

## CS 288 2018S Section 006

### Homework 07 - Milestone 1

**Due:** At 11:45pm on Wednesday April 25<sup>th</sup>.

Write a UNIX Shell script that repeats the following steps every hour for each state among a select number of states:

1. Set a shell variable to the current date & time using the format:  
`yyyy-mm-dd-hh-mm-ss-AB.html`  
where AB is the 2-letter state abbreviation.
2. Use `curl` or `wget` to download a Web page containing the current weather conditions for a major city in the given state and rename it to the string you've obtained in the previous step.
3. Use TagSoup, or any similar tool, to generate a `.xhtml` that corresponds to the downloaded `.html` file. TagSoup is written in Java. As such, you have to have a Java JRE installed on your system in order to run it. The command to run TagSoup is:  
`java -jar tagsoup-1.2.1.jar --files yyyy-mm-dd-hh-mm-ss-AB.html`  
Use `curl` or `wget` to download the `tagsoup-1.2.1.jar` file if it isn't present in the current directory.
4. Run a Python script that uses the `xml.dom.minidom` module to traverse the `.xhtml` document and extract the relevant values.
5. Store the extracted values in a `.csv` file as comma-separated values based on the Sample Input/Output given below.

Your deliverables are the shell and Python scripts. Before submitting your solutions via Moodle, zip the scripts and name the archive, if your name is Harry Houdini, for example, `HW7a_HarryHoudini.zip`. You may use `gzip` if that's more convenient.

#### Sample Input/Output:

The following rendering of the `.html` page for New York City, Central Park, for example:



should produce the following output where the first line is reserved as a header containing a list of field names:

```
state,city,weather,temperature,humidity,pressure  
NY,New York,Clear,48,37,30.2
```