Creating URLs path for Displaying in the Webapp

Initially, I saved 10% of the training dataset (Stanford Dog Dataset) into our project image folder for displaying matched dog images in our web app. However, the folder is still significantly large, so I decided to upload the dataset into the Github repository then get all image links. So I can use URLs to get images in the web app without uploading large images.

In [1]:
```
import numpy as np
import pandas as pd
import os
```

I uploaded the training dataset into my Google drive first.

In [2]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

Now, I will get the path where I saved my dataset.

In [3]:
```
sample_image_folder_path = '/content/drive/MyDrive/dogdata/Images'
```

Now, let's loop over each images.

In [4]:
```
path_list = []
dog_name = []

# loop over all 120 dog folders
for root, directories, files in os.walk(sample_image_folder_path):
  # inside each dog folder
  for link in files:
    get_path = os.path.join(root, link)
    path_list.append(os.path.join(root, link))
    dog_name.append(root.split('-')[-1])
```

Let's check the first path.

In [8]:
```
path_list[0]
```

Out[8]:
```
'/content/drive/MyDrive/dogdata/Images/n02085620-Chihuahua/n02085620_1205.jpg'
```

We do not need `/content/drive/MyDrive/SampleData/val/` at the front, so we should remove it.

In [10]:
```
dog_path=[]
for x in path_list:
  dog_path.append(x[38:])
```

Now, let's check the first path to make sure.

In [11]:
```
dog_path[0]
```

Out[11]:
```
'n02085620-Chihuahua/n02085620_1205.jpg'
```

Next, I will change all the dog names to lowercase.

In [12]:
```
dog_name = np.asarray(dog_name)
# change to lowercase for names
dog_name = np.array([x.lower() if isinstance(x, str) else x for x in dog_name])
```

Let's check the first name.

In [13]:
```
dog_name[0]
```

Out[13]:
```
'chihuahua'
```

Then, I will change all the paths to lowercase.

In [14]:
```
# change to lowercase for paths
dog_path = np.array([x.lower() if isinstance(x, str) else x for x in dog_path])
```

Let's check it again.

In [15]:
```
dog_path[0]
```

Out[15]:
```
'n02085620-chihuahua/n02085620_1205.jpg'
```

I noticed at my GitHub repository [pic16b-stanford-dog-dataset (https://github.com/PengWu2626/pic16b-stanford-dog-dataset/tree/main/images)](https://github.com/PengWu2626/pic16b-stanford-dog-dataset/tree/main/images) where I uploaded the dataset that each image link begins with `repo_path` below.

In [16]:
```
repo_path = str("https://raw.githubusercontent.com/PengWu2626/pic16b-stanford-dog-dataset/main/image
```

Let's add the `repo_path` to all paths we got from early.

In [17]:
```
repo_path = np.asarray(repo_path)
dog_image_path = np.char.add(repo_path, dog_path)
```

Let's double check it.

In [ ]:
```
dog_image_path[0]
```

Out[21]:
```
'https://raw.githubusercontent.com/PengWu2626/pic16b-stanford-dog-dataset/main/images/n02085620-chihuah
ua/n02085620_1205.jpg'
```

Excellent, now we can click that link to get the image.

Let's create a dataframe for each image link with the name of the associated dog breed.

In [18]:
```
df = pd.DataFrame({'name':dog_name, 'path':dog_image_path})
```

Let's check the dataframe.

In [19]:
```
df.head()
```

Out[19]:

	name	path
0	chihuahua	https://raw.githubusercontent.com/PengWu2626/p...
1	chihuahua	https://raw.githubusercontent.com/PengWu2626/p...
2	chihuahua	https://raw.githubusercontent.com/PengWu2626/p...
3	chihuahua	https://raw.githubusercontent.com/PengWu2626/p...
4	chihuahua	https://raw.githubusercontent.com/PengWu2626/p...

Finally, let's saved the dataframe called `dog_sample_images_path.csv` so I can use it in our web app.

In [20]:
```
df.to_csv('dog_sample_images_path.csv', index=False)
```