from urllib.request import urlopen from bs4 import BeautifulSoup html = urlopen('http://www.pythonscraping.com/pages/warandpeace.html') bs = BeautifulSoup(html, 'html.parser') print(bs)

In [1]:
```
from urllib.request import urlopen
from bs4 import BeautifulSoup
html = urlopen('http://www.pythonscraping.com/pages/warandpeace.html')
bs = BeautifulSoup(html, "html.parser")
```

In [2]:
```
nameList = bs.findAll('span', {'class': 'green'})
for name in nameList:
    print(name.get_text())
```

```
Anna
Pavlovna Scherer
Empress Marya
Fedorovna
Prince Vasili Kuragin
Anna Pavlovna
St. Petersburg
the prince
Anna Pavlovna
Anna Pavlovna
the prince
the prince
the prince
Prince Vasili
Anna Pavlovna
Anna Pavlovna
the prince
Wintzingerode
King of Prussia
le Vicomte de Mortemart
Montmorencys
Rohans
Abbe Morio
the Emperor
the prince
Prince Vasili
Dowager Empress Marya Fedorovna
the baron
Anna Pavlovna
the Empress
the Empress
Anna Pavlovna's
Her Majesty
Baron
Funke
The prince
Anna
Pavlovna
the Empress
The prince
Anatole
the prince
The prince
Anna
Pavlovna
Anna Pavlovna
```

In [3]:
```
titles = bs.find_all(['h1', 'h2','h3','h4','h5','h6'])
print([title for title in titles])
```

```
[<h1>War and Peace</h1>, <h2>Chapter 1</h2>]
```

In [8]:
```
allText = bs.find_all('span', {'class':{'green', 'red'}})
print([text for text in allText])
```

```
[<span class="red">Well, Prince, so Genoa and Lucca are now j
ust family estates of the
Buonapartes. But I warn you, if you don't tell me that this m
eans war,
if you still try to defend the infamies and horrors perpetrat
ed by
that Antichrist- I really believe he is Antichrist- I will ha
ve
nothing more to do with you and you are no longer my friend,
 no longer
my 'faithful slave,' as you call yourself! But how do you do?
I see
I have frightened you- sit down and tell me all the news.</sp
an>, <span class="green">Anna
Pavlovna Scherer</span>, <span class="green">Empress Marya
Fedorovna</span>, <span class="green">Prince Vasili Kuragin</
span>, <span class="green">Anna Pavlovna</span>, <span class
="green">St. Petersburg</span>, <span class="red">If you have
nothing better to do, Count [or Prince], and if the
prospect of spending an evening with a poor invalid is not to
o
terrible, I shall be very charmed to see you tonight between
 7 and 10-
Annette Scherer.</span>, <span class="red">Heavens! what a vi
rulent attack!</span>, <span class="green">the prince</span>,
<span class="green">Anna Pavlovna</span>, <span class="red">F
irst of all, dear friend, tell me how you are. Set your frien
d's
mind at rest,</span>, <span class="red">Can one be well while
suffering morally? Can one be calm in times
like these if one has any feeling?</span>, <span class="gree
n">Anna Pavlovna</span>, <span class="red">You are
staying the whole evening, I hope?</span>, <span class="red">
And the fete at the English ambassador's? Today is Wednesday.
I
must put in an appearance there,</span>, <span class="green">
the prince</span>, <span class="red">My daughter is
coming for me to take me there.</span>, <span class="red">I t
hought today's fete had been canceled. I confess all these
festivities and fireworks are becoming wearisome.</span>, <sp
an class="red">If they had known that you wished it, the ente
rtainment would
have been put off,</span>, <span class="green">the prince</sp
an>, <span class="red">Don't tease! Well, and what has been d
ecided about Novosiltsev's
dispatch? You know everything.</span>, <span class="red">What
can one say about it?</span>, <span class="green">the prince
</span>, <span class="red">What has been decided? They have d
ecided that
Buonaparte has burnt his boats, and I believe that we are rea
dy to
burn ours.</span>, <span class="green">Prince Vasili</span>,
 <span class="green">Anna Pavlovna</span>, <span class="gree
n">Anna Pavlovna</span>, <span class="red">Oh, don't speak to
me of Austria. Perhaps I don't understand
things, but Austria never has wished, and does not wish, for
 war.
She is betraying us! Russia alone must save Europe. Our graci
ous
```

