#!/usr/bin/python

#credits: Rishabh Purwar ([rpurwar24@gmail.com](mailto:rpurwar24@gmail.com))

import MySQLdb

import httplib2

from bs4 import BeautifulSoup, SoupStrainer

**db = MySQLdb.connect("localhost", "<username>", "<password>", "<db>")**

cursor = db.cursor()

http = httplib2.Http()

def fetch\_data(source\_table, destination\_table):

sql = "CREATE TABLE IF NOT EXISTS `%s` (`id` int(11) NOT NULL AUTO\_INCREMENT,`name` varchar(100) DEFAULT NULL,`mobile` varchar(30) DEFAULT NULL,`address` text,`rating` varchar(10) DEFAULT NULL,`votes` varchar(10) DEFAULT NULL,`established` varchar(20) DEFAULT NULL,`website` varchar(200) DEFAULT NULL, `mon\_time` varchar(100) DEFAULT NULL, `tues\_time` varchar(100) DEFAULT NULL, `wed\_time` varchar(100) DEFAULT NULL, `thurs\_time` varchar(100) DEFAULT NULL, `fri\_time` varchar(100) DEFAULT NULL, `sat\_time` varchar(100) DEFAULT NULL, `sun\_time` varchar(100) DEFAULT NULL, PRIMARY KEY (`id`)) ENGINE=InnoDB AUTO\_INCREMENT=1 DEFAULT CHARSET=latin1;" % (destination\_table)

try:

cursor.execute(sql)

db.commit()

except:

db.rollback()

print "exception"

sql\_query = "SELECT \* FROM %s where visited = 0" % (source\_table)

cursor.execute(sql\_query)

results = cursor.fetchall()

for rows in results:

wp\_id = rows[0]

# Center name

name = rows[1]

request\_link = rows[2]

status, response = http.request(request\_link)

soup = BeautifulSoup(response)

# Telephone number

try:

a\_tag = soup.select(".tel")[0]

mobile = a\_tag.get\_text(strip=True)

except Exception, e:

mobile = "Null"

# Full address

try:

span\_tag = soup.select("#fulladdress")[0]

address = span\_tag.get\_text(strip=True)

except Exception, e:

address = "Null"

# Established year

try:

span\_tag = soup.select(".alstdul")[-1]

established = span\_tag.get\_text(strip=True)

except Exception, e:

established = "Null"

# Rating

try:

span\_tag = soup.select(".value-titles")[0]

rating = span\_tag.get\_text(strip=True)

except Exception, e:

rating = "Null"

# Votes

try:

span\_tag = soup.select(".votes")[0]

votes = span\_tag.get\_text(strip=True)

except Exception, e:

votes = "Null"

# Website

try:

p\_tag = soup.select(".mreinfp")[1]

website = p\_tag.a['href']

except Exception, e:

website = "Null"

try:

outer\_tag = soup.select('#statHr')

inner\_tag = outer\_tag[0].select('li')

span = inner\_tag[0].select("span")

mon\_time = span[1].text

span = inner\_tag[1].select("span")

tues\_time = span[1].text

span = inner\_tag[2].select("span")

wed\_time = span[1].text

span = inner\_tag[3].select("span")

thurs\_time = span[1].text

span = inner\_tag[4].select("span")

fri\_time = span[1].text

span = inner\_tag[5].select("span")

sat\_time = span[1].text

span = inner\_tag[6].select("span")

sun\_time = span[1].text

except Exception, e:

mon\_time = "Null"

tues\_time = "Null"

wed\_time = "Null"

thurs\_time = "Null"

fri\_time = "Null"

sat\_time = "Null"

sun\_time = "Null"

sql\_query = "INSERT INTO %s(name, mobile, address, established, rating, votes, website, \

mon\_time, tues\_time, wed\_time, thurs\_time, fri\_time, sat\_time, sun\_time) \

VALUES ('%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s', '%s')" % (destination\_table, name, mobile, address, established, rating, votes, website, mon\_time, tues\_time, wed\_time, thurs\_time, fri\_time, sat\_time, sun\_time)

try:

cursor.execute(sql\_query)

db.commit()

except Exception, e:

print("error in inserting data for id = %d") % (wp\_id)

print(e)

db.rollback()

sql = "update %s set visited = 1 where id = %s" % (source\_table, wp\_id)

try:

cursor.execute(sql)

db.commit()

except Exception, e:

print("error in updating data for id = %d") % (wp\_id)

print(e)

db.rollback()

print("succsessful for %d \n") % (wp\_id)

#print("Mobile - %s \n address - %s \n established in - %s \n rating - %s\nwebsite - %s\n working days - %s\nworking time - %s")%(mobile,address,established,ratings,website,working\_days,working\_time)

def fetch\_link(page\_no, table\_name, col1, col2):

sql = "CREATE TABLE IF NOT EXISTS `%s` (`id` int(11) NOT NULL AUTO\_INCREMENT,`%s` varchar(70) DEFAULT NULL,`%s` text, `visited` bit(1) DEFAULT 0, PRIMARY KEY (`id`),UNIQUE KEY `name\_UNIQUE` (`%s`)) ENGINE=InnoDB AUTO\_INCREMENT=1 DEFAULT CHARSET=latin1;" % (

table\_name, col1, col2, col1)

try:

cursor.execute(sql)

db.commit()

except:

db.rollback()

**# CHANGE\_THIS\_SEED\_URL**

**request\_link = "http://www.justdial.com/Delhi-Ncr/Martial-Art-Training-Centres/ct-465593/page-%d" % (**

**page\_no)**

status, response = http.request(request\_link)

soup = BeautifulSoup(response)

span\_tag\_col = soup.find\_all("span", attrs={"class": "jcn"})

for span\_tag in span\_tag\_col:

wp\_name = span\_tag.a.string

wp\_link = span\_tag.a['href']

print wp\_name

print wp\_link

sql = "INSERT INTO %s(%s,%s) VALUES ('%s','%s')" % (

table\_name, col1, col2, wp\_name, wp\_link)

try:

cursor.execute(sql)

db.commit()

except:

db.rollback()

for i in range(1, 27):

fetch\_link(i, "jd\_link", "name", "link")

# fetch\_data("jd\_link", "jd\_data")

db.close()