

1. Goal: I want to get some links from a web page, like this:

加强心理疏导，携手抗击疫情		
必修	【专家引领】五分钟和孩子说清疫情这件事	✓
必修	【专家引领】面对疫情你恐慌了吗？	!
必修	【专家引领】个体在应激中的生理反应原理	!
必修	【专家引领】群体性危机事件中个体的行为表现	!
必修	【专家引领】面对疫情，大众群体会产生怎样的应激状态？	!

2. Then I find the source code:

```
<tr class="normal act_for_line" id="ce04f5177f000001d6adb42b12d38885_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5187f000001d6adb42b90a5d50a_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5157f000001d6adb42b07db6ba4_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f51d7f000001d6adb42be9641456_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5167f000001d6adb42b50b5c62c_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5107f000001d6adb42b74fd77d1_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f51a7f000001d6adb42bdb3460b5_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5187f000001d6adb42bb6993e31_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5167f000001d6adb42be5bd795a_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f51a7f000001d6adb42bfeb3a160_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5157f000001d6adb42bd45ba7b3_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f51d7f000001d6adb42b81528930_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5147f000001d6adb42b64a5c3fc_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5197f000001d6adb42b94130614_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f51c7f000001d6adb42b188026ce_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f5197f000001d6adb42bd7613f68_BACT004">...</tr>
<tr class="normal act_for_line" id="ce04f51d7f000001d6adb42bda629db6_BACT004">...</tr>
```

3. I want to use webdriver, so I just want the code after id. I use an online software:

<https://ocr.wdku.net/>.

Then I copy the text I need and paste it in an Excel file:

id="ce04f5177f000001d6adb42b12d38885_BACT004">^</tr>
id="ce04f5187f000001d6adb42b90a5d50a_BACT004">„</tr>
id="ce04f5157f000001d6adb42b07db6ba4_BACT004">„</tr>
id="ce04f51d7f000001d6adb42be9641456_BACT004">„</tr>
id="ce04f5167f000001d6adb42b50b5c62c_BACT004">„</tr>
id="ce04f5107f000001d6adb42b74fd77d1_BACT004">„</tr>
id="ce04f51a7f000001d6adb42bdb3460b5_BACT004">„</tr>
id="ce04f5187f000001d6adb42bb6993e31_BACT004">„</tr>
id="ce04f5167f000001d6adb42be5bd795a_BACT004">„</tr>
id="ce04f51a7f000001d6adb42bfeb3a160_BACT004">„</tr>
id="ce04f5157f000001d6adb42bd45ba7b3_BACT004">„</tr>
id="ce04f51d7f000001d6adb42b81528930_BACT004">„</tr>
id="ce04f5147f000001d6adb42b64a5c3fc_BACT004">„</tr>
id="ce04f5197f000001d6adb42b94130614_BACT004">„</tr>
id="ce04f51c7f000001d6adb42b188026ce_BACT004">„</tr>
id="ce04f5197f000001d6adb42bd7613f68_BACT004">„</tr>
id="ce04f51d7f000001d6adb42bda629db6_BACT004">^</tr>

4. Then I open it in python. By the way, I also learn to read a csv but not add a header, and output a csv but not add an index. Such two problem is really easy, but I always try to

avoid these problem, and avoid to learn something new which can help me solve my problem quickly. *Haste makes waste*. I learn it again.

5. The results

```
[61] > MI
link = pd.read_csv('id.csv',header=None,names=['0'])
link1 = link['0'].apply(lambda x: x[4:44])

[62] > MI
link.head()

      0
0 id="ce04f5177f000001d6adb42bl2d38885_BACT004,,...
1 id="ce04f5187f000001d6adb42b90a5d50a_BACT004">...
2 id="ce04f5157f000001d6adb42b07db6ba4_BACT004">...
3 id="ce04f51d7f000001d6adb42be9641456_BACT004">...
4 id="ce04f5167f000001d6adb42b50b5c62c_BACT004">...

[63] > MI
link1.head()

0    ce04f5177f000001d6adb42bl2d38885_BACT004
1    ce04f5187f000001d6adb42b90a5d50a_BACT004
2    ce04f5157f000001d6adb42b07db6ba4_BACT004
3    ce04f51d7f000001d6adb42be9641456_BACT004
4    ce04f5167f000001d6adb42b50b5c62c_BACT004
Name: 0, dtype: object

[64] > MI
link1.to_csv('link1.csv',index=False)
```

6. However, the webpage can't be operated by webdriver, maybe it is encrypted by someone hhhh.