sovereign recognizes his high vocation and will be true to i
t. That is
the one thing I have faith in! Our good and wonderful soverei
gn has to
perform the noblest role on earth, and he is so virtuous and
 noble
that God will not forsake him. He will fulfill his vocation a
nd
crush the hydra of revolution, which has become more terrible
than
ever in the person of this murderer and villain! We alone mus
t
avenge the blood of the just one.... Whom, I ask you, can we
 rely
on?... England with her commercial spirit will not and cannot
understand the Emperor Alexander's loftiness of soul. She has
refused to evacuate Malta. She wanted to find, and still seek
s, some
secret motive in our actions. What answer did Novosiltsev ge
t? None.
The English have not understood and cannot understand the
self-abnegation of our Emperor who wants nothing for himself,
but only
desires the good of mankind. And what have they promised? Not
hing! And
what little they have promised they will not perform! Prussia
has
always declared that Buonaparte is invincible, and that all E
urope
is powerless before him.... And I don't believe a word that H
ardenburg
says, or Haugwitz either. This famous Prussian neutrality is
 just a
trap. I have faith only in God and the lofty destiny of our a
dored
monarch. He will save Europe!, I thi
nk,, the prince, <span clas
s="red">that if you had been
sent instead of our dear Wintzingerode</s
pan> you would have captured the
King of Prussia's consent by assau
lt. You are so eloquent. Will you
give me a cup of tea?, Wintzingero
de, King of Prussia, <span
 class="red">In a moment. A propos,, <span class="re
d">I am
expecting two very interesting men tonight, <span class="gree
n">le Vicomte de Mortemart,
who is connected with the Montmorencys</s
pan> through the Rohans, one of
the best French families. He is one of the genuine emigres, t
he good
ones. And also the Abbe Morio. Do
 you know that profound thinker? He
has been received by the Emperor.
 Had you heard?, le Vicomte de Mor
temart, Montmorencys, <span
class="green">Rohans, Abbe Morio</
span>, the Emperor, <span class="r
ed">I shall be delighted to meet them,, <span class="g
reen">the prince, But tell me,, is it true that the Dowager Empress want
s Baron Funke
to be appointed first secretary at Vienna? The baron by all a

```
ccounts
is a poor creature.</span>, <span class="green">Prince Vasili
</span>, <span class="green">Dowager Empress Marya Fedorovna
</span>, <span class="green">the baron</span>, <span class="g
reen">Anna Pavlovna</span>, <span class="green">the Empress</
span>, <span class="red">Baron Funke has been recommended to
 the Dowager Empress by her
sister,</span>, <span class="green">the Empress</span>, <span
class="green">Anna Pavlovna's</span>, <span class="green">Her
Majesty</span>, <span class="green">Baron
Funke</span>, <span class="green">The prince</span>, <span cl
ass="green">Anna
Pavlovna</span>, <span class="green">the Empress</span>, <spa
n class="red">Now about your family. Do you know that since y
our daughter came
out everyone has been enraptured by her? They say she is amaz
ingly
beautiful.</span>, <span class="green">The prince</span>, <sp
an class="red">I often think,</span>, <span class="red">I oft
en think how unfairly sometimes the
joys of life are distributed. Why has fate given you two such
splendid
children? I don't speak of <span class="green">Anatole</span
>, your youngest. I don't like
him,</span>, <span class="green">Anatole</span>, <span class
="red">Two such charming children. And really you appreciate
them less than anyone, and so you don't deserve to have them.
</span>, <span class="red">I can't help it,</span>, <span cla
ss="green">the prince</span>, <span class="red">Lavater would
have said I
lack the bump of paternity.</span>, <span class="red">Don't j
oke; I mean to have a serious talk with you. Do you know I
am dissatisfied with your younger son? Between ourselves</spa
n>, <span class="red">he was mentioned at Her
Majesty's and you were pitied....</span>, <span class="gree
n">The prince</span>, <span class="red">What would you have m
e do?</span>, <span class="red">You know I did all
a father could for their education, and they have both turned
out
fools. Hippolyte is at least a quiet fool, but Anatole is an
 active
one. That is the only difference between them.</span>, <span
 class="red">And why are children born to such men as you? If
you were not a
father there would be nothing I could reproach you with,</spa
n>, <span class="green">Anna
Pavlovna</span>, <span class="red">I am your faithful slave a
nd to you alone I can confess that my
children are the bane of my life. It is the cross I have to b
ear. That
is how I explain it to myself. It can't be helped!</span>, <s
pan class="green">Anna Pavlovna</span>]
```

