

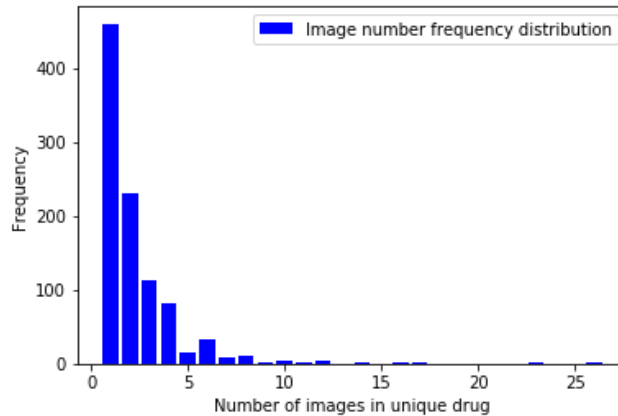
```
In [50]: import time
import os
import pro
import pandas as pd
import numpy as np
from IPython.display import HTML
from IPython.display import display
from IPython.display import SVG
from IPython.display import display_pretty, display_html, display_jpeg, display_png, display_json, display_latex, display_svg
from IPython.display import Image
import PIL
from PIL import Image
import matplotlib
%matplotlib inline
import matplotlib.pyplot as plt
```

```
In [51]: class Drug(object):
        """__init__() functions as the class constructor"""
        def __init__(self, name=None, text=None, label=None):
            self.name = name
            self.text = text
            self.label = label
n, t, n_r, t_r = [],[],[],[]
```

```
In [52]: drug_list=[]#list of drugs to store fileName,Text and label information
with open('data/ocr.txt', 'r') as rf:
    for line in rf:
        split_r = pro.split_fname_texts(line)
        drug_list.append(Drug(split_r['name'], split_r['text'], split_r['label']))
print ('Totally: ' + str(len(drug_list)) + ' images loaded')
```

Totally: 2198 images loaded

```
In [53]: df = pd.DataFrame([vars(f) for f in drug_list])
df.columns = ["label", "name", "text"]
count_fre = pro.frequency_distribution(drug_list)
plt.bar(count_fre.keys(),count_fre.values(), label="Image number frequency distribution",
        color='b')
plt.legend()
plt.xlabel('Number of images in unique drug')
plt.ylabel('Frequency')
plt.show()
```



```
In [54]: final_list = pro.frequency(drug_list, 4)#get the records with count more than n
print ('Unique drug number: ' + str((len(final_list))))
df = df[df['label'].isin(final_list)]
print ('Total images number: ' + str(len(df)))
```

Drugs containing more than 4 images are selected
Unique drug number: 81
Total images number: 618

```
In [55]: print ('Computing similarity score with each records in reference dataset.....
\n')
print ('..... \n')
print ('..... \n')
r1, r2, n, t, n_r, t_r= pro.cal_top(df, 15, 1)
print ('Computing finish!')
```

Computing similarity score with each records in reference dataset.....

.....

.....

Computing finish!

```
In [56]: print ('Following are top score images found:')
r1[['result_label','result_score']]
```

Following are top score images found:

Out[56]:

	result_label	result_score
29	13bb8267-1cab-43e5-acae-55a4d957630a	200
27	13bb8267-1cab-43e5-acae-55a4d957630a	110
28	13bb8267-1cab-43e5-acae-55a4d957630a	110
26	13bb8267-1cab-43e5-acae-55a4d957630a	110
80	2bcd627-51e3-4e5c-a23c-30deb86ff37b	85
159	47924917-bbd5-45de-96cb-c7ff9754da97	84
23	10db92f9-2300-4a80-836b-673e1ae91610	83
419	a0949773-c75f-4ef0-a69c-1abd978340d0	83
82	2bcd627-51e3-4e5c-a23c-30deb86ff37b	83
260	60fcfb61-f594-4e7e-89a6-155965a143a9	82
46	205be197-5ad5-4849-bc78-5fda9927b531	82
263	60fcfb61-f594-4e7e-89a6-155965a143a9	82
264	60fcfb61-f594-4e7e-89a6-155965a143a9	82
538	e45c75dc-d381-475b-b649-a871c8a36e60	82
261	60fcfb61-f594-4e7e-89a6-155965a143a9	82

```
In [57]: print('Following are top candidates found: ')
r2
```

