Assignment 2

CSCI 3172 – Web-Centric Computing

Christian Gilbert-Paszkowski

B00726153

**Note: For the sake of time, I have only searched for results in the products section of the website (not the services, etc.) and used the first child category of every parent category to reduce the number of results and not retrieved products beyond the first page of results. The fully functional alternative is included but commented out.

**Note: Not all products on the Best Buy website have a corresponding code/part number and were therefore not included in some results.

The plan for how my web crawler is constructed is as follows:

1. Using the web address of the Best Buy Canada website, retrieve the homepage html.
2. Send the html to my first method which looks for all the product categories and the links are stored in a queue and passed.
3. The queue is passed to a second method which recursively looks for any additional sub-categories for each link in the queue. This repeats until no more sub-categories are found and the final link is put in a queue which is has all the product list pages for each category.
4. This queue of product lists is then passed to a method which collects links for each individual product in each page.
5. The products are then visited individually and the raw data is extracted, formatted and then printed.