Computer Networks and Programming

(ECE 5650)

Project 3

Web page Downloader using multi-threading

Rajeev Bhupathiraju (gd4707@wayne.edu)

Sai Sundeep Innamuri (fx3898@wayne.edu)

Wayne State University

Winter 2017

Introduction

HTML (Hyper Text Markup Language):

Hyper Text Markup Language (HTML) is the standard markup language for creating web pages and web applications. With Cascading Style Sheets (CSS), and JavaScript, it forms a triad of cornerstone technologies for the World Wide Web. Web browsers receive HTML documents from a webserver or from local storage and render them into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.

Objective

The aim of the project is:

1. Enhancing webpage downloader in project 2 by supporting multithreading and obtaining all required features

2. Comparing the performance of multi-threaded web page downloader with previous project and tabulating the results and plot graphs

Code:

project2.py project3.py

```
######################################################################################
####
# Course        : Computer Networks and Programming, Winter 2017
# project       : project 3
# author        : Sai Sundeep Innamuri - 004507888, Rajeev Bhupathiraju - 004570122
# file name     : project3.py
# date Written  : Mar 26 2017
######################################################################################
####

from HTMLParser import HTMLParser
from socket import *
import sys
import os
import time
from datetime import datetime
from datetime import timedelta
import threading

HTTPGETRequestFormat = "GET {} HTTP/1.1\nHost: {}\n\n"
httpPort = 80
# serverSocket = socket(AF_INET, SOCK_STREAM)
domain_names = ('com','org','edu','in','uk')

# create a subclass and override the handler methods
class MyHTMLParser(HTMLParser):
    title = None
    srcs= list()
    next_tag_data_type = None

    def getAttrValue(self, attrs, key):
        for attr in attrs:
             if attr[0] == key: return  attr[1]
        return None

    def handle_starttag(self, tag, attrs):

        if tag == 'title': self.next_tag_data_type = 'title'
        elif tag == 'img':
            src = self.getAttrValue(attrs, "src")
            if src not in self.srcs:
                self.srcs.append(self.getAttrValue(attrs,"src"))
        else:
            src = self.getAttrValue(attrs,"src")
            href= self.getAttrValue(attrs,"href")
            attrval = None
            if src is not None: attrval = src
            if href is not None: attrval = href
            if attrval is None: return
            if attrval.find('http') != -1: return
            # print attrval
            self.srcs.append(attrval)


    def handle_data(self, data):
        if self.next_tag_data_type == 'title' :
            self.title = data.replace(" ","_").replace("\n","").replace(":", "_")
        self.next_tag_data_type = None


def writeFile(path, file, body):
    if not os.path.exists(path):
        # print os.path.basename(path)
        os.mkdir(path)
    elif file.rfind("/") != -1:
        os.mkdir(path + "/" + file[:file.rfind("/")])

    htmlfile = open(path + '/' + file, "wb")
    htmlfile.write(body)
    htmlfile.close()


#function to seperate header and body from a html response
def getHeadersNBody(response):

    headers = response[: response.index("\r\n\r\n")]
    body = response[response.index("\r\n\r\n") + 4:]
    headers = headers.split("\r\n")
    headersDict = dict()
    headersDict["status"]=headers[0][9:12]
    headers = headers[1:]
    # print headers
    #making a dictonary object for headers
    for header in headers:
        try:
            sp = header.split(": ")
            headersDict[sp[0]] = sp[1]
        except Exception, e:
            # print "cannot parse header" + header
            None
    return headersDict, body


#function to send and receive HTTP request
def sendRecvHTTPReq(host, HTTPRequest):
    # print HTTPRequest[:-2]
    serverSocket = socket(AF_INET, SOCK_STREAM)
    serverSocket.connect((host, httpPort))
    serverSocket.settimeout(1)
    #sending the http request
    serverSocket.send(HTTPRequest)
    #receving the response
    response = serverSocket.recv(10000)
    headers, body = getHeadersNBody(response)
    # print "status : " + headers["status"]

    # handling differt status codes
    if headers["status"] in ('302', '301', '307' ):
        # print "Encountered redirection. Not sipported!"
        serverSocket.close()
        raise Exception
        # protocal, host, request = extractDetailFromURL(headers["Location"])
        # serverSocket.close()
        # headers, body = sendRecvHTTPReq(host, HTTPGETRequestFormat.format(request,
host))
        # return headers, body

    elif headers["status"] != '200':
        # print response
        # print "Some error in request. HTTP error code " + headers["status"]
        # print "Exiting"
        serverSocket.close()
        raise Exception
        # sys.exit(2)

    #receving more messages if the entire conteent is not received in one request
    contentReceived = len(body)
    if "Content-Length" in headers.keys():
        # print "content is chunked. downloading in Iterations, Content length in
header keys"
        while (contentReceived < int(headers["Content-Length"])):
            body = body + serverSocket.recv(10000)
            contentReceived = len(body)
    elif "Transfer-Encoding" in headers.keys() and headers["Transfer-Encoding"] ==
"chunked":
        # print "content is chunked. downloading in Iterations. no content length"
        while (1):
            newChunk = serverSocket.recv(10000)
            if newChunk == "": break
            body = body + newChunk
            time.sleep(1)

    serverSocket.close()
    return headers,body

def checkForDomainInString(request):
    for domain in domain_names:
        if domain_names.find('.'+domain) != -1 : return domain
        else: continue
    return None


thread_list = []
srcs_to_replace = dict()


def objectDownloadThread(homepage, initreq, host, folderToSave):
    try:
        request = initreq
        request = request.replace(" ", "%20")

        if request[0] != "/":
            req = HTTPGETRequestFormat.format(homepage + request, host)
        elif request.find("//") != -1:
            nprotocal, nhost, nrequest = extractDetailFromURL(request)
            req = HTTPGETRequestFormat.format(nrequest, nhost)
        else:
            req = HTTPGETRequestFormat.format(request, host)

        headers, body = sendRecvHTTPReq(host, req)
        if initreq[-1] == "/": initreq = initreq[:-1]
        if initreq.rfind("/") != -1:
            savefile = initreq.replace("/", "_")
            # print "Save file "+ savefile
            if initreq not in srcs_to_replace.keys(): srcs_to_replace[initreq] = 
savefile
            writeFile(folderToSave, savefile, body)
        else:
            writeFile(folderToSave, initreq, body)
    except:
        None
        # print "Exception, Proceeding wtith next"


#function to download an array of bojects from src tag
def processReqArray(homepage, reqArray, host, folderToSave):
    # print reqArray
    numObjects = 0

    for initreq in reqArray:
        t = threading.Thread(target=objectDownloadThread, args=(homepage, initreq, 
host, folderToSave))
        thread_list.append(t)

    for thread in thread_list:
        thread.start()

    for i in range (0, len(thread_list)):
        thread_list[i].join()
        sys.stdout.write("\rThread Joined for Object: %d" %(i+1))
        sys.stdout.flush()

    print ""
    return numObjects, srcs_to_replace


#to extract details from URL
def extractDetailFromURL(htmlReq):
    protocal = htmlReq.split("//")
    host = protocal[1].split("/")[0]
    request = protocal[1][protocal[1].index('/'):]
    protocal = protocal[0]
    return protocal, host, request


def main():
    parser = MyHTMLParser()
    try:
        start_time = datetime.now()
        print "start_time : " + str(start_time)
        #storing the web site request in to a variable
        protocal, host, indexReq = extractDetailFromURL(sys.argv[1])
        headers,body = sendRecvHTTPReq(host, HTTPGETRequestFormat.format(indexReq,
host))
        websiteHome = indexReq if indexReq == "/" else
indexReq[:indexReq[1:].index("/")+2]
        #to parse the HTML files
        parser.feed(body)

        if os.path.exists(parser.title):
            for root, dirs, files in os.walk(parser.title, topdown=False):
                for name in files:
                    os.remove(os.path.join(root, name))
                for name in dirs:
                    os.rmdir(os.path.join(root, name))
        else:  os.mkdir(parser.title)
        # htmlfile = open(parser.title + "/" + parser.title +".html" , "wb")

        #to download the images
        num_objects, srcs_to_replace = processReqArray(websiteHome, parser.srcs, host,
parser.title)

        end_time = datetime.now()
        print "end_time : " + str(end_time)

        for k,v in srcs_to_replace.items():
            # print "replacing keyy " +k+ " with "+v
            body = body.replace(k, v)
        # print "writing file : " + parser.title +".html"
        writeFile(parser.title, parser.title +".html", body )
        print "totaltime : "+ str(timedelta.total_seconds(end_time - start_time))
    except Exception,e :
        None
        print str(e) + " ignored"
    finally:
        None
        if os.path.exists(parser.title):
            print ""
            print "please check {} folder for downloaded html
website".format(parser.title)

if __name__ == '__main__':
    main()
```

Testing procedure:

1) We have code for project 2 and project 3 in this submission
2) Project 2 code is named project2.py
    a. To run project 2: python project2.py <any URL>
3) Project 3 code with multi-threading is named project3.py
    a. To run project 3: python project3.py <any URL>
4) We have tested with four URL. They are

1. http://www.ece.eng.wayne.edu/~nabil/
2. http://ece.eng.wayne.edu/~hying/
3. http://ece.eng.wayne.edu/~apandya/
4. http://imgur.com/