In [4]:
```
nameList = bs.find_all(text='the prince')
print(len(nameList))
```

```
7
```

In [5]:
```
title = bs.find_all(id='title', class_='text')
print([text for text in allText])
```

```
[]
```

In [6]:
```
from urllib.request import urlopen
from bs4 import BeautifulSoup
```

```
html = urlopen('http://www.pythonscraping.com/pages/page3.html')
bs = BeautifulSoup(html, 'html.parser')

for child in bs.find('table',{'id':'giftList'}).children:
    print(child)
```

```
<tr><th>
Item Title
</th><th>
Description
</th><th>
Cost
</th><th>
Image
</th></tr>


<tr class="gift" id="gift1"><td>
Vegetable Basket
</td><td>
This vegetable basket is the perfect gift for your health conscious (or overweight) friends!
<span class="excitingNote">Now with super-colorful bell peppers!</span>
</td><td>
$15.00
</td><td>
<img src="../img/gifts/img1.jpg"/>
</td></tr>


<tr class="gift" id="gift2"><td>
Russian Nesting Dolls
</td><td>
Hand-painted by trained monkeys, these exquisite dolls are priceless! And by "priceless," we mean "extremely expensive"! <span class="excitingNote">8 entire dolls per set! Octuple the presents!</span>
</td><td>
$10,000.52
</td><td>
<img src="../img/gifts/img2.jpg"/>
</td></tr>


<tr class="gift" id="gift3"><td>
Fish Painting
</td><td>
If something seems fishy about this painting, it's because it's a fish! <span class="excitingNote">Also hand-painted by trained monkeys!</span>
</td><td>
$10,005.00
</td><td>
<img src="../img/gifts/img3.jpg"/>
</td></tr>


<tr class="gift" id="gift4"><td>
Dead Parrot
</td><td>
This is an ex-parrot! <span class="excitingNote">Or maybe h
```

```
This is an ex-parrot! <span class= excitingNote >Or maybe h
e's only resting?</span>
</td><td>
$0.50
</td><td>
<img src="../img/gifts/img4.jpg"/>
</td></tr>


<tr class="gift" id="gift5"><td>
Mystery Box
</td><td>
If you love suprises, this mystery box is for you! Do not pla
ce on light-colored surfaces. May cause oil staining. <span c
lass="excitingNote">Keep your friends guessing!</span>
</td><td>
$1.50
</td><td>
<img src="../img/gifts/img6.jpg"/>
</td></tr>
```

In [26]:

```
from urllib.request import urlopen
from bs4 import BeautifulSoup

html = urlopen('http://www.pythonscraping.com/pages/page3.htm
l')
bs = BeautifulSoup(html, 'html.parser')

for sibling in bs.find('table', {'id':'giftList'}).tr.next_si
blings:
    print(sibling)
```

```
<tr class="gift" id="gift1"><td>
Vegetable Basket
</td><td>
This vegetable basket is the perfect gift for your health con
scious (or overweight) friends!
<span class="excitingNote">Now with super-colorful bell peppe
rs!</span>
</td><td>
$15.00
</td><td>
<img src="../img/gifts/img1.jpg"/>
</td></tr>


<tr class="gift" id="gift2"><td>
Russian Nesting Dolls
</td><td>
Hand-painted by trained monkeys, these exquisite dolls are pr
iceless! And by "priceless," we mean "extremely expensive"! <
span class="excitingNote">8 entire dolls per set! Octuple the
presents!</span>
</td><td>
$10,000.52
</td><td>
<img src="../img/gifts/img2.jpg"/>
</td></tr>


<tr class="gift" id="gift3"><td>
Fish Painting
```

```
</td><td>
If something seems fishy about this painting, it's because i
t's a fish! <span class="excitingNote">Also hand-painted by t
rained monkeys!</span>
</td><td>
$10,005.00
</td><td>
<img src="../img/gifts/img3.jpg"/>
</td></tr>


<tr class="gift" id="gift4"><td>
Dead Parrot
</td><td>
This is an ex-parrot! <span class="excitingNote">Or maybe h
e's only resting?</span>
</td><td>
$0.50
</td><td>
<img src="../img/gifts/img4.jpg"/>
</td></tr>


<tr class="gift" id="gift5"><td>
Mystery Box
</td><td>
If you love suprises, this mystery box is for you! Do not pla
ce on light-colored surfaces. May cause oil staining. <span c
lass="excitingNote">Keep your friends guessing!</span>
</td><td>
$1.50
</td><td>
<img src="../img/gifts/img6.jpg"/>
</td></tr>
```

