Inspection II

Problem: Can you correct this html?

Hint: You will need to write a script

Given: inspection.html

Note:

Steps:
1) Opening the html file in browser and view its source code. We see “flag”. So, maybe, the flag string is wrapped in tag.

```
<html><head>
<meta http-equiv="content-type" content="text/html; charset=UTF-8"></head>
<body>
<h1>Third Subdue Lesser</h1>
<h2>Sixth <e>zero</e> Rule Unto Good</h2>

<p>Make one sixth light fruitful their air light <e>one</e> kind us
<em>flag</em> that multiply his all thing Seas for may <e>one</e> said
creature. Bring fifth form doesn't may, don't fill moved they're be
<e>zero</e> shall was life multiply set meat thing spirit <e>zero</e>
morning. Winged man. Replenish multiply. Can't every fruit <e>one</e> place
```

2) Try cutting out all of the italics word using grep. We do not have the flag, but we are on the right track.

```
[qijun@glap web]$ grep -o '<em>[^/]*</em>' inspection.html
<em>flag</em>
<em>is</em>
<em>not</em>
<em>this</em>
<em>line</em>
<em>but</em>
<em>you</em>
<em>think</em>
<em>right</em>
<em>way</em>
```

3) After further inspection, we see many “<e>one</e>” and “<e>zero</e>” in the source code. It looks like some sort of code. So we use grep to find all such patterns from the source code. Yes, there is a long sequence of zero and one.

```
[qijun@glap web]$ grep -o '<e>[^/]*</e>' inspection.html
<e>zero</e>
<e>one</e>
<e>one</e>
<e>zero</e>
<e>zero</e>
<e>one</e>
<e>one</e>
<e>zero</e>
<e>zero</e>
```

4) Now, let's make a script to convert zeros and ones to 1 and 0. It prints out the string before and after the replacement.

```
#!/usr/bin/python
# -*- coding: utf-8 -*-

import re
f = open('inspection.html')
a = f.read()
b = ''.join(re.findall("<e>([^/]*)</e>", a))
print b
b = b.replace('zero', '0')
b = b.replace('one', '1')
print b
print len(b)
```

```
[qijun@glap web]$ ./inspection2.py
zerooneonezerozerooneonezerozerooneonezerooneonezerozerozerooneonezerozerozeroze
roonezerooneonezerozerooneonezeroonezerooneoneoneoneonezerooneonezeroonezeroz
eroonezerooneoneonezerozerooneonezeroonezerooneoneoneoneonezerooneoneonezeroonez
erozerozerooneonezerooneonezerozerozerozerooneonezerooneonezerozerooneonezerooneoneonezer
ozerooneonezerooneonezerooneoneoneoneonezerooneonezerooneonezerozerooneonezerooneoneoneze
rozerooneonezerooneonezerooneoneoneoneonezerooneonezerozerozerozerooneonezeroonezeroon
eoneoneoneonezerooneoneonezerozerooneonezerooneonezerooneonezerozerooneonezerooneoneze
rooneonezerooneonezerooneoneonezerozerozerozerozerooneonezerooneonezerozerozerooneo
nezerozerooneonezerooneonezerooneonezerooneone
01100110011011000110000101100111010111110110100101110011010111110111010001101000
01101001011100110101111101101001011100110101111101100001010111110111001101101001
0110110101110000011011000110010101011
197
```

(5) We get a binary string. We find that every 8 bits has a staring 0 bit. It indicates every 8 bits may be an printable ASCII character. So, we then need to convert the binary string to a character string...