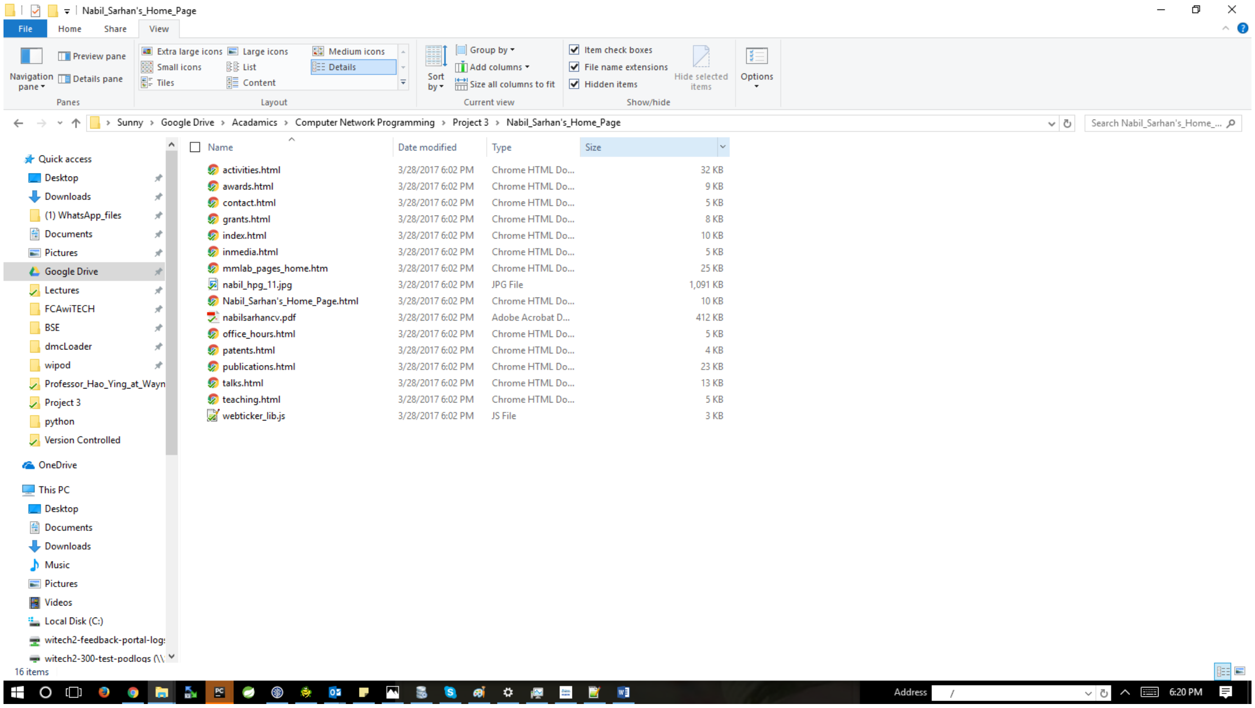
we are submitting project3.py and project2.py code but we are copying only project3.py code in this document to reduce length of the document

URL #1: http://www.ece.eng.wayne.edu/~nabil/

Max object size = 1091 KB

Min object size = 3 KB

Items Downloaded:

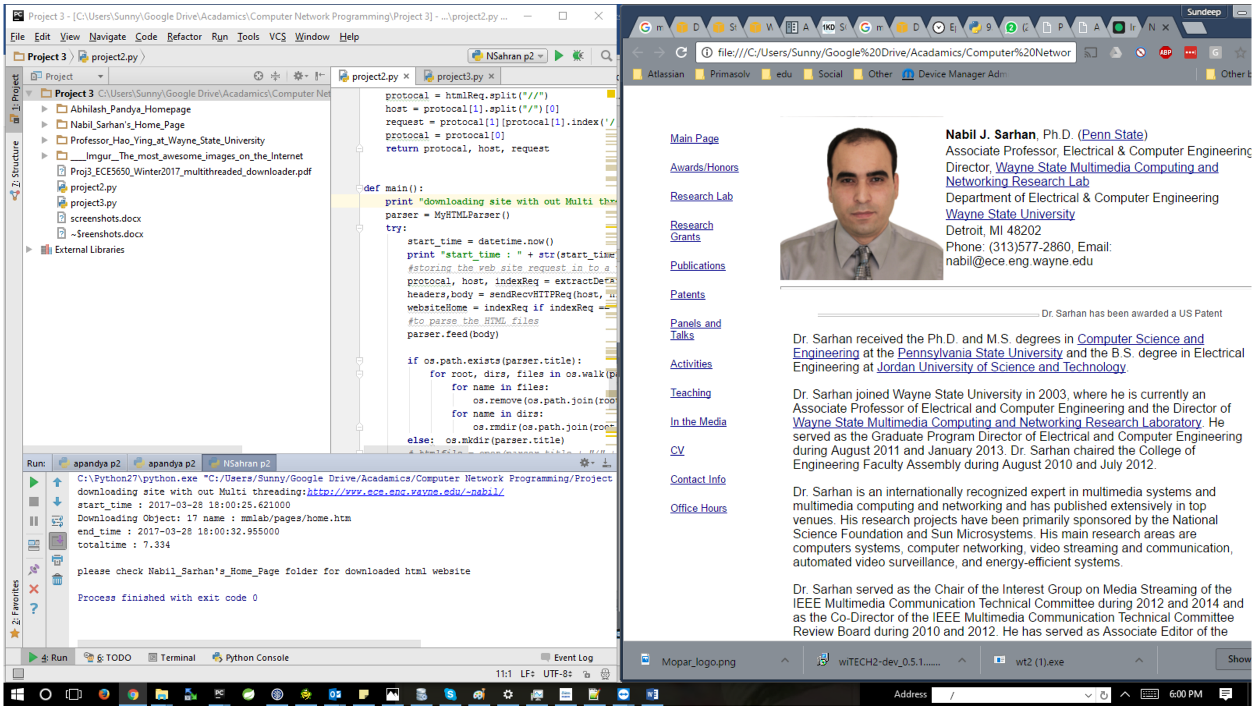


Sunny > Google Drive > Acadamics > Computer Network Programming > Project 3 > Nabil_Sarhan's_Home_Page

Name	Date modified	Type	Size
activities.html	3/28/2017 6:02 PM	Chrome HTML Do...	32 KB
awards.html	3/28/2017 6:02 PM	Chrome HTML Do...	9 KB
contact.html	3/28/2017 6:02 PM	Chrome HTML Do...	5 KB
grants.html	3/28/2017 6:02 PM	Chrome HTML Do...	8 KB
index.html	3/28/2017 6:02 PM	Chrome HTML Do...	10 KB
inmedia.html	3/28/2017 6:02 PM	Chrome HTML Do...	5 KB
mmlab_pages_home.htm	3/28/2017 6:02 PM	Chrome HTML Do...	25 KB
nabil_hpg_11.jpg	3/28/2017 6:02 PM	JPG File	1,091 KB
Nabil_Sarhan's_Home_Page.html	3/28/2017 6:02 PM	Chrome HTML Do...	10 KB
nabilsarhancv.pdf	3/28/2017 6:02 PM	Adobe Acrobat D...	412 KB
office_hours.html	3/28/2017 6:02 PM	Chrome HTML Do...	5 KB
patents.html	3/28/2017 6:02 PM	Chrome HTML Do...	4 KB
publications.html	3/28/2017 6:02 PM	Chrome HTML Do...	23 KB
talks.html	3/28/2017 6:02 PM	Chrome HTML Do...	13 KB
teaching.html	3/28/2017 6:02 PM	Chrome HTML Do...	5 KB
webticker_lib.js	3/28/2017 6:02 PM	JS File	3 KB

16 items

Project 2 output


Project 3 - [C:\Users\Sunny\Google Drive\Acadamics\Computer Network Programming\Project 3] - ...\project2.py ...
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Project 3 project2.py
NSahran p2
project2.py
project3.py
Project 3 C:\Users\Sunny\Google Drive\Acadamics\Computer Net
Abhilash_Pandya_Homepage
Nabil_Sarhan's_Home_Page
Professor_Hao_Ying_at_Wayne_State_University
___Imgur__The_most_awesome_images_on_the_Internet
Proj3_ECE5650_Winter2017_multithreaded_downloader.pdf
project2.py
project3.py
screenshots.docx
~$reenshots.docx
External Libraries

```
    protocal = htmlReq.split("//")
    host = protocal[1].split("/")[0]
    request = protocal[1][protocal[1].index('/
    protocal = protocal[0]
    return protocal, host, request


def main():
    print "downloading site with out Multi thr
    parser = MyHTMLParser()
    try:
        start_time = datetime.now()
        print "start_time : " + str(start_time
        #storing the web site request in to a
        protocal, host, indexReq = extractDeta
        headers,body = sendRecvHTTPReq(host,
        websiteHome = indexReq if indexReq ==
        #to parse the HTML files
        parser.feed(body)

        if os.path.exists(parser.title):
            for root, dirs, files in os.walk(pa
                for name in files:
                    os.remove(os.path.join(roo
                for name in dirs:
                    os.rmdir(os.path.join(roo
        else:  os.mkdir(parser.title)
```

```
C:\Python27\python.exe "C:/Users/Sunny/Google Drive/Acadamics/Computer Network Programming/Project
downloading site with out Multi threading:http://www.ece.eng.wayne.edu/~nabil/
start_time : 2017-03-28 18:00:25.621000
Downloading Object: 17 name : mmlab/pages/home.htm
end_time : 2017-03-28 18:00:32.955000
totaltime : 7.334

please check Nabil_Sarhan's_Home_Page folder for downloaded html website

Process finished with exit code 0
```

