

# four\_webCrawler

```
import requests
import lxml
from bs4 import BeautifulSoup
from xlwt import *
workbook = Workbook(encoding = 'utf-8')
table = workbook.add_sheet('data')
table.write(0, 0, 'Number')
table.write(0, 1, 'movie_url')
table.write(0, 2, 'movie_name')
table.write(0, 3, 'movie_introduction')
line = 1
url = "https://www.rottentomatoes.com/top/bestofrt/"
headers = {
'User-Agent': 'Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like
Gecko) Chrome/63.0.3239.132 Safari/537.36 QIHU 360SE'
}
f = requests.get(url, headers = headers)
movies_lst = []
soup = BeautifulSoup(f.content, 'lxml')
movies = soup.find_all('a', {
'class': 'js-tile-link'
})
num = 0
for anchor in movies:
try:
urls = 'https://www.rottentomatoes.com' + anchor.get('href')
movies_lst.append(urls)
num += 1
movie_url = urls
movie_f = requests.get(movie_url, headers = headers)
movie_soup = BeautifulSoup(movie_f.content, 'lxml')
movie_content = movie_soup.find('div', {
'class': 'movie_synopsis clamp clamp-6 js-clamp'
})
print(num, urls, '\n', 'Movie:' + urls.split('/')[-1])
print('Movie info:' + movie_content.string.strip())
table.write(line, 0, num)
table.write(line, 1, urls)
```

```
table.write(line, 2, urls.split('/')[1])
table.write(line, 3, movie_content.string.strip())
line += 1
workbook.save('movies_top100.xls')
except TypeError:
pass
```

## #SAMPLE OUTPUT

```
# 1 https://www.rottentomatoes.com/m/hold_me_tight_2021
# Movie:hold_me_tight_2021
# Movie info:In Hold Me Tight, Vicky Krieps (Phantom Thread, Bergman Island) gives another
riveting
# performance as Clarisse, a woman on the run from her family for reasons that aren't
immediately clear.
# Widely renowned as one of France's great contemporary actors but less well-known in
North America for his
# equally impressive work behind the camera, Mathieu Amalric's sixth feature as director is
his most
# ambitious to date. This virtuosic, daringly fluid portrait of a woman in crisis alternates
between
# Clarisse's adventures on the road and scenes of her abandoned husband Marc (Arieh
Worthalter) as he
# struggles to take care of their children at home. Amalric's film keeps viewers uncertain as to
the
# reality of what they're seeing until the final moments of this moving, unpredictable, and
richly rewarding family drama.
# 2 https://www.rottentomatoes.com/m/my_fathers_dragon_2022
# Movie:my_fathers_dragon_2022
# Movie info:Struggling to cope after a move to the city with his mother, Elmer runs away in
search
# of Wild Island and a young dragon who waits to be rescued. Elmer's adventures introduce
him to
# ferocious beasts, a mysterious island and the friendship of a lifetime.
# 3 https://www.rottentomatoes.com/m/fire_of_love
# Movie:fire_of_love
# Movie info:Fire of Love tells the story of two French lovers, Katia and Maurice Krafft, who
died
```

# in a volcanic explosion doing the very thing that brought them together: unraveling the mysteries

# of our planet, while simultaneously capturing the most explosive volcano imagery ever recorded.

# Along the way, they changed our understanding of the natural world, and saved tens of thousands of lives.

# Previously unseen hours of pristine 16-millimeter film and thousands of photographs reveal the birth of

# modern volcanology through an unlikely lens -- the love of its two pioneers.