Lab 9

Instructions:

Provide your solutions in a file named **lab9.py**. Make sure you run the doctest.

Problems:

In HTML headings are indicated by using the tags `h1,h2,…h6`. For example:

```
<h3>This is a heading</h3>
```

1. Write class **HeadingParser** that is a subclass of the HTMLParser class. It will find and collect the contents of all the headings in an HTML file fed to it. The parser works by identifying when a heading tag has been encountered and setting a boolean variable in the class to indicate that. When the data handler for the class is called and the boolean in the class indicates that a heading is currently open, the data inside the heading is added to a list. Finally, when a closing heading tag is encountered the boolean variable is unset. To implement this parser you will need to write the following methods (some override HTMLParser methods):

    a. **__init__:** calls the HTMLParser __init__, initializes an empty list and sets the boolean variable to False
    b. **handle_starttag** : If the tag that resulted in this method being called is a heading, the heading indicator should be set.
    c. **handle_endtag** : If the tag that resulted in this method being called is a heading, the heading indicator should be unset.
    d. **handle_data** : If the parser is currently inside a heading, then the data should be added to the list of headings contents. Make sure that you strip any extra spaces or newlines off the contents of the heading before adding it to the list.
    e. **getheadings()**: returns the list of headings

2. Write a test function **testHP** (url) that given a url (as a string), opens it, feeds the html to a heading parser and returns the obtained headings.

Together your code should make the following work. Note that the headings returned may vary somewhat depending on the current state of the web page.

```
>>>
testHP('http://www.warnerbros.com/archive/spacejam/movie/cmp/sitemap.h
tml')
[]
>>> testHP('http://www.pmichaud.com/toast/')
['Strawberry Pop-Tart Blow-Torches', 'Author', 'Abstract',
'Introduction', 'Materials Used', 'Experiment Preparation', 'The
Experiment and Observations', 'Summary and Recommendations',
'Acknowledgements', 'Followup Comments']
>>> testHP('http://www.kli.org/')
['Home Page', 'The Klingon Language Institute', "Registration for
qep'a' cha'maH cha'DIch (July 23rd - 25th, 2015) is now open!", 'Join
the KLI today!', 'Klingon in a nutshell', 'Learn Klingon!']
>>> testHP('http://home.mcom.com/home/welcome.html')
['For More Information about Netscape...', 'Welcome to the world of
Netscape', 'and the Internet.', 'Enjoy!']
>>>
testHP('http://usatoday30.usatoday.com/sports/baseball/sbfant.htm')
['Fantasy baseball home page', 'Daily Best/Worst', "Baseball Weekly's
John Hunt", 'Player position eligibility', 'Player ratings and
projections', 'Other links of interest']
>>>
```