

Webscraping Geospatial Data with Python/Beautiful Soup

Hal Mueller, MaptimeSEA, May 7, 2024

Downloads

- <https://www.anaconda.com/download>
- <https://www.google.com/chrome/>
- <https://github.com/halmueller/SoupTalk>

Beautiful Soup

‘Beautiful Soup, so rich and green,
Waiting in a hot tureen!
Who for such dainties would not
stoop?
Soup of the evening, beautiful Soup!
Soup of the evening, beautiful Soup!
Beau--ootiful Soo--oop!
Beau--ootiful Soo--oop!
Soo--oop of the e--e--evening,
Beautiful, beautiful Soup!’

– Lewis Carroll

Beautiful Soup

- “Tag soup”: https://en.wikipedia.org/wiki/Tag_soup
- Python Package: “Beautiful Soup”
- [https://en.wikipedia.org/wiki/Beautiful_Soup_\(HTML_parser\)](https://en.wikipedia.org/wiki/Beautiful_Soup_(HTML_parser))
- Leonard Richardson + Tidelift: <https://www.crummy.com/software/BeautifulSoup/>
- Python 3, BeautifulSoup4 4.12.3: <https://pypi.org/project/beautifulsoup4/>
- Alternative: <https://scrapy.org/>

“Don’t try this at home!”

- Web API if you can: Wikipedia, Seattle City Data
- Be kind to servers: render cost
- Test often

Activity

- <https://sfdlive.com/>
- <https://web.seattle.gov/sfd/realtime911/>
- <https://data.seattle.gov/Public-Safety/Seattle-Real-Time-Fire-911-Calls/kzjm-xkqj/data>

Python with Anaconda

- Environment manager with Python3 installation
- Package manager

Package and environment setup

- Launch Anaconda
- Add “conda-forge” to Channels
- Create environment for soup course
- Import GeoPy, BS4, requests to the environment
- Launch command line from environment “play” button
- `python3 scripts/check_install.py`
- Windows: `python scripts/check_install.py`

Geocoding

- GeoPy: <https://geopy.readthedocs.io/en/stable/#module-geopy.geocoders>, wraps 32 different services
- Terms of service
- Licensing of data
- Alternative: King County government has local geocoder, with web API

python3 scripts/nominatim_demo.py

```
from geopy.geocoders import Nominatim

geocoder = Nominatim(user_agent="SoupTalk_add_your_name")

# TPC Ballard
# 5101 14th Ave NW
# Suite 201
# Seattle, WA 98103

location1 = geocoder.geocode("5101 14th Avenue NW")

print(location1.address)

print((location1.latitude, location1.longitude))

print(location1.raw)
```