

Project description for Web Scraping and Social Media
Scraping classes.

Spring 2020/2021

Rafał Łobacz

384892

Paulina Żabska

385257

Description of the topic and the web page.

Main idea of the topic was to scrape data about cat species. Every cat is different and require special treatment/approach. Future owner of the cat should think before obtaining one, selecting the one which would fit to him/her. Our scrapers gather data about 42 cat species. Obtained species database is quite detailed and can be used in many ways. Our main idea of data usage is to create recommendation system based on specific needs of the clients and their personal characteristic. Similar task could be done for dogs or other animal. For the sake of purpose we only scraped cats.

Initial web page that we scraped, in the main body contains links to all cat species. Although site shows only few species based on starting letter. It is not a problem because all links are available in the source code. Web page hides other cats only for visual purpose. Then based on gathered links, we move to each link and scrape each cat description.

Description of scraper mechanics.

All of three scrapers have following mechanism. First they scrape starting web page with links to each cat species. Then they crawl to each link and extract all information about cats. First header elements are scraped – name, age, weight and then body elements: traits in numeric format and text informations like history, fun facts or grooming. BeautifulSoup use html elements to retrieve those part which we want, scrapy and selenium use xpath. Each scraper scrapes only 43 web pages, so we did not put limit of 100 sites at the beginning.

BeautifulSoup took 10 seconds to run, scrapy 2 seconds and selenium 100 seconds.

Share of work.

<u>Rafał</u>	<u>Paulina</u>
Writing BS scraper	Testing BS scraper
Writing Scrapy scraper	Testing Scrapy scraper
Testing Selenium scraper	Writing Selenium scraper
Creating descriptions	Creating descriptions
Data analysis	Data analysis

Technical description of the output.

Each scraper produce csv file containing scraped data. Database description is presented in the table below.

<u>Output Dataset Description</u>		
Name	Definition	Type
name	species name	character
lifespan	length of cat life	character(can be easily transformed to numeric)
weight	weight of cat	character(can be easily transformed to numeric)
lap_cat	eagerness to sit on laps (0-5 scale, 0-little, 5-very much)	numeric
intelligence	species smartness/intelligence (0-5 scale, 0-bad, 5-very good)	numeric
ease_of_training	easiness to train (0-5 scale, 0-easy, 5-very hard)	numeric
grooming_requirements	needed amount of cleaning (0-5 scale, 0-little, 5-very much)	numeric
shedding	tendency to lose hair (0-5 scale, 0-little, 5-very much)	numeric
good_with_children	how good cat is with children (0-5 scale, 0-bad, 5-very good)	numeric
good_with_dogs	how cat handle dogs (0-5 scale, 0-bad, 5-very good)	numeric
chattiness	how much cat like to chat/talk (0-5 scale, 0-little 5-very much)	numeric
history	history of species	character
behavior	behavior concerns	character
look	physical description	character
grooming	groomng requirements	character
nutritrional_needs	suggested nutritional needs	character
fun_facts	fun facts about species	character

Elementary data analysis.

As we said previously, one can build recommendation system based on client needs. Let's assume that someone wants cat that is dog friendly. For such person we should propose cat that has the highest score in good_with_dogs column. If there are more than one cat with the same value we propose the one with highest average score out of them. Such analysis could be seen in the table below.

name	average	lap_cat	intelligence	ease_of_training	grooming_requirements	shedding	good_with_children	good_with_dogs	chattiness
Maine Coon	4.000	5	4	4	3	4	5	5	2
American Bobtail	3.875	5	4	4	3	3	5	5	2
Devon Rex	3.625	3	5	5	1	1	4	5	5
Tonkinese	3.500	1	5	4	1	2	5	5	5
Birman	3.000	5	3	3	2	2	3	5	1


As we can see in the table above. Maine Coon is on average the best cat within the cats that deal with dogs the best. In the next step we can build dashboard that contains data about the best cat with picture and link to shop/shelter. Such summary of the cat could be seen below.

Maine Coon

Lifespan: 12 - 15 years | | Height: | Weight: 8 - 20 pounds

Maine Coon Traits

Lap Cat	● ● ● ● ●
Intelligence	● ● ● ● ●
Ease of Training	● ● ● ● ●
Grooming Requirements	● ● ● ● ●
Shedding	● ● ● ● ●
Good with Children	● ● ● ● ●
Good with Dogs	● ● ● ● ●
Chattiness	● ● ● ● ●



Fun Facts of Maine Coons

- Enjoys many nicknames, including "Gentle Giant," "Shags," and "Feline Greeters of the World."

Look of Maine Coons

- Resembles a small bobtail with large frame, barrel chest, its tuft ears, full mane, shaggy weatherproof coat and bushy tail.
- The head is large and square with big, expressive eyes that are wide set.

There are more things that we can include but one page requirement limit us to this brief conspect.