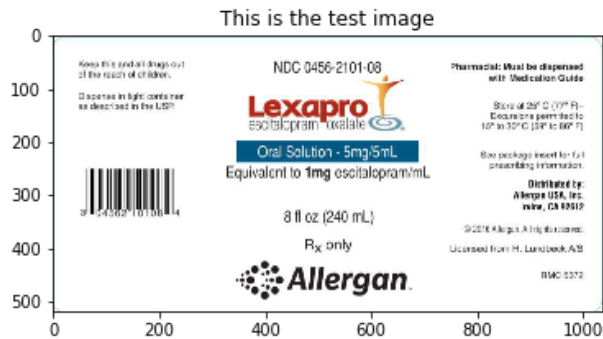
Following are top candidates found:

Out[57]:

	result_label	total_score	probability
1	13bb8267-1cab-43e5-acae-55a4d957630a	530	0.37
5	60fcfb61-f594-4e7e-89a6-155965a143a9	328	0.23
3	2bcd627-51e3-4e5c-a23c-30deb86ff37b	168	0.12
4	47924917-bbd5-45de-96cb-c7ff9754da97	84	0.06
0	10db92f9-2300-4a80-836b-673e1ae91610	83	0.06
6	a0949773-c75f-4ef0-a69c-1abd978340d0	83	0.06
2	205be197-5ad5-4849-bc78-5fda9927b531	82	0.06
7	e45c75dc-d381-475b-b649-a871c8a36e60	82	0.06

```
In [58]: img = Image.open(os.path.join('data/images/', n[0]))
plt.figure()
plt.title('This is the test image')
plt.imshow(img)
print('Extracted Text: ' + t[0])
```

Extracted Text: Keep this and all drugs out of the reach of children. Dispense in tight container as described in the USP. T ©! SC NDC 0456-2101-08. Lgxapro a lopram "oxalate E Oral Solution - 5mg/5mL Equivalent to 1mg escitalopram/mL 8 f l 0z (240 mL) Ry only Allergan. Pharmacist: Must be dispensed with Medication Guide Store at 25° C (77° F)- Excursions permitted to 15°10 30° C (59° to 86° F) See package insert for full prescribing information. Distributed by: Allergan U SA, Inc. Irvine, CA 92612 © 2016 Allergan. Al rights reserved. Licensed from H. Lundbeck A/S RMC 5372



```
In [59]: print ('Following are the top images found')
s = ''
for i,val in enumerate(n_r):
    path = os.path.join('data/images/', val)
    s += ''
    s += '<p>' + 'Extracted Text: ' + t_r[i] + '</p>' + '<hr style="height:1px"
/>'
h = HTML(s); h
```

display

<http://localhost:8892/nbconvert/html/display.ipyn...>

Following are the top images found

Out[59]:



Extracted Text: Keep this and all drugs out of the reach of children. Dispense in tight container as described in the USP. T ©! SC NDC 0456-2101-08. Lgxapro alopram "oxalate E Oral Solution - 5mg/5mL Equivalent to 1mg escitalopram/mL 8 fl Oz (240 mL) Ry only Allergan. Pharmacist: Must be dispensed with Medication Guide Store at 25° C (77° F)- Excursions permitted to 15° 10 30° C (59° to 86° F) See package insert for full prescribing information. Distributed by: Allergan USA, Inc. Irvine, CA 92612 © 2016 Allergan. Al rights reserved. Licensed from H. Lundbeck A/S RMC 5372



Extracted Text: NDC 0456-2010-01 o esci(a\op(amlo\alate g. Tablets—10 mg Equivalent to 10 mg escitalopram 100 Tablets Rx only Allergan. Pharmacist: Must be dispensed with Medication Guide Usual dosage: Seg package insert for full prescribing information. Keep this and all drugs out of the reach of children. Each tablet contains 10mg escitalopram Dispense in tight container as described in the USP. Store at 25° C (77° F)— Excursions permitted to 15° 0 30° C (59° to 86° F) Licensed from H. Lundbeck A/S MPO1



Extracted Text: 40841 wz (I NDC 0456-2020-01 escitalopram loxa\ate g. Tablets—20 mg Equivalent to 20 mg escitalopram 100 Tablets Rx only Allergan. Pharmacist: Must be dispensed with Medication Guide Usual dosage: Sea package insert for full prescribing information. Keep this and all drugs out of the reach of children Each tablet contains 20mg escitalopram Dispense in tight container as described in the USP. Store at 25° C (77° F)— Excursions permitted to 15° 10 30° C (59° to 86° F) Licensed from H. Lundbeck A/S MPO1



In []: