

CS 101 - Assignment 4 - Working with html  
Fall 2018

Algorithm Due : **Oct 7th, 2018**

Program Due : **Oct 14th, 2018**

**All work submitted must be your own.**

**Deliverables :**

You must use functions to modularize your work in a logical way.

You should use exception handling where necessary as well.

All submitted work must be your own.

50 points off for programs that crash on expected input.

## HTML Files

Hypertext Markup Language is the basic language used for web pages. It is just a text file, so we can use python to read an html. The markup in html uses tags for sections and formatting. Tags have an opening and a closing tag. For instance, in the <head> tag, there is normally a title tag that gives the title of the web page.

Example

```
<html>
  <head>
    <title>A sample web page</title>
  </head>
  <body>
    <p>This is a paragraph in the body.</p>
  </body>
</html>
```

The indentation in the above section is only for readability. There may not be any indentation. The important part.

## Headers

The header tag in html indicates the head of a section, and has formatting to make it stand out before its section. <h1> is the highest level, with lower tags <h2>, <h3>, <h4> etc.

```
<h1>My Heading</h1>
<p>Big section of information</p>
<h2>Smaller sub heading</h2>
<p>Big section of information</p>
<h1>Another heading</h1>
<p>Big section of information</p>
```

## Requirements

- Ask user for a valid input file. Must respond to File Not Found Error appropriately.
- Ask user for file to output to. Must respond to IO Errors appropriately.
- Read through input html file for any header tags and export the result as unordered lists to the output file.

## Example

Enter HTML file to open ==> invalidfile.txt

File not found invalidfile.txt

Enter HTML file to open ==> websample1.html

Enter name of html file to write to ==> out\*bad.txt

IO Error out\*bad.txt

Enter name of html file to write to ==> output.html

Successfully output to file output.html

Enter HTML file to open ==>