

AIST 2120
Lab 8 – Chapter 12

Find the Funnies, A.K.A., Web Scraping

Introduction

Chapter 12 introduces us to the wonderful world of scraping. It's a useful capability and can be quite fun.

Assignment & Discussion

Your mission is to practice using the web scraping tools encountered in Chapter 12.

Use the lab template, `lastN_lab8_Ch12.py`; it works in concert with the directions for **Project: Downloading All XKCD Comics** on page 286 of the text (2nd edition).

Tasks

Your program should accomplish the following:

1. Ensure current working directory is the **lab8** directory on desktop.
2. Load the **XKCD** home page. See "Project: Downloading All XKCD Comics" in your book or at <https://automatetheboringstuff.com/2e/chapter12/>.
3. Save/copy the comic image on that page into a directory on the **lab8** folder. Name the directory `lastN_xkcd`. Don't hardcode the name in each statement; instead set a variable name early in the program identifying the name of this directory and use it thereafter.
4. Follow the Previous Comic link.
5. Repeat until it reaches the first comic number ending with **00** (which will be 2400).

Extra Challenge (+5 points): After the program runs with the other requirements, in Step 4, can you modify the script with a try-except-else block to continue program execution if an image download fails?

Notes

- Study the Project code. Seek to understand the purpose and behavior of each block of code. Include comments demonstrating that understanding of the code.
- Pay attention to the 'divide-and-conquer' approach to this project.
- Work to understand, not just copy.

Requirements

☐ Application

- Write your program in Implement your Python 3 program using the development environment of your choice.
- Download the provided template `lastN_lab8_Ch12.py` and save your file with your name.
- Pay close attention to comments and white space in the template code. Your goal is to write, or rearrange, source code that is easy to read and understand.
- Remember to make the output pleasing to the program user.

☐ Submission

AIST 2120
Lab 8 – Chapter 12

- Lab exercises are evaluated through demonstration. Let me know when you are ready to demo your code. D2L submission is required. No email submissions will be accepted.
- Retain a complete copy of your program source code file. It may come in handy later.

☐ **Due Date:** Per D2L instructions.

☐ **Late Penalty:** Per D2L instructions.

Style Guide

☐ **File headers.** Include a header in the below format at the top of all .py source code files.

```
# Your Name
# AIST 2120, Lab Exercise 8 ( Chap. 12)
# xx/xx/202x
```

☐ **Sample Screen Output**

```
-----
---      AIST 2120      ---
---    Lab Exercise 8, Ch. 12    ---
---      Find the Funnies      ---
---    A.k.a., Web Scraping    ---
-----

Current working directory is /Users/Mistony/Desktop .
FILE(S) DOWNLOAD TO COMMENCE ...

Downloading page http://xkcd.com...
  Downloading image http://imgs.xkcd.com/comics/alien_visitors_2.png...
Downloading page http://xkcd.com/2477/...
  Downloading image http://imgs.xkcd.com/comics/alien_visitors.png...
Downloading page http://xkcd.com/2476/...
  Downloading image http://imgs.xkcd.com/comics/base_rate.png...
Downloading page http://xkcd.com/2475/...
  Downloading image http://imgs.xkcd.com/comics/health_drink.png...
Downloading page http://xkcd.com/2474/...
  Downloading image http://imgs.xkcd.com/comics/first_time_since_early_2020.png...
Downloading page http://xkcd.com/2473/...
  Downloading image http://imgs.xkcd.com/comics/product_launch.png...
Downloading page http://xkcd.com/2472/...
  Downloading image http://imgs.xkcd.com/comics/fuzzy_blob.png...
Downloading page http://xkcd.com/2471/...
  Downloading image http://imgs.xkcd.com/comics/hippo_attacks.png...
Downloading page http://xkcd.com/2470/...
  Downloading image http://imgs.xkcd.com/comics/next_slide_please.png...
Downloading page http://xkcd.com/2469/...
  Downloading image http://imgs.xkcd.com/comics/astronomy_status_board.png...
Downloading page http://xkcd.com/2468/...
  Downloading image http://imgs.xkcd.com/comics/inheritance.png...
Downloading page http://xkcd.com/2467/...
  Downloading image http://imgs.xkcd.com/comics/wikipedia_caltrops.png...
```