Nabil J. Sarhan, Ph.D. (Penn State)
Associate Professor, Electrical & Computer Engineering
Director, Wayne State Multimedia Computing and Networking Research Lab
Department of Electrical & Computer Engineering
Wayne State University
Detroit, MI 48202
Phone: (313)577-2860, Email: nabil@ece.eng.wayne.edu

Dr. Sarhan has been awarded a US Patent

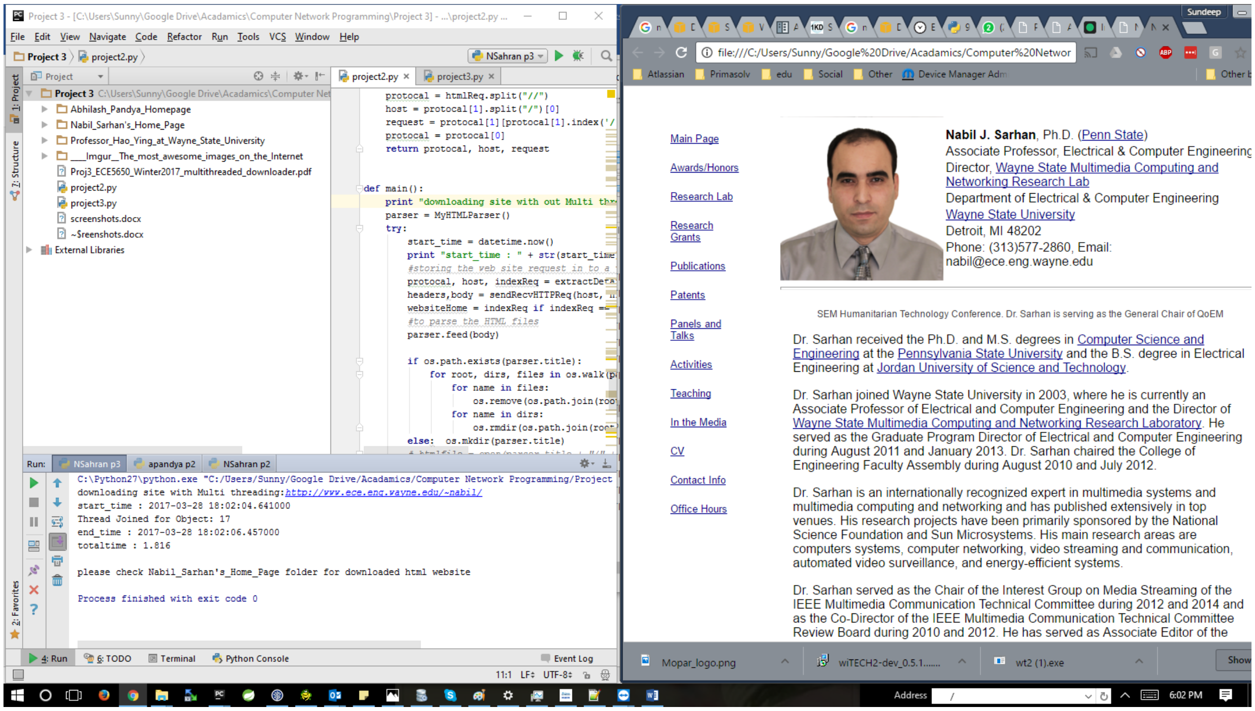
Dr. Sarhan received the Ph.D. and M.S. degrees in Computer Science and Engineering at the Pennsylvania State University and the B.S. degree in Electrical Engineering at Jordan University of Science and Technology.

Dr. Sarhan joined Wayne State University in 2003, where he is currently an Associate Professor of Electrical and Computer Engineering and the Director of Wayne State Multimedia Computing and Networking Research Laboratory. He served as the Graduate Program Director of Electrical and Computer Engineering during August 2011 and January 2013. Dr. Sarhan chaired the College of Engineering Faculty Assembly during August 2010 and July 2012.

Dr. Sarhan is an internationally recognized expert in multimedia systems and multimedia computing and networking and has published extensively in top venues. His research projects have been primarily sponsored by the National Science Foundation and Sun Microsystems. His main research areas are computers systems, computer networking, video streaming and communication, automated video surveillance, and energy-efficient systems.

Dr. Sarhan served as the Chair of the Interest Group on Media Streaming of the IEEE Multimedia Communication Technical Committee during 2012 and 2014 and as the Co-Director of the IEEE Multimedia Communication Technical Committee Review Board during 2010 and 2012. He has served as Associate Editor of the

Project 3 Output


Project 3 - [C:\Users\Sunny\Google Drive\Acadamics\Computer Network Programming\Project 3] - ...\project2.py ...
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Project 3 project2.py
NSahran p3
project2.py
project3.py
Project 3 C:\Users\Sunny\Google Drive\Acadamics\Computer Net
Abhilash_Pandya_Homepage
Nabil_Sarhan's_Home_Page
Professor_Hao_Ying_at_Wayne_State_University
___Imgur__The_most_awesome_images_on_the_Internet
Proj3_ECE5650_Winter2017_multithreaded_downloader.pdf
project2.py
project3.py
screenshots.docx
~$reenshots.docx
External Libraries

```
    protocal = htmlReq.split("//")
    host = protocal[1].split("/")[0]
    request = protocal[1][protocal[1].index('/
    protocal = protocal[0]
    return protocal, host, request


def main():
    print "downloading site with out Multi thr
    parser = MyHTMLParser()
    try:
        start_time = datetime.now()
        print "start_time : " + str(start_time
        #storing the web site request in to a
        protocal, host, indexReq = extractDeta
        headers,body = sendRecvHTTPReq(host,
        websiteHome = indexReq if indexReq ==
        #to parse the HTML files
        parser.feed(body)

        if os.path.exists(parser.title):
            for root, dirs, files in os.walk(pa
                for name in files:
                    os.remove(os.path.join(roo
                for name in dirs:
                    os.rmdir(os.path.join(roo
        else:  os.mkdir(parser.title)
```

```
C:\Python27\python.exe "C:/Users/Sunny/Google Drive/Acadamics/Computer Network Programming/Project
downloading site with Multi threading:http://www.ece.eng.wayne.edu/~nabil/
start_time : 2017-03-28 18:02:04.641000
Thread Joined for Object: 17
end_time : 2017-03-28 18:02:06.457000
totaltime : 1.816

please check Nabil_Sarhan's_Home_Page folder for downloaded html website

Process finished with exit code 0
```

Nabil J. Sarhan, Ph.D. (Penn State)
Associate Professor, Electrical & Computer Engineering
Director, Wayne State Multimedia Computing and Networking Research Lab
Department of Electrical & Computer Engineering
Wayne State University
Detroit, MI 48202
Phone: (313)577-2860, Email: nabil@ece.eng.wayne.edu

SEM Humanitarian Technology Conference. Dr. Sarhan is serving as the General Chair of QoEM

Dr. Sarhan received the Ph.D. and M.S. degrees in Computer Science and Engineering at the Pennsylvania State University and the B.S. degree in Electrical Engineering at Jordan University of Science and Technology.

Dr. Sarhan joined Wayne State University in 2003, where he is currently an Associate Professor of Electrical and Computer Engineering and the Director of Wayne State Multimedia Computing and Networking Research Laboratory. He served as the Graduate Program Director of Electrical and Computer Engineering during August 2011 and January 2013. Dr. Sarhan chaired the College of Engineering Faculty Assembly during August 2010 and July 2012.

Dr. Sarhan is an internationally recognized expert in multimedia systems and multimedia computing and networking and has published extensively in top venues. His research projects have been primarily sponsored by the National Science Foundation and Sun Microsystems. His main research areas are computers systems, computer networking, video streaming and communication, automated video surveillance, and energy-efficient systems.