In [7]:
```
from urllib.request import urlopen
from bs4 import BeautifulSoup

html = urlopen('http://www.pythonscraping.com/pages/page3.htm
l')
bs = BeautifulSoup(html, 'html.parser')
print(bs.find('img',
              {'src':'../img/gifts/img1.jpg'})
      .parent.previous_sibling.get_text())
```

```
$15.00
```

In [8]:
```
from urllib.request import urlopen
from bs4 import BeautifulSoup
import re

html = urlopen('http://www.pythonscraping.com/pages/page3.htm
l')
bs = BeautifulSoup(html, 'html.parser')
images = bs.find_all('img', {'src':re.compile('\.\.\/img\/gif
ts/img.*\.jpg')})
for image in images:
    print(image['src'])
```

```
../img/gifts/img1.jpg
../img/gifts/img2.jpg
../img/gifts/img3.jpg
```

```
../img/gifts/img4.jpg
../img/gifts/img6.jpg
```

In [30]:
```
bs.find_all(lambda tag: len(tag.attrs) == 2)
```

Out[30]:
```
[<img src="../img/gifts/logo.jpg" style="float:left;"/>,
 <tr class="gift" id="gift1"><td>
 Vegetable Basket
 </td><td>
 This vegetable basket is the perfect gift for your health co
nscious (or overweight) friends!
 <span class="excitingNote">Now with super-colorful bell pepp
ers!</span>
 </td><td>
 $15.00
 </td><td>
 <img src="../img/gifts/img1.jpg"/>
 </td></tr>,
 <tr class="gift" id="gift2"><td>
 Russian Nesting Dolls
 </td><td>
 Hand-painted by trained monkeys, these exquisite dolls are p
riceless! And by "priceless," we mean "extremely expensive"!
 <span class="excitingNote">8 entire dolls per set! Octuple t
he presents!</span>
 </td><td>
 $10,000.52
 </td><td>
 <img src="../img/gifts/img2.jpg"/>
 </td></tr>,
 <tr class="gift" id="gift3"><td>
 Fish Painting
 </td><td>
 If something seems fishy about this painting, it's because i
t's a fish! <span class="excitingNote">Also hand-painted by t
rained monkeys!</span>
 </td><td>
 $10,005.00
 </td><td>
 <img src="../img/gifts/img3.jpg"/>
 </td></tr>,
 <tr class="gift" id="gift4"><td>
 Dead Parrot
 </td><td>
 This is an ex-parrot! <span class="excitingNote">Or maybe h
e's only resting?</span>
 </td><td>
 $0.50
 </td><td>
 <img src="../img/gifts/img4.jpg"/>
 </td></tr>,
 <tr class="gift" id="gift5"><td>
 Mystery Box
 </td><td>
 If you love suprises, this mystery box is for you! Do not pl
ace on light-colored surfaces. May cause oil staining. <span
 class="excitingNote">Keep your friends guessing!</span>
 </td><td>
 $1.50
 </td><td>
 <img src="../img/gifts/img6.jpg"/>
 </td></tr>]
```

In [9]:
```
bs.find_all(lambda tag: tag.get_text() == 'Or maybe he\'s onl
```

```
y resting?')
```

Out[9]:
```
[<span class="excitingNote">Or maybe he's only resting?</span
>]
```

In [10]:
```
bs.find_all('', text='Or maybe he\'s only resting?')
```

Out[10]:
```
["Or maybe he's only resting?"]
```

In [ ]:
```

```