• • •

AIST 2120

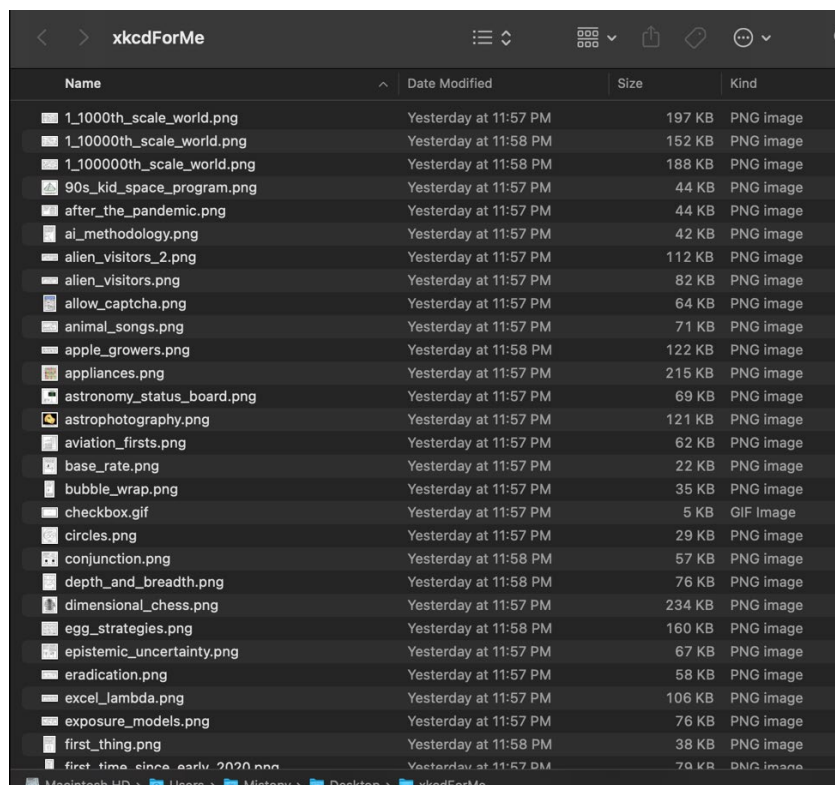
Lab 8 – Chapter 12

```
...
Downloading page http://xkcd.com/2413/...
  Downloading image http://imgs.xkcd.com/comics/pulsar_analogy.png...
Downloading page http://xkcd.com/2412/...
  Downloading image http://imgs.xkcd.com/comics/1_100000th_scale_world.png...
Downloading page http://xkcd.com/2411/...
  Downloading image http://imgs.xkcd.com/comics/1_100000th_scale_world.png...
Downloading page http://xkcd.com/2410/...
  Downloading image http://imgs.xkcd.com/comics/apple_growers.png...
Downloading page http://xkcd.com/2409/...
  Downloading image http://imgs.xkcd.com/comics/steepen_the_curve.png...
Downloading page http://xkcd.com/2408/...
  Downloading image http://imgs.xkcd.com/comics/egg_strategies.png...
Downloading page http://xkcd.com/2407/...
  Downloading image http://imgs.xkcd.com/comics/depth_and_breadth.png...
Downloading page http://xkcd.com/2406/...
  Downloading image http://imgs.xkcd.com/comics/viral_vector_immunity.png...
Downloading page http://xkcd.com/2405/...
  Downloading image http://imgs.xkcd.com/comics/flash_gatsby.png...
Downloading page http://xkcd.com/2404/...
  Downloading image http://imgs.xkcd.com/comics/first_thing.png...
Downloading page http://xkcd.com/2403/...
  Downloading image http://imgs.xkcd.com/comics/wrapping_paper.png...
Downloading page http://xkcd.com/2402/...
  Downloading image http://imgs.xkcd.com/comics/into_my_veins.png...
Downloading page http://xkcd.com/2401/...
  Downloading image http://imgs.xkcd.com/comics/conjunction.png...

-----
---      Lab Exercise 8      ---
---      Complete          ---
---  That, that's all folks!  ---
-----

Press the ENTER key to exit
>>>
```

☐ Sample Abbreviated Folder Output



Name	Date Modified	Size	Kind
1_1000th_scale_world.png	Yesterday at 11:57 PM	197 KB	PNG image
1_10000th_scale_world.png	Yesterday at 11:58 PM	152 KB	PNG image
1_100000th_scale_world.png	Yesterday at 11:58 PM	188 KB	PNG image
90s_kid_space_program.png	Yesterday at 11:57 PM	44 KB	PNG image
after_the_pandemic.png	Yesterday at 11:57 PM	44 KB	PNG image
ai_methodology.png	Yesterday at 11:57 PM	42 KB	PNG image
alien_visitors_2.png	Yesterday at 11:57 PM	112 KB	PNG image
alien_visitors.png	Yesterday at 11:57 PM	82 KB	PNG image
allow_captcha.png	Yesterday at 11:57 PM	64 KB	PNG image
animal_songs.png	Yesterday at 11:57 PM	71 KB	PNG image
apple_growers.png	Yesterday at 11:58 PM	122 KB	PNG image
appliances.png	Yesterday at 11:57 PM	215 KB	PNG image
astronomy_status_board.png	Yesterday at 11:57 PM	69 KB	PNG image
astrophotography.png	Yesterday at 11:57 PM	121 KB	PNG image
aviation_firsts.png	Yesterday at 11:57 PM	62 KB	PNG image
base_rate.png	Yesterday at 11:57 PM	22 KB	PNG image
bubble_wrap.png	Yesterday at 11:57 PM	35 KB	PNG image
checkbox.gif	Yesterday at 11:57 PM	5 KB	GIF image
circles.png	Yesterday at 11:57 PM	29 KB	PNG image
conjunction.png	Yesterday at 11:58 PM	57 KB	PNG image
depth_and_breadth.png	Yesterday at 11:58 PM	76 KB	PNG image
dimensional_chess.png	Yesterday at 11:57 PM	234 KB	PNG image
egg_strategies.png	Yesterday at 11:58 PM	160 KB	PNG image
epistemic_uncertainty.png	Yesterday at 11:57 PM	67 KB	PNG image
eradication.png	Yesterday at 11:57 PM	58 KB	PNG image
excel_lambda.png	Yesterday at 11:57 PM	106 KB	PNG image
exposure_models.png	Yesterday at 11:57 PM	76 KB	PNG image
first_thing.png	Yesterday at 11:58 PM	38 KB	PNG image
first_time_since_early_2020.png	Yesterday at 11:57 PM	79 KB	PNG image

AIST 2120
Lab 8 – Chapter 12

□ Example XKCD Comic.

