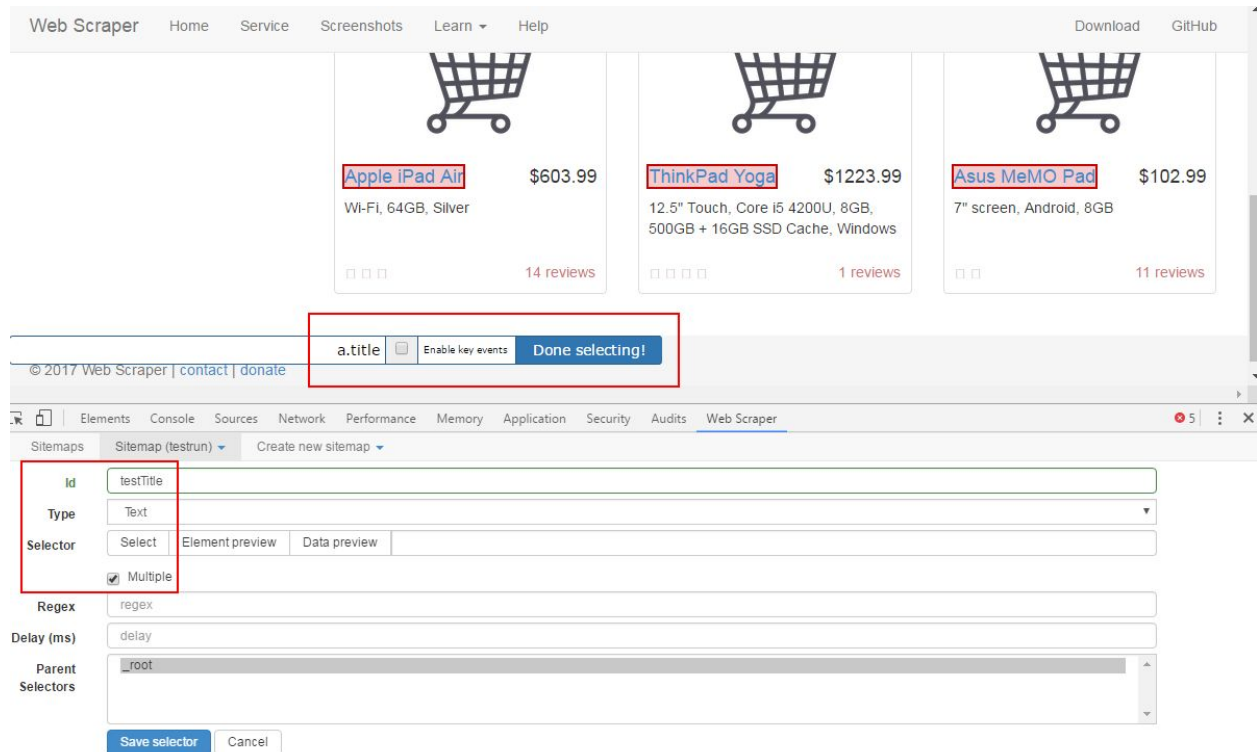
1. GIT repo: <https://github.com/csjacob2/cs511.git>
2. Add extension to Chrome.
  - a. Tools -> More tools -> Extensions
  - b. Check "Developer Mode"
  - c. Click "Load unpacked extensions"
  - d. Navigate to folder "web-scraper-chrome-extension/extension", click "ok"
  - e. Ensure "Enabled" is checked and "Collect errors" is disabled



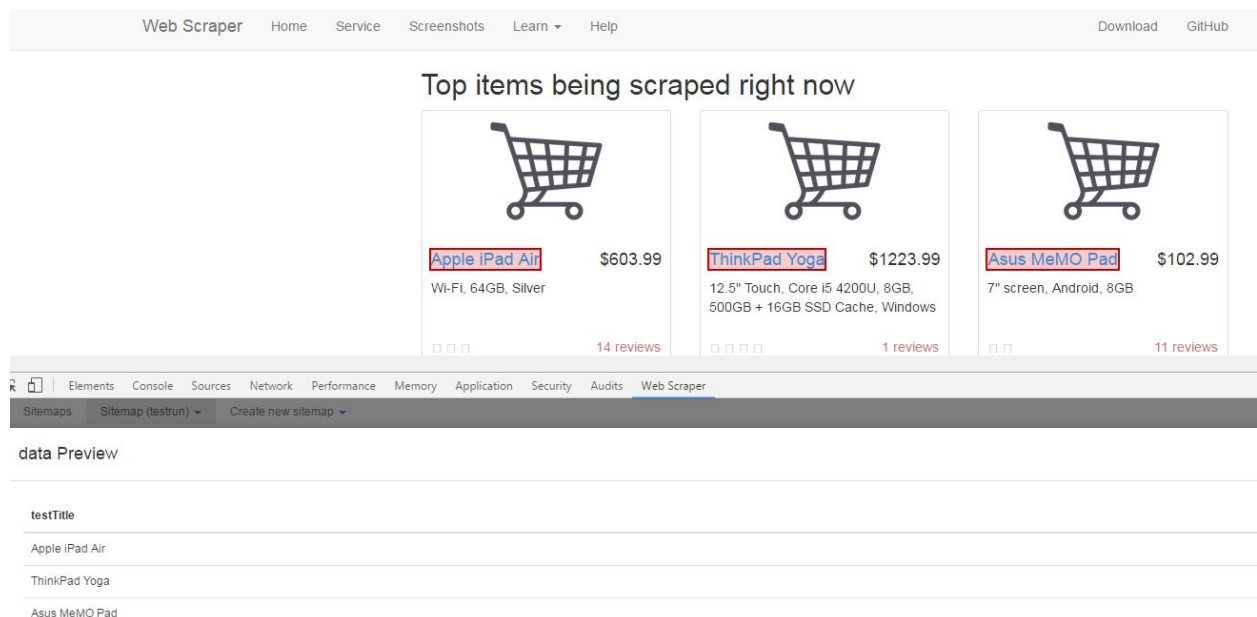
2. To use:
  - a. Navigate to test site for easy use first:  
<http://webscraper.io/test-sites/e-commerce/allinone>
  - b. Open web tools (F12), then select "Web Scraper" tab:



- c. Click “Create new sitemap -> Create sitemap”, then enter a test name and the test URL from above, then click “Create Sitemap”.
  - i. **Sitemap name: testrun**
  - ii. **Start URL: <http://webscraper.io/test-sites/e-commerce/allinone>**
- d. Click “Add new selector”, then enter in test data. Select the data you want to scrape, ideally the “Title” of the products in the test page. This is done by highlighting the product and clicking on it. You want to click on at least two so the extension can be more specific. When finished selecting, click on “Done selecting!” in the top pane. Then click “Save selector”
  - i. **Id: testTitle**
  - ii. **Type: Text**
  - iii. **Selector: Select**
  - iv. **Multiple: Checked**
  - v. **Regex: LEAVE BLANK**
  - vi. **Delay (ms): LEAVE BLANK**
  - vii. **Parent Selectors: \_root (selected)**



- e. Selecting “Element preview” in the Sitemap screen will highlight the elements that fit the selector criteria. Selecting “Data preview” will give a preliminary data scrape.



- f. To run the Scrape, select “Sitemap” (make sure your sitemap name is next to it in the main menu), then “Scrape”, then “Start scraping”. A list will be produced.

g. The data can also be exported as a CSV.

To tokenize and classify:

- a) Under “sitemap”, select “Send to Labeler”.
- b) Data will be posted to the dev tools window in JSON format.