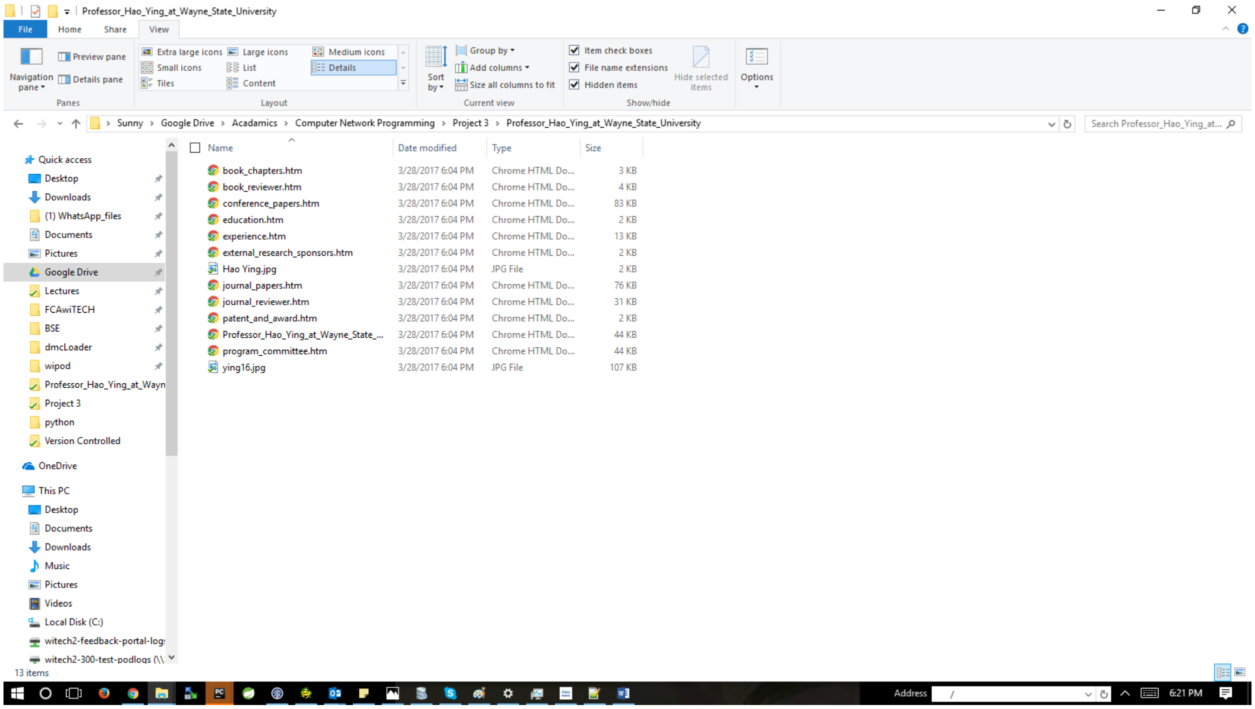
Dr. Sarhan served as the Chair of the Interest Group on Media Streaming of the IEEE Multimedia Communication Technical Committee during 2012 and 2014 and as the Co-Director of the IEEE Multimedia Communication Technical Committee Review Board during 2010 and 2012. He has served as Associate Editor of the

URL #2: http://ece.eng.wayne.edu/~hying/

Max object size: 107 KB

Min object size: 2 KB

Objects downloaded:

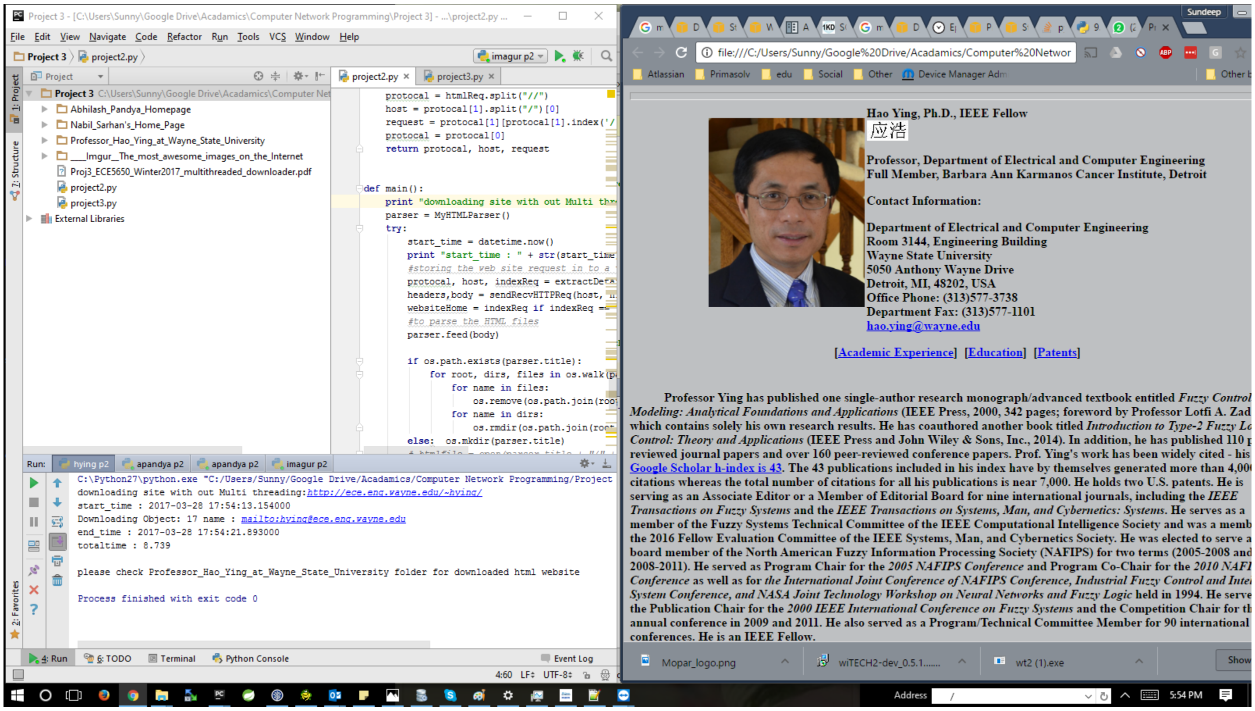


Sunny › Google Drive › Acadamics › Computer Network Programming › Project 3 › Professor_Hao_Ying_at_Wayne_State_University

Name	Date modified	Type	Size
book_chapters.htm	3/28/2017 6:04 PM	Chrome HTML Do...	3 KB
book_reviewer.htm	3/28/2017 6:04 PM	Chrome HTML Do...	4 KB
conference_papers.htm	3/28/2017 6:04 PM	Chrome HTML Do...	83 KB
education.htm	3/28/2017 6:04 PM	Chrome HTML Do...	2 KB
experience.htm	3/28/2017 6:04 PM	Chrome HTML Do...	13 KB
external_research_sponsors.htm	3/28/2017 6:04 PM	Chrome HTML Do...	2 KB
Hao Ying.jpg	3/28/2017 6:04 PM	JPG File	2 KB
journal_papers.htm	3/28/2017 6:04 PM	Chrome HTML Do...	76 KB
journal_reviewer.htm	3/28/2017 6:04 PM	Chrome HTML Do...	31 KB
patent_and_award.htm	3/28/2017 6:04 PM	Chrome HTML Do...	2 KB
Professor_Hao_Ying_at_Wayne_State_...	3/28/2017 6:04 PM	Chrome HTML Do...	44 KB
program_committee.htm	3/28/2017 6:04 PM	Chrome HTML Do...	44 KB
ying16.jpg	3/28/2017 6:04 PM	JPG File	107 KB

13 items

Project 2 output`



```
    protocal = htmlReq.split("//")
    host = protocal[1].split("/")[0]
    request = protocal[1][protocal[1].index('/
    protocal = protocal[0]
    return protocal, host, request


def main():
    print "downloading site with out Multi thr
    parser = MyHTMLParser()
    try:
        start_time = datetime.now()
        print "start_time : " + str(start_time
        #storing the web site request in to a
        protocal, host, indexReq = extractDeta
        headers,body = sendRecvHTTPReq(host, i
        websiteHome = indexReq if indexReq ==
        #to parse the HTML files
        parser.feed(body)

        if os.path.exists(parser.title):
            for root, dirs, files in os.walk(p
                for name in files:
                    os.remove(os.path.join(roo
                for name in dirs:
                    os.rmdir(os.path.join(roo
        else:  os.mkdir(parser.title)
```

```
C:\Python27\python.exe "C:/Users/Sunny/Google Drive/Acadamics/Computer Network Programming/Project
downloading site with out Multi threading:http://ece.eng.wayne.edu/~hying/
start_time : 2017-03-28 17:54:13.154000
Downloading Object: 17 name : mailto:hying@ece.eng.wayne.edu
end_time : 2017-03-28 17:54:21.893000
totaltime : 8.739

please check Professor_Hao_Ying_at_Wayne_State_University folder for downloaded html website

Process finished with exit code 0
```

Hao Ying, Ph.D., IEEE Fellow
应浩

Professor, Department of Electrical and Computer Engineering
Full Member, Barbara Ann Karmanos Cancer Institute, Detroit

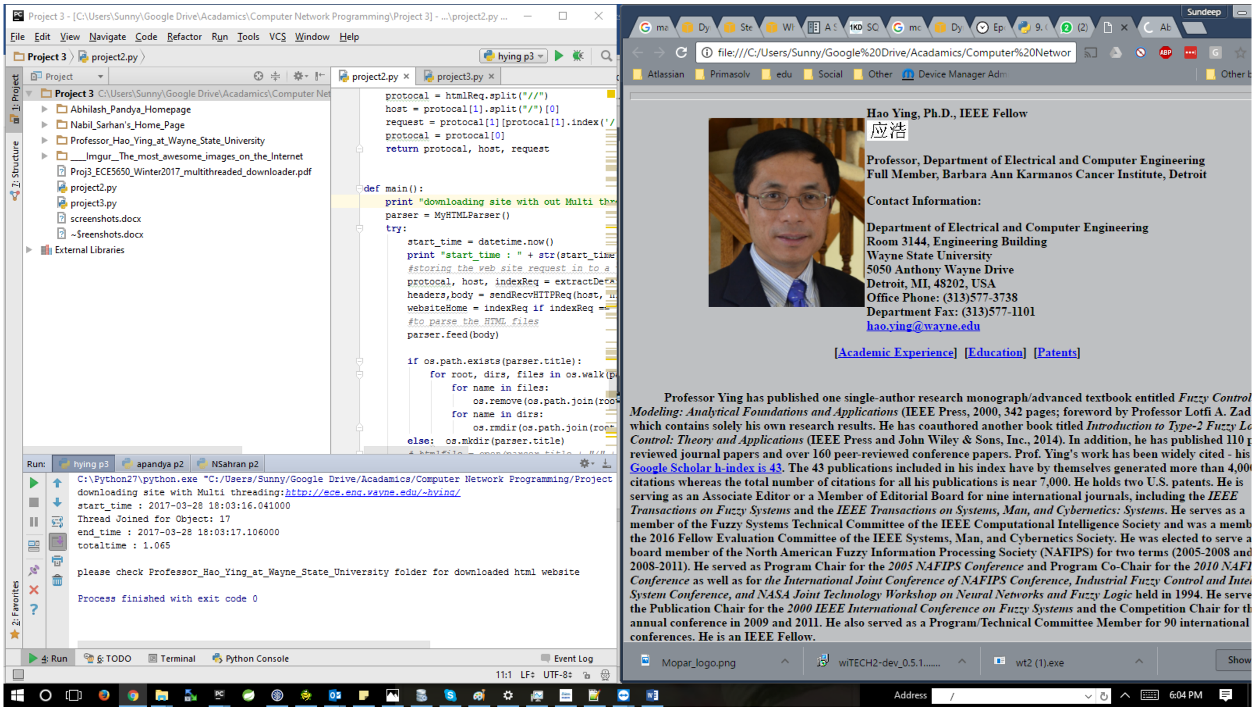
Contact Information:

