

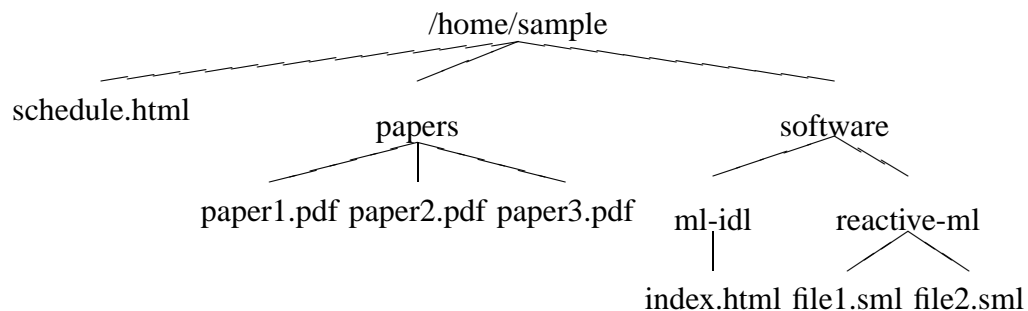
This homework requires some familiarity with HTML. See the web classic

*<http://archive.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html>*

for a quick introduction.

The purpose of this homework is to develop a script to walk down a directory hierarchy, adding in “table of contents” HTML files in every directory. Why is this useful, you ask? On web servers with virtual directories, this is a quick way to present a hierarchy of files (for example, if you want people to browse the source code to your massive 100+ source files app).

Here is the idea. Assume you have the following directory hierarchy:



You want to add a file *index.html* in each of those directories, containing a list of all the files and subdirectories in that directory. Each file and directory name should be a link to respectively the actual file or the file *index.html* in that subdirectory. Moreover, if the file *index.html* does not exist in a subdirectory, you want to create it, recursively.

Here is a sample *index.html* created for the toplevel directory above:

```
<html>
  <head>
    <title>Content of /home/sample</title>
  </head>
  <body>
    <ul>
      <li><a href="schedule.html">schedule.html</a></li>
      <li><a href="papers/index.html">papers/</a></li>
      <li><a href="software/index.html">software/</a></li>
    </ul>
  </body>
</html>
```

A few things are worth noticing about the above:

- the title contains the full path to the directory being indexed;
- an HTML link for a file (for example, *schedule.html*) is of the form

```
<li><a href="file">file</a></li>
```

- an HTML link for a subdirectory (for example, *papers*) is of the form

```
<li><a href="dir/index.html">dir</a></li>
```

(Roughly speaking, a link in HTML is interpreted as follows: when you select the text between `<a>` and `</a>` on the web page, the browser loads the file given by *href*.)

For this homework, you have to write a script *buildindex* taking no arguments, and building a file *index.html* in the current directory, according to the above specifications, as well as building a file *index.html* in every subdirectory that applies. (You can use the sample *index.html* file as a template.) Some remarks:

- For every subdirectory you encounter, if that directory contains no file called *index.html*, you should recursively create an index file in that directory; if there already is a file *index.html*, *do not* recursively index the content of that directory.
- If you do not have write permission in a subdirectory (i.e. you can't create the file *index.html*), do not recursively index that subdirectory. Ideally, don't link to it either, unless there is already a file *index.html* in it.
- You can write more than one script if it helps you (in which case, submit them all to get due credit).
- Don't worry about loops in the directory hierarchy due to symbolic links.

You can use *lynx*, the WWW text browser, to see the results of your script. The command *lynx index.html* (or whatever the path is to the starting index file) will allow you to browse your generated indices.

The *~cs214/HW1* directory on babbage will contain the sample tree directory above, along with an "after" tree directory showing the files that have been created. Your script should build something like that.

Submit all your scripts according to the instructions on the web page.