Homework 9

Due: Wednesday 10/19 at 11:59pm

Submit the following with handin as assignment hwk9.

1. (10 points) Complete the “dead link finder” you started in lab 6. Your program should take a URL from the command line to serve as the basis of its search. It should examine all links on that page to make sure that they refer to valid webpages. In addition, all links to the same “site”, defined as being on the same host and starting with the same path as the base URL, should be examined recursively. (Thus, if you start at http://faculty.knox.edu/dbunde, then all reachable pages within dbunde would be examined for dead links as well.) Visit these sites in a breadth-first search (BFS) order. To prevent overwhelming the server, your program should include a half second delay between requests; use Thread.sleep(500) to add this delay. (Note that you’ll need to catch the exception InterruptedException for this call.)

As it is running, your program should print out each URL that it is examining. Once it has finished, it should print a summary including the number of pages examined and a list of dead links found. (For these, list the page on which they were found and the identity of the dead link.)