Department of Electrical and Computer Engineering
Room 3144, Engineering Building
Wayne State University
5050 Anthony Wayne Drive
Detroit, MI, 48202, USA
Office Phone: (313)577-3738
Department Fax: (313)577-1101
hao.ying@wayne.edu

[Academic Experience] [Education] [Patents]

Professor Ying has published one single-author research monograph/advanced textbook entitled *Fuzzy Control Modeling: Analytical Foundations and Applications* (IEEE Press, 2000, 342 pages; foreword by Professor Lotfi A. Zad which contains solely his own research results. He has coauthored another book titled *Introduction to Type-2 Fuzzy Lo Control: Theory and Applications* (IEEE Press and John Wiley & Sons, Inc., 2014). In addition, he has published 110 p reviewed journal papers and over 160 peer-reviewed conference papers. Prof. Ying's work has been widely cited - his Google Scholar h-index is 43. The 43 publications included in his index have by themselves generated more than 4,00 citations whereas the total number of citations for all his publications is near 7,000. He holds two U.S. patents. He is serving as an Associate Editor or a Member of Editorial Board for nine international journals, including the *IEEE Transactions on Fuzzy Systems* and the *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. He serves as a member of the Fuzzy Systems Technical Committee of the IEEE Computational Intelligence Society and was a memb the 2016 Fellow Evaluation Committee of the IEEE Systems, Man, and Cybernetics Society. He was elected to serve a board member of the North American Fuzzy Information Processing Society (NAFIPS) for two terms (2005-2008 and 2008-2011). He served as Program Chair for the *2005 NAFIPS Conference* and Program Co-Chair for the *2010 NAFI Conference* as well as for *the International Joint Conference of NAFIPS Conference, Industrial Fuzzy Control and Inte System Conference, and NASA Joint Technology Workshop on Neural Networks and Fuzzy Logic* held in 1994. He serve the Publication Chair for the *2000 IEEE International Conference on Fuzzy Systems* and the Competition Chair for th annual conference in 2009 and 2011. He also served as a Program/Technical Committee Member for 90 international conferences. He is an IEEE Fellow.

Project 3 output



```
    protocal = htmlReq.split("//")
    host = protocal[1].split("/")[0]
    request = protocal[1][protocal[1].index('/
    protocal = protocal[0]
    return protocal, host, request


def main():
    print "downloading site with out Multi thr
    parser = MyHTMLParser()
    try:
        start_time = datetime.now()
        print "start_time : " + str(start_time
        #storing the web site request in to a
        protocal, host, indexReq = extractDeta
        headers,body = sendRecvHTTPReq(host, i
        websiteHome = indexReq if indexReq ==
        #to parse the HTML files
        parser.feed(body)

        if os.path.exists(parser.title):
            for root, dirs, files in os.walk(p
                for name in files:
                    os.remove(os.path.join(roo
                for name in dirs:
                    os.rmdir(os.path.join(roo
        else:  os.mkdir(parser.title)
```

```
C:\Python27\python.exe "C:/Users/Sunny/Google Drive/Acadamics/Computer Network Programming/Project
downloading site with Multi threading:http://ece.eng.wayne.edu/~hying/
start_time : 2017-03-28 18:03:16.041000
Thread Joined for Object: 17
end_time : 2017-03-28 18:03:17.106000
totaltime : 1.065

please check Professor_Hao_Ying_at_Wayne_State_University folder for downloaded html website

Process finished with exit code 0
```

Hao Ying, Ph.D., IEEE Fellow
应浩

Professor, Department of Electrical and Computer Engineering
Full Member, Barbara Ann Karmanos Cancer Institute, Detroit

Contact Information:

Department of Electrical and Computer Engineering
Room 3144, Engineering Building
Wayne State University
5050 Anthony Wayne Drive
Detroit, MI, 48202, USA
Office Phone: (313)577-3738
Department Fax: (313)577-1101
hao.ying@wayne.edu

[Academic Experience] [Education] [Patents]

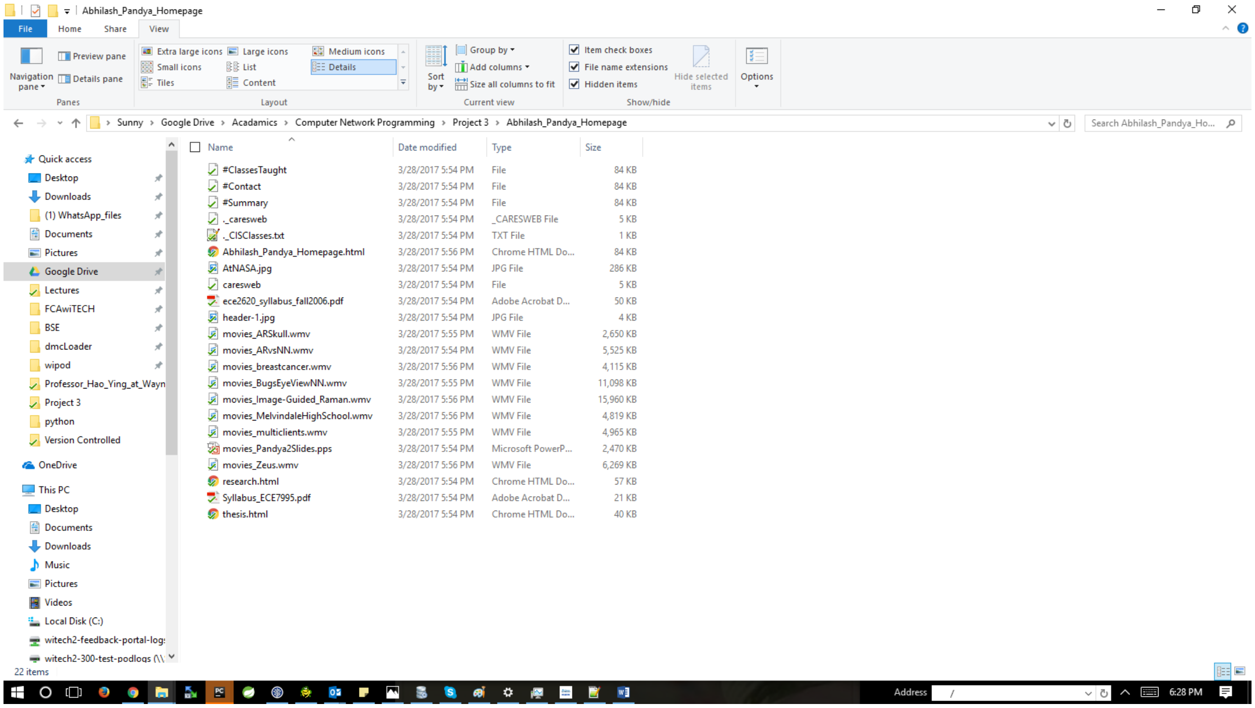
Professor Ying has published one single-author research monograph/advanced textbook entitled *Fuzzy Control Modeling: Analytical Foundations and Applications* (IEEE Press, 2000, 342 pages; foreword by Professor Lotfi A. Zad which contains solely his own research results. He has coauthored another book titled *Introduction to Type-2 Fuzzy Lo Control: Theory and Applications* (IEEE Press and John Wiley & Sons, Inc., 2014). In addition, he has published 110 p reviewed journal papers and over 160 peer-reviewed conference papers. Prof. Ying's work has been widely cited - his Google Scholar h-index is 43. The 43 publications included in his index have by themselves generated more than 4,00 citations whereas the total number of citations for all his publications is near 7,000. He holds two U.S. patents. He is serving as an Associate Editor or a Member of Editorial Board for nine international journals, including the *IEEE Transactions on Fuzzy Systems* and the *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. He serves as a member of the Fuzzy Systems Technical Committee of the IEEE Computational Intelligence Society and was a memb the 2016 Fellow Evaluation Committee of the IEEE Systems, Man, and Cybernetics Society. He was elected to serve a board member of the North American Fuzzy Information Processing Society (NAFIPS) for two terms (2005-2008 and 2008-2011). He served as Program Chair for the *2005 NAFIPS Conference* and Program Co-Chair for the *2010 NAFI Conference* as well as for *the International Joint Conference of NAFIPS Conference, Industrial Fuzzy Control and Inte System Conference, and NASA Joint Technology Workshop on Neural Networks and Fuzzy Logic* held in 1994. He serve the Publication Chair for the *2000 IEEE International Conference on Fuzzy Systems* and the Competition Chair for th annual conference in 2009 and 2011. He also served as a Program/Technical Committee Member for 90 international conferences. He is an IEEE Fellow.

