In [ ]:
```
# %%bash
# for dir in /home/amazigh/MLFinal/OpenImages/*;
#     do
#     cd $dir
#     for dir2 in */; do
#         rsync -avz pascal/ images/ "$dir"/ --delete-after;
#     done
#     cd ..
# done
```

In [ ]:
```
# %%bash

# # if globstar is not enabled, you'll need it.
# shopt -s globstar
# for file in */**; do [ -f "$file" ] && mv -i "$file" "${file//\//-}"; done
# # get rid of the now-empty subdirectories.
# find . -type d -empty -delete
```

In [ ]:
```
import os
from random import choice
import shutil

#arrays to store file namesMyImages
imgs =[]
xmls =[]

#setup dir namesMyImages
trainPath = 'train'
valPath = 'val'
testPath = 'test'
crsPath = 'OpenImgs' #dir where images and annotations stored

#setup ratio (val ratio = rest of the files in origin dir after splitting into train and test)
train_ratio = 0.7
test_ratio = 0.2


#total count of imgs
totalImgCount = len(os.listdir(crsPath))/2
```

In [ ]:
```
#soring files to corresponding arrays
for (dirname, dirs, files) in os.walk(crsPath):
    for filename in files:
        if filename.endswith('.xml'):
            xmls.append(filename)
        else:
            imgs.append(filename)
```

In [ ]:
```
#counting range for cycles
countForTrain = int(len(imgs)*train_ratio)
countForTest = int(len(imgs)*test_ratio)

print(countForTest)
print(countForTrain)
```

In [ ]:
```
#cycle for train dir
for x in range(countForTrain):

    fileJpg = choice(imgs) # get name of random image from origin dir
    fileXml = fileJpg[:-4] +'.xml' # get name of corresponding annotation file

    #move both files into train dir
    shutil.move(os.path.join(crsPath, fileJpg), os.path.join(trainPath, fileJpg))
    shutil.move(os.path.join(crsPath, fileXml), os.path.join(trainPath, fileXml))

    #remove files from arrays
    imgs.remove(fileJpg)
    xmls.remove(fileXml)
```

In [ ]:
```
#rest of files will be validation files, so rename origin dir to val dir
os.rename(crsPath, valPath)

#summary information after splitting
print('Total images: ', totalImgCount)
print('Images in train dir:', len(os.listdir(trainPath))/2)
print('Images in test dir:', len(os.listdir(testPath))/2)
print('Images in validation dir:', len(os.listdir(valPath))/2)
```