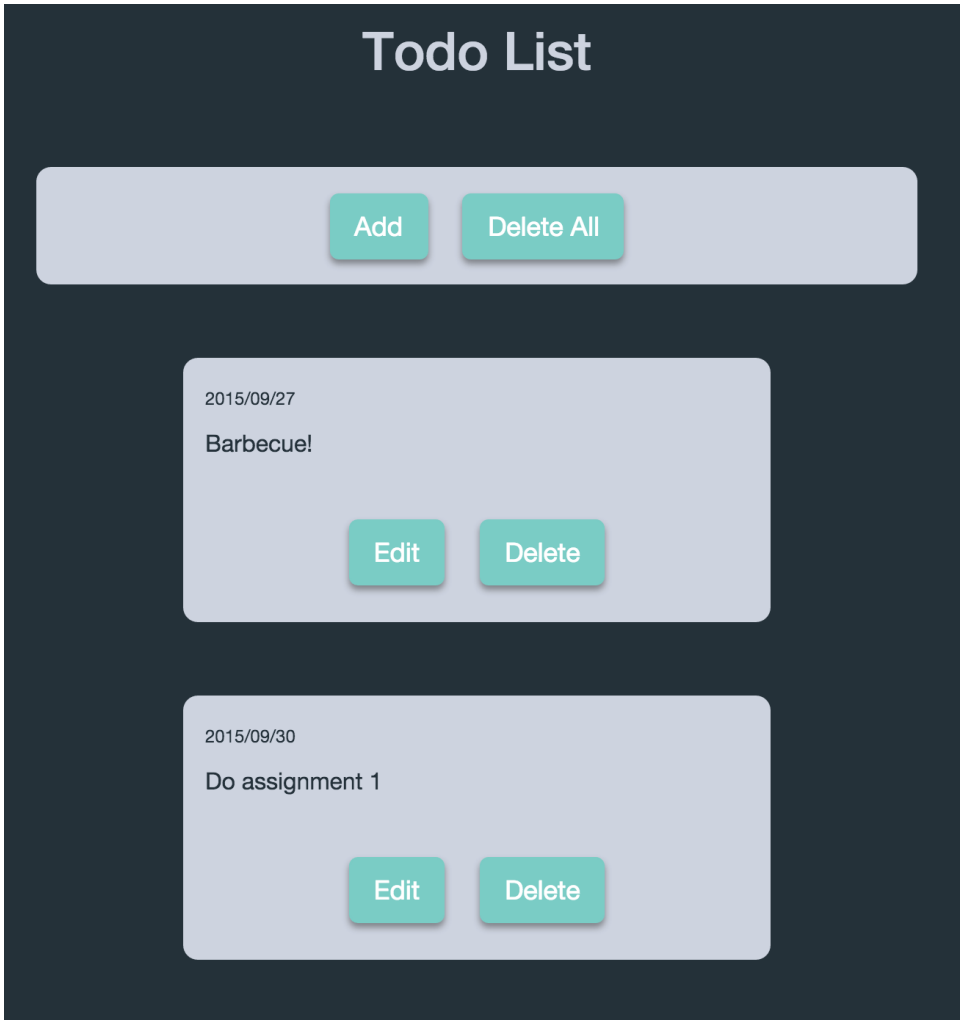
Social Computing Application Design
Assignment 1 : Tool Practice

Honor Code

Any cheating will be handled seriously in compliance with the university rules. All assigned work is expected to be individual, except where explicitly written otherwise (e.g., term project). You are encouraged to discuss with your classmates; however, what you hand in should be your own work.

1. Write a “To-do-list” web page that can record what the user has to do in the future. This web page should allow the user to create and edit items with the date and the content, and also delete items. Below is an example of what it may look like, and you should design your own page with a distinct style on your own. You’re encouraged to explore using CSS style, jQuery or the other technical skills to beautify or enhance your visual design. Bonus credits may be given to the extra work.



2. Use Python to extract n-word sequences from “big-sequence.txt” which is posted on iLMS. Count the number of occurrence for each n-word sequence, then show the top five sequences with the highest counts for each n-word sequence (n = 1 to 5). An n-word sequence is a sequence that is made up of n words. For example, for the sentence “I will attend the meeting tomorrow”, following shows its n-word sequences:

- 1-word sequence: I, will, attend, the, meeting, tomorrow
- 2-word sequence: I will, will attend, attend the, the meeting, meeting tomorrow
- 3-word sequence: I will attend, will attend the, attend the meeting, the meeting tomorrow
- 4-word sequence: I will attend the, will attend the meeting, attend the meeting tomorrow
- 5-word sequence: I will attend the meeting, will attend the meeting tomorrow

The output is look like below:

```
----Top five 1-words sequences----
the     72009
of      39540
and     36671
to      28130
a       19920
----Top five 2-words sequences----
of the  11791
in the  5486
to the  4117
and the         2913
on the  2105
----Top five 3-words sequences----
the United States       342
one of the      313
out of the      227
I don t         209
of the United   196
----Top five 4-words sequences----
of the United States    170
at the same time        83
as a result of  76
the commander in chief  67
the end of the  65
----Top five 5-words sequences----
History of the United States    40
in the region of the    26
Project Gutenberg Literary Archive Foundation   23
the other side of the   22
on the same lines as    22
```

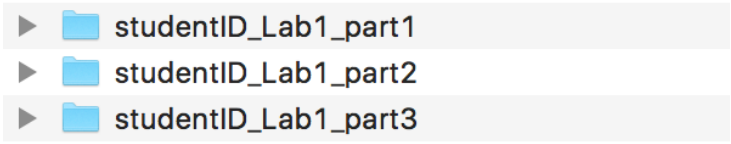
3. Use Python to write an HTML parser. This program can parse a given web site, extract information items from the web page and store the items in a data structure, then output the results as an .html file that displays the data structure. Parse the comments of this PTT post (https://www.ptt.cc/bbs/StupidClown/M.1443521969.A.C78.html). You need to extract all of the comments including the status (upvote/downvote/neutral), commentators, dates, titles, and contents (You don't need to represent any pictures on the page). The HTML output may look like the sample below.