URL #3: http://ece.eng.wayne.edu/~apandya/

Max object size: 15960 KB

Min object size: 1KB

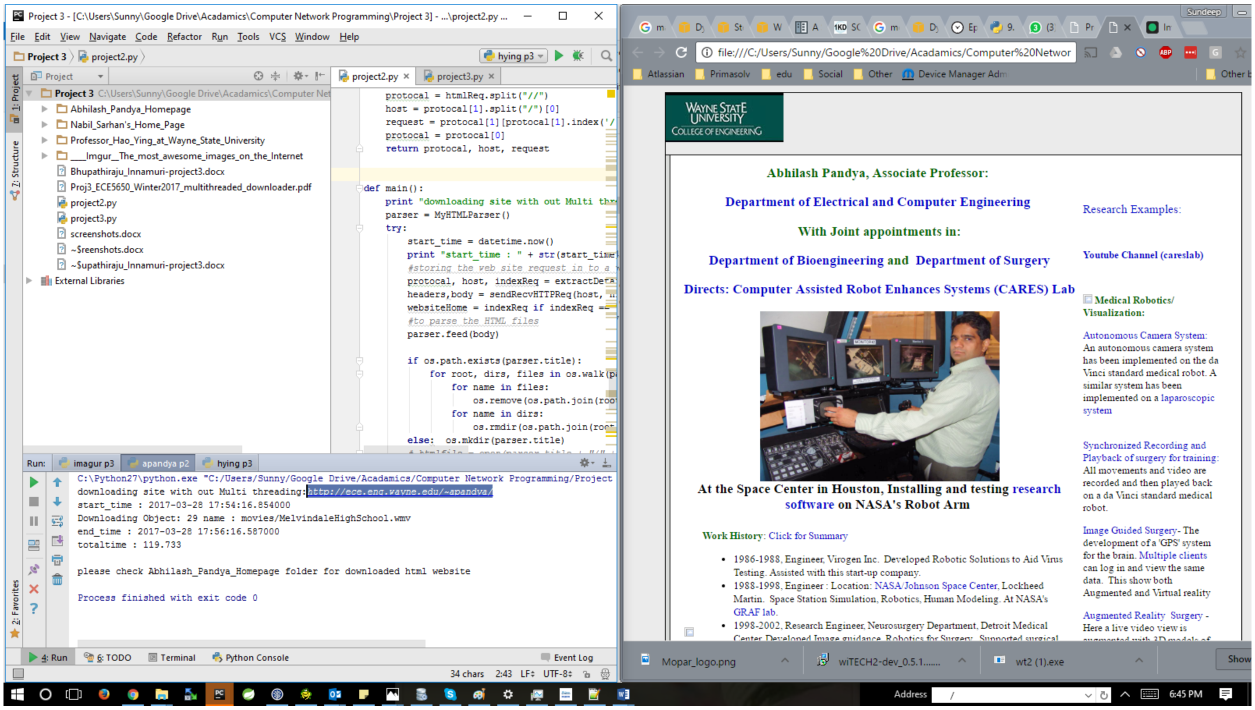
Items Downloaded:



Sunny › Google Drive › Acadamics › Computer Network Programming › Project 3 › Abhilash_Pandya_Homepage

Name	Date modified	Type	Size
#ClassesTaught	3/28/2017 5:54 PM	File	84 KB
#Contact	3/28/2017 5:54 PM	File	84 KB
#Summary	3/28/2017 5:54 PM	File	84 KB
._caresweb	3/28/2017 5:54 PM	_CARESWEB File	5 KB
._CISClasses.txt	3/28/2017 5:54 PM	TXT File	1 KB
Abhilash_Pandya_Homepage.html	3/28/2017 5:56 PM	Chrome HTML Do...	84 KB
AtNASA.jpg	3/28/2017 5:54 PM	JPG File	286 KB
caresweb	3/28/2017 5:54 PM	File	5 KB
ece2620_syllabus_fall2006.pdf	3/28/2017 5:54 PM	Adobe Acrobat D...	50 KB
header-1.jpg	3/28/2017 5:54 PM	JPG File	4 KB
movies_ARSkull.wmv	3/28/2017 5:55 PM	WMV File	2,650 KB
movies_ARvsNN.wmv	3/28/2017 5:54 PM	WMV File	5,525 KB
movies_breastcancer.wmv	3/28/2017 5:56 PM	WMV File	4,115 KB
movies_BugsEyeViewNN.wmv	3/28/2017 5:55 PM	WMV File	11,098 KB
movies_Image-Guided_Raman.wmv	3/28/2017 5:56 PM	WMV File	15,960 KB
movies_MelvindaleHighSchool.wmv	3/28/2017 5:56 PM	WMV File	4,819 KB
movies_multiclients.wmv	3/28/2017 5:55 PM	WMV File	4,965 KB
movies_Pandya2Slides.pps	3/28/2017 5:54 PM	Microsoft PowerP...	2,470 KB
movies_Zeus.wmv	3/28/2017 5:56 PM	WMV File	6,269 KB
research.html	3/28/2017 5:54 PM	Chrome HTML Do...	57 KB
Syllabus_ECE7995.pdf	3/28/2017 5:54 PM	Adobe Acrobat D...	21 KB
thesis.html	3/28/2017 5:54 PM	Chrome HTML Do...	40 KB

Project 2 output


Project 3 - [C:\Users\Sunny\Google Drive\Acadamics\Computer Network Programming\Project 3] - ...\project2.py ...
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Project 3 C:\Users\Sunny\Google Drive\Acadamics\Computer Net
Abhilash_Pandya_Homepage
Nabil_Sarhan's_Home_Page
Professor_Hao_Ying_at_Wayne_State_University
___Imgur__The_most_awesome_images_on_the_Internet
Bhupathiraju_Innamuri-project3.docx
Proj3_ECE5650_Winter2017_multithreaded_downloader.pdf
project2.py
project3.py
screenshots.docx
~$reenshots.docx
~$upathiraju_Innamuri-project3.docx
External Libraries
hying p3
project2.py
project3.py
Run: imagur p3 apandya p2 hying p3
4: Run 6: TODO Terminal Python Console Event Log
34 chars 2:43 LF UTF-8
file:///C:/Users/Sunny/Google%20Drive/Acadamics/Computer%20Networ
Atlassian Primasolv edu Social Other Device Manager Adm
Mopar_logo.png wiTECH2-dev_0.5.1...... wt2 (1).exe Show
Address /
6:45 PM

```
    protocal = htmlReq.split("//")
    host = protocal[1].split("/")[0]
    request = protocal[1][protocal[1].index('/
    protocal = protocal[0]
    return protocal, host, request


def main():
    print "downloading site with out Multi thr
    parser = MyHTMLParser()
    try:
        start_time = datetime.now()
        print "start_time : " + str(start_time
        #storing the web site request in to a
        protocal, host, indexReq = extractDeta
        headers,body = sendRecvHTTPReq(host, 
        websiteHome = indexReq if indexReq ==
        #to parse the HTML files
        parser.feed(body)

        if os.path.exists(parser.title):
            for root, dirs, files in os.walk(p
                for name in files:
                    os.remove(os.path.join(roo
                for name in dirs:
                    os.rmdir(os.path.join(roo
        else:  os.mkdir(parser.title)
```

```
C:\Python27\python.exe "C:/Users/Sunny/Google Drive/Acadamics/Computer Network Programming/Project
downloading site with out Multi threading:http://ece.eng.wayne.edu/~apandya/
start_time : 2017-03-28 17:54:16.854000
Downloading Object: 29 name : movies/MelvindaleHighSchool.wmv
end_time : 2017-03-28 17:56:16.587000
totaltime : 119.733

please check Abhilash_Pandya_Homepage folder for downloaded html website

Process finished with exit code 0
```

WAYNE STATE UNIVERSITY COLLEGE OF ENGINEERING

Abhilash Pandya, Associate Professor:

Department of Electrical and Computer Engineering

With Joint appointments in:

Department of Bioengineering and Department of Surgery

Directs: Computer Assisted Robot Enhances Systems (CARES) Lab

At the Space Center in Houston, Installing and testing research software on NASA's Robot Arm

Work History: Click for Summary

- 1986-1988, Engineer, Virogen Inc. Developed Robotic Solutions to Aid Virus Testing. Assisted with this start-up company.
- 1988-1998, Engineer : Location: NASA/Johnson Space Center, Lockheed Martin. Space Station Simulation, Robotics, Human Modeling. At NASA's GRAF lab.
- 1998-2002, Research Engineer, Neurosurgery Department, Detroit Medical Center. Developed Image guidance, Robotics for Surgery. Supported surgical

Research Examples:

Youtube Channel (careslab)

Medical Robotics/ Visualization:

