

## Exercise 1 <<20 points>>

You are provided with an XML file with information about products in the store. This is a sample file:

```
<?xml version="1.0"?>
<shop name="Uncle SNAFU's Computer Store">
  <item id="1" stock="7" discounted="y">
    <name>FOOBAR keyboard</name>
    <price>39.99</price>
  </item>
  <item id="2" stock="15" discounted="y">
    <name>Grouchobyte hard drive</name>
    <price>149.99</price>
  </item>
  <item id="3" stock="81">
    <name>Wumpus repellent</name>
    <price>49.99</price>
  </item>
</shop>
```

For each product, the following information is stored:

- *Product ID*: Each product is assigned a unique id.
- *Items in stock*: The number of products immediately available in the store.
- *Name*: The product's name.
- *Price*: The product's price.
- *Discounted product*: Indicates whether this product is currently offered at a discounted price (-15% discount).

Write a Python program that reads an inventory file and determines the total value of all the items in stock. In other words, for each product, multiply its price by the stock (applying a discount if necessary), and sum all the values. The program's output should be the total value.

The output for the above sample file should be:

**6199.50**

The output for the provided **inventory.xml** file should be:

**34202.42**

## Exercise 2 <<20 points + 5 bonus points>>

In this problem you will process an XML document acquired from an NPR RSS news feed to find a list of frequent news keywords and the corresponding stories:

Use the XML file nprNews.xml from the lab5.zip file to develop your program.

Find the five named entities that appear most frequently in the title and description elements in this news feed. A named entity is a word that identifies a person, place, location, organization, or other entity. (Named entities can also be phrases, such as "White House", but we will only consider single words here).

Identifying named entities is a task until itself, but we will use this simple metric: A named entity is any word that is capitalized.

So for example, "Kennedy" and "National" are keywords, but "national" is not.

Again, note that you need to only search for keywords in the "title" and "description" elements of this document:

```
<item>
  <title>Kennedy Dies; Universal Care Dream Lives On</title>
  <description>Kennedy's death leaves behind a rich legacy of health
legislation but his drive for universal health coverage unfulfilled.</description>

  <pubDate>Wed, 26 Aug 2009 09:10:00 -0400</pubDate>

  <link>http://www.npr.org/templates/story/story.php?storyId=112236406&ft=1&
mp;f=1001</link>

  <guid>http://www.npr.org/templates/story/story.php?storyId=112236406&ft=1&
mp;f=1001</guid>
  <content:encoded><![CDATA...]></content:encoded>
</item>
```

After you have found the five most common named entity keywords, list each keyword with the number of times it appears.

Your output for the sample input provided should look something like this:

```
Kennedy 25
Sen 9
Senate 8
```

Ted 7  
Senator 6

**Hints:**

You can use the ElementTree API to complete this problem.

Test individual components in the interactive python mode first. Here are the things that might be worth testing separately

- Extracting the text from the title and description elements of the xml file.
- Use regular expressions to find all the capitalized words in a string.
- Write a function extractMax that finds the max value in a dictionary, returns a tuple containing the max value and the corresponding key (and perhaps deletes the entry corresponding to the max key from the dictionary)

**Bonus (5pts)**

For each of the top 5 entity keywords, write the title of the news item where it appears. Your output should look something like this:

*Kennedy 25*

*Kennedy Dies; Universal Care Dream Lives On  
Edward Kennedy, Senate's Liberal Lion, Dies  
Leahy: Kennedy Was A Senator's Senator  
Obama: Honored To Call Teddy A Friend  
Sen. Kennedy's Talent: Reaching Across The Aisle  
Influential Sen. Ted Kennedy Dies  
Sen. Ted Kennedy, Last Surviving Kennedy Son Dies  
Health Care Debate Will Miss Sen. Ted Kennedy  
Republicans Join Democrats In Mourning Kennedy  
Sen 9*

*Edward Kennedy, Senate's Liberal Lion, Dies  
Leahy: Kennedy Was A Senator's Senator  
Obama: Honored To Call Teddy A Friend  
Sen. Kennedy's Talent: Reaching Across The Aisle  
Bernanke Must Decide When To Raise Interest Rates  
Influential Sen. Ted Kennedy Dies  
Sen. Ted Kennedy, Last Surviving Kennedy Son Dies  
Health Care Debate Will Miss Sen. Ted Kennedy  
Republicans Join Democrats In Mourning Kennedy  
Senate 8*

*Edward Kennedy, Senate's Liberal Lion, Dies  
Leahy: Kennedy Was A Senator's Senator  
Sen. Kennedy's Talent: Reaching Across The Aisle  
Bernanke Must Decide When To Raise Interest Rates  
Health Care Debate Will Miss Sen. Ted Kennedy*

*Ted 7*

*Edward Kennedy, Senate's Liberal Lion, Dies*

*Obama: Honored To Call Teddy A Friend*

*Sen. Kennedy's Talent: Reaching Across The Aisle*

*Influential Sen. Ted Kennedy Dies*

*Sen. Ted Kennedy, Last Surviving Kennedy Son Dies*

*Health Care Debate Will Miss Sen. Ted Kennedy*

*Senator 6*

*Leahy: Kennedy Was A Senator's Senator*

*Influential Sen. Ted Kennedy Dies*

*Sen. Ted Kennedy, Last Surviving Kennedy Son Dies*

*Health Care Debate Will Miss Sen. Ted Kennedy*