```
# coding: utf-8

# # Configuration + Imports

# In[7]:

from Bio import Entrez, SeqIO
import time
from lxml import html
import requests
import urllib
import sys

Entrez.email = "vincent.deruaz@master.hes-so.ch"


# # Populate db
# ## Class

# In[8]:

class db_population():
    def get_informations_from_phage_db(self, phage_name):
        page = requests.get('http://phagesdb.org/phages/%s' % (phage_name))
        tree = html.fromstring(page.content)

        host = tree.xpath('//*[@id="phageDetails"]/tbody/tr/td[2]/a/em/text()')[0]

        return host

    def get_list_Phage_from_ebi(self, ):
        pass_first = 0
        access_ids = []

        for line in open('./phages_list_1.txt','r'):
            if pass_first < 2:
                pass_first += 1
                pass
            else:
                index_tab = line.index("\t")
                access_ids.append(line[0:index_tab])

        return access_ids

    def write_gb_file(self, GIs, sub_dir):
        counter = 0

        for genome_id in GIs:
            counter += 1
            try:
                record = Entrez.efetch(db="nuccore", id=genome_id, rettype="gb", retmode=
                "text")

                filename = 'generated/{}genBankRecord_{}.gb'.format(sub_dir, genome_id)
                #print('{} - Writing:{}'.format(counter, filename))
                sys.stdout.write('\r%d - %s' % (counter, filename))
```

```
                with open(filename, 'w') as f:
                    f.write(record.read())
            except urllib.error.HTTPError as err:
                print(err.code)
        print('\n')

db_population = db_population()


# ## Uses

# In[9]:



access_ids = db_population.get_list_Phage_from_ebi()
print('Number Of Access Ids: %d' % (len(access_ids)))

#db_population.write_gb_file(access_ids[:10])
#db_population.write_gb_file(access_ids[11:20], '17052016-1038/')
#db_population.write_gb_file(access_ids[21:31], '17052016-1146/')
#db_population.write_gb_file(access_ids[32:62], '17052016-1224/')
#db_population.write_gb_file(access_ids[63:163], '17052016-1244/')
#db_population.write_gb_file(access_ids[164:1000], '17052016-1425/')
db_population.write_gb_file(access_ids, 'all_from_list_1')


# In[ ]:
```