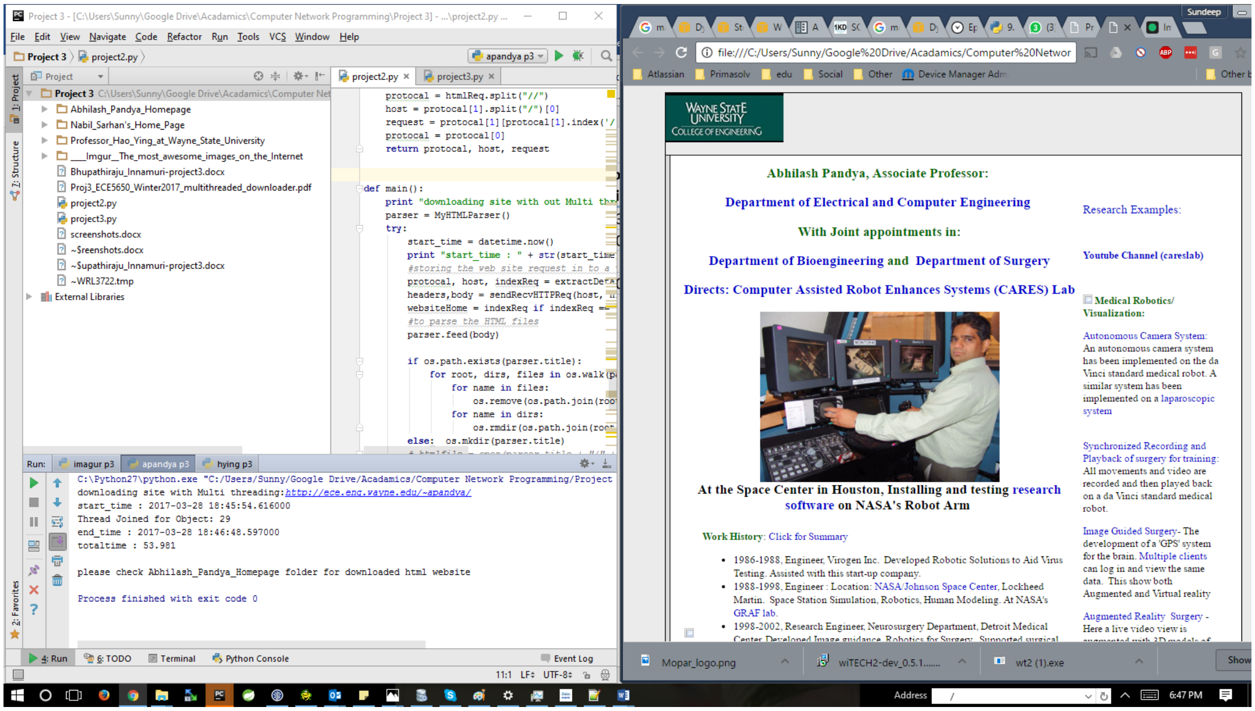
Autonomous Camera System: An autonomous camera system has been implemented on the da Vinci standard medical robot. A similar system has been implemented on a laparoscopic system

Synchronized Recording and Playback of surgery for training: All movements and video are recorded and then played back on a da Vinci standard medical robot.

Image Guided Surgery- The development of a 'GPS' system for the brain. Multiple clients can log in and view the same data. This show both Augmented and Virtual reality

Augmented Reality Surgery - Here a live video view is

Project 3 output


Project 3 - [C:\Users\Sunny\Google Drive\Acadamics\Computer Network Programming\Project 3] - ...\project2.py ...
~WRL3722.tmp
apandya p3
Run: imagur p3 apandya p3 hying p3
11:1 LF UTF-8
6:47 PM

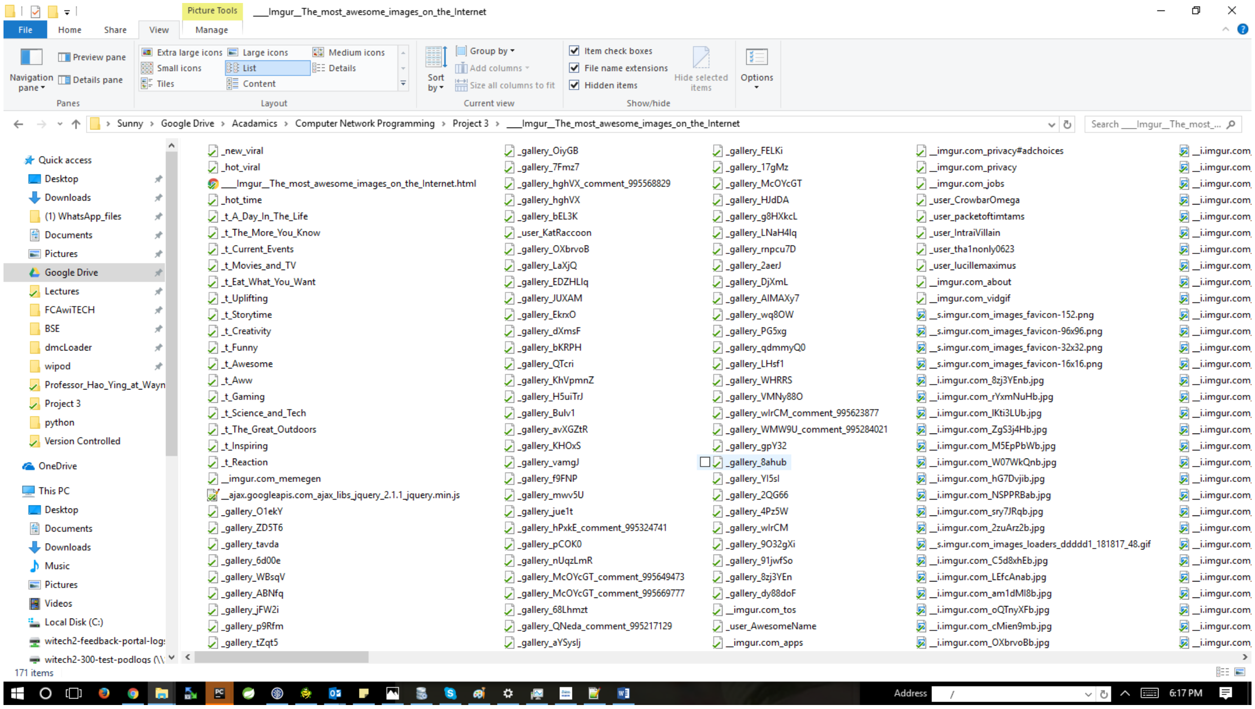
```
C:\Python27\python.exe "C:/Users/Sunny/Google Drive/Acadamics/Computer Network Programming/Project
downloading site with Multi threading:http://ece.eng.wayne.edu/~apandya/
start_time : 2017-03-28 18:45:54.616000
Thread Joined for Object: 29
end_time : 2017-03-28 18:46:48.597000
totaltime : 53.981

please check Abhilash_Pandya_Homepage folder for downloaded html website

Process finished with exit code 0
```