推 144
噓 5
→ 14

推 anita811038: 原來後面有鏡子XDDDD 09/29 18:27
推 chuegou: 鏡子wwwwwwwwww 09/29 18:29
推 wangsynnex: 所以男友是她的左右手 09/29 18:38
→ Doughnut0425: 我真的以為是反串XDD 09/29 18:44
推 blueskymaple: 真討厭我男朋友就是鏡子，而且眼裡只有我XD 09/29 18:51
推 tp6d93vup: 好恐怖!! 09/29 18:56
推 tsubakichen: 五樓講的好恐怖… 09/29 18:58
推 misakimisaki: 手機就是男友啦 09/29 19:16
→ lunhsuan: 淡淡的哀傷 09/29 19:21
推 louisyang: 鏡子完全露餡XDDDD 09/29 19:22
推 qwer35736: 哈哈何苦呢 09/29 19:29
推 waynehandsom: 苦命女子 09/29 19:33
推 littlelinsyu: 這魯的很入味，魯到有點哀傷了 09/29 19:36
推 hungryuu: 蝦妹XDDDD 09/29 19:36
推 elvisleeee: 本文中肯必推 虛偽啊 09/29 19:37
推 kevinhsu113: 魯蛇何必為難魯蛇 09/29 19:45
推 fvbcgn124578: 有點哀傷 09/29 19:49
推 Ying880118: 已朝聖 09/29 19:59
推 sw953122: 感覺反串 09/29 20:01
推 sky85924: #秀恩愛 XDD 09/29 20:05
推 hugo520: 男友只有她看得見！？ 09/29 20:08
推 ShieChi: 呵呵 09/29 20:08

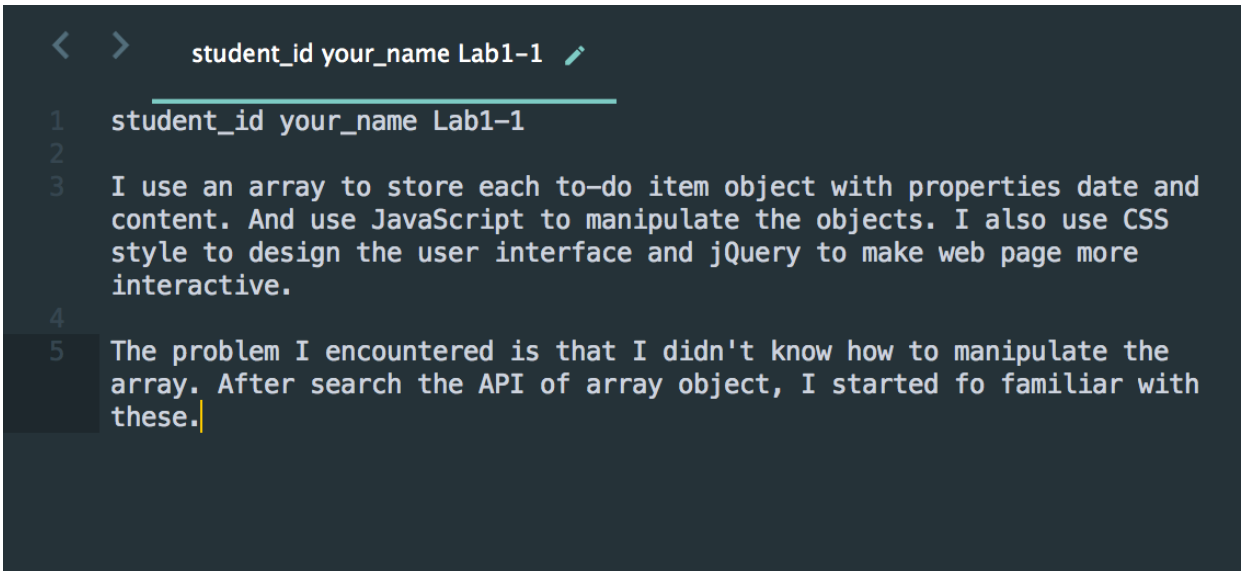
Notice:

1. **Deadline : 2015/10/11 23:59. If you submit in 10/12 00:00 ~ 23: 59, you'll get partial credits (70% of the original score). No credits if submitting afterward.**



2. To submit your assignment, you should follow the form like below:
   Put the source code, "Readme" document in each part of assignment. Then zip all of the folders to a **zip** file named as "id_lab1.zip" and upload to iLMS.

3. For each source code file, you have to add the comments to explain your code.



4. The "Readme" files have to include the brief explanation of your work, the problem you encountered and how did you solve the problem.