URL: http://imgur.com/

Max object Size = 4 KB

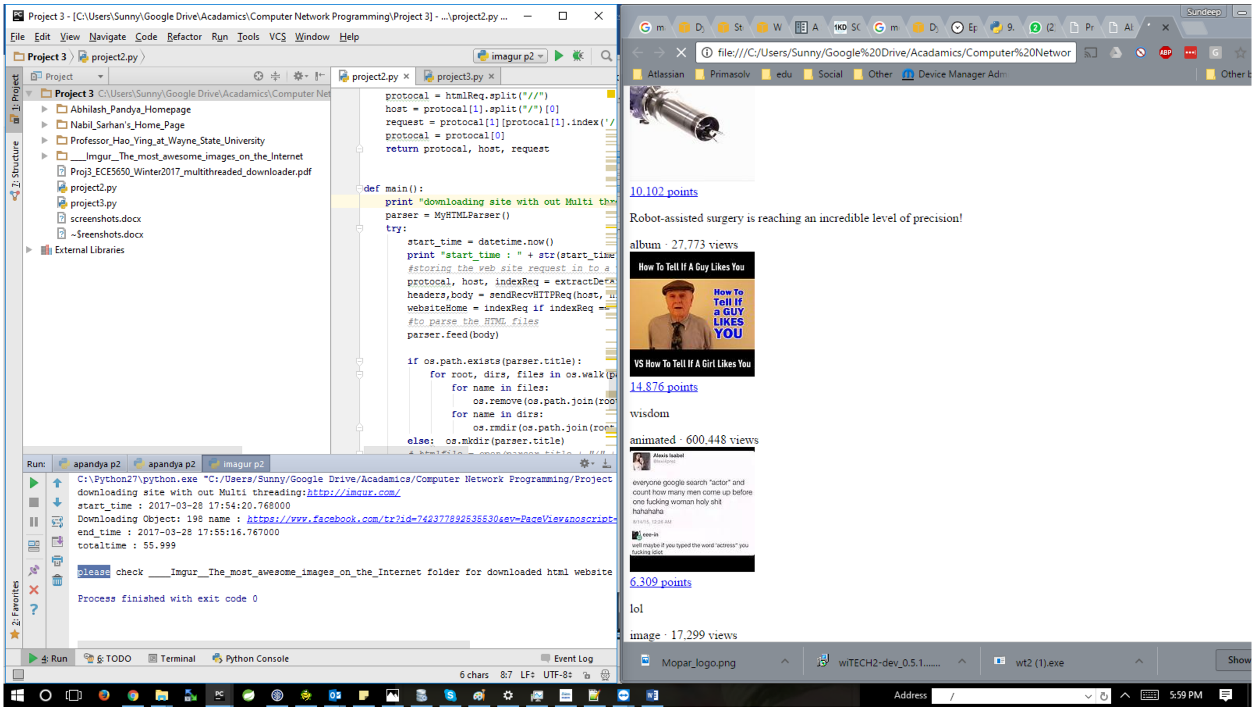
Min object size = 136 Kb

Items Downloaded:



```
_new_viral
_hot_viral
___Imgur__The_most_awesome_images_on_the_Internet.html
_hot_time
_t_A_Day_In_The_Life
_t_The_More_You_Know
_t_Current_Events
_t_Movies_and_TV
_t_Eat_What_You_Want
_t_Uplifting
_t_Storytime
_t_Creativity
_t_Funny
_t_Awesome
_t_Aww
_t_Gaming
_t_Science_and_Tech
_t_The_Great_Outdoors
_t_Inspiring
_t_Reaction
__imgur.com_memegen
__ajax.googleapis.com_ajax_libs_jquery_2.1.1_jquery.min.js
_gallery_O1ekY
_gallery_ZD5T6
_gallery_tavda
_gallery_6d00e
_gallery_WBsqV
_gallery_ABNfq
_gallery_jFW2i
_gallery_p9Rfm
_gallery_tZqt5
_gallery_OiyGB
_gallery_7Fmz7
_gallery_hghVX_comment_995568829
_gallery_hghVX
_gallery_bEL3K
_user_KatRaccoon
_gallery_OXbrvoB
_gallery_LaXjQ
_gallery_EDZHLIq
_gallery_JUXAM
_gallery_EkrxO
_gallery_dXmsF
_gallery_bKRPH
_gallery_QTcri
_gallery_KhVpmnZ
_gallery_H5uiTrJ
_gallery_Bulv1
_gallery_avXGZtR
_gallery_KHOxS
_gallery_vamgJ
_gallery_f9FNP
_gallery_mwv5U
_gallery_jue1t
_gallery_hPxkE_comment_995324741
_gallery_pCOK0
_gallery_nUqzLmR
_gallery_McOYcGT_comment_995649473
_gallery_McOYcGT_comment_995669777
_gallery_68Lhmzt
_gallery_QNeda_comment_995217129
_gallery_aYSyslj
_gallery_FELKi
_gallery_17gMz
_gallery_McOYcGT
_gallery_HJdDA
_gallery_g8HXkcL
_gallery_LNaH4lq
_gallery_rnpcu7D
_gallery_2aerJ
_gallery_DjXmL
_gallery_AIMAXy7
_gallery_wq8OW
_gallery_PG5xg
_gallery_qdmmyQ0
_gallery_LHsf1
_gallery_WHRRS
_gallery_VMNy88O
_gallery_wlrCM_comment_995623877
_gallery_WMW9U_comment_995284021
_gallery_gpY32
_gallery_8ahub
_gallery_YI5sl
_gallery_2QG66
_gallery_4Pz5W
_gallery_wlrCM
_gallery_9O32gXi
_gallery_91jwfSo
_gallery_8zj3YEn
_gallery_dy88doF
__imgur.com_tos
_user_AwesomeName
__imgur.com_apps
__imgur.com_privacy#adchoices
__imgur.com_privacy
__imgur.com_jobs
_user_CrowbarOmega
_user_packetoftimtams
_user_IntraiVillain
_user_tha1nonly0623
_user_lucillemaximus
__imgur.com_about
__imgur.com_vidgif
__s.imgur.com_images_favicon-152.png
__s.imgur.com_images_favicon-96x96.png
__s.imgur.com_images_favicon-32x32.png
__s.imgur.com_images_favicon-16x16.png
__i.imgur.com_8zj3YEnb.jpg
__i.imgur.com_rYxmNuHb.jpg
__i.imgur.com_IKti3LUb.jpg
__i.imgur.com_ZgS3j4Hb.jpg
__i.imgur.com_M5EpPbWb.jpg
__i.imgur.com_W07WkQnb.jpg
__i.imgur.com_hG7Dvjib.jpg
__i.imgur.com_NSPPRBab.jpg
__i.imgur.com_sry7JRqb.jpg
__i.imgur.com_2zuArz2b.jpg
__s.imgur.com_images_loaders_ddddd1_181817_48.gif
__i.imgur.com_C5d8xhEb.jpg
__i.imgur.com_LEfcAnab.jpg
__i.imgur.com_am1dMl8b.jpg
__i.imgur.com_oQTnyXFb.jpg
__i.imgur.com_cMien9mb.jpg
__i.imgur.com_OXbrvoBb.jpg
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
__i.imgur.com
```

Project 2 output



```
        protocal = htmlReq.split("//")
        host = protocal[1].split("/")[0]
        request = protocal[1][protocal[1].index('/
        protocal = protocal[0]
        return protocal, host, request


def main():
        print "downloading site with out Multi thr
        parser = MyHTMLParser()
        try:
            start_time = datetime.now()
            print "start_time : " + str(start_time
            #storing the web site request in to a
            protocal, host, indexReq = extractDeta
            headers,body = sendRecvHTTPReq(host, i
            websiteHome = indexReq if indexReq ==
            #to parse the HTML files
            parser.feed(body)

            if os.path.exists(parser.title):
                for root, dirs, files in os.walk(p
                    for name in files:
                        os.remove(os.path.join(roo
                    for name in dirs:
                        os.rmdir(os.path.join(roo
            else:  os.mkdir(parser.title)
```

```
C:\Python27\python.exe "C:/Users/Sunny/Google Drive/Acadamics/Computer Network Programming/Project
downloading site with out Multi threading:http://imgur.com/
start_time : 2017-03-28 17:54:20.768000
Downloading Object: 198 name : https://www.facebook.com/tr?id=742377892535530&ev=PageView&noscript=
end_time : 2017-03-28 17:55:16.767000
totaltime : 55.999

please check ____Imgur__The_most_awesome_images_on_the_Internet folder for downloaded html website

Process finished with exit code 0
```

10,102 points

Robot-assisted surgery is reaching an incredible level of precision!

album · 27,773 views

14,876 points

wisdom

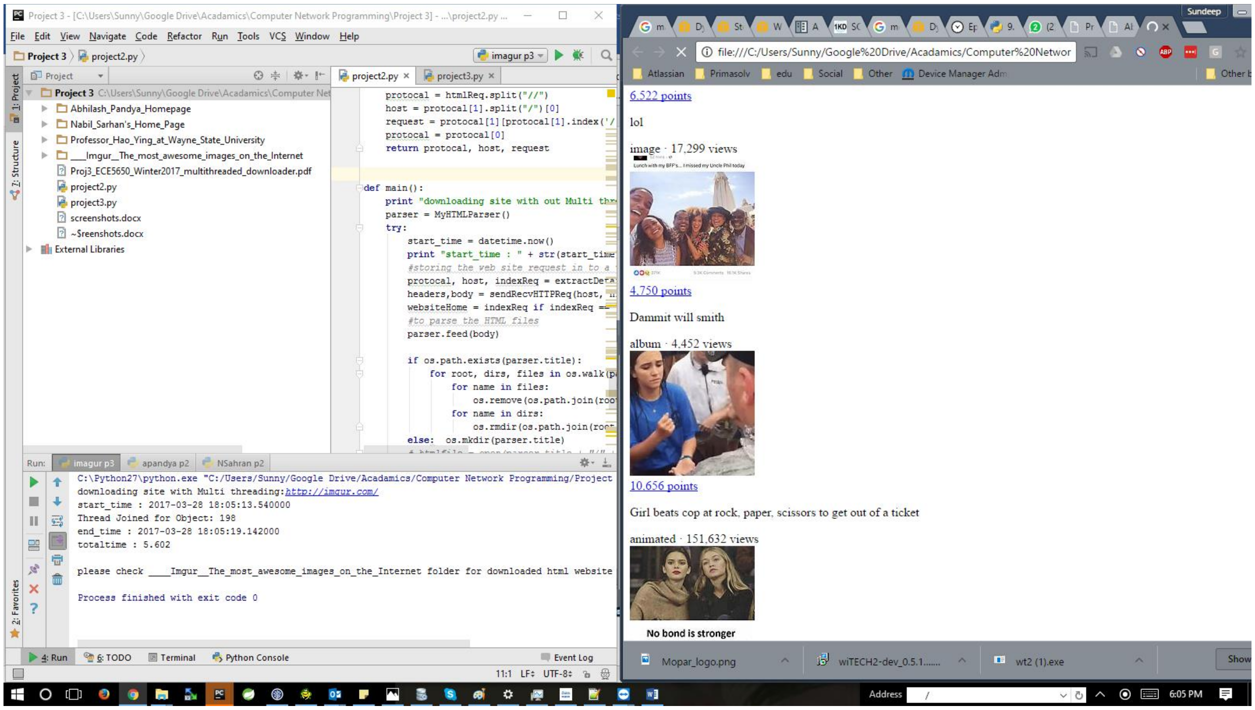
animated · 600,448 views

6,309 points

lol

image · 17,299 views

Project 3 output


Project 3 - [C:\Users\Sunny\Google Drive\Acadamics\Computer Network Programming\Project 3] - ...\project2.py ...
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Project 3
project2.py
imagur p3
Project
Project 3 C:\Users\Sunny\Google Drive\Acadamics\Computer Net
Abhilash_Pandya_Homepage
Nabil_Sarhan's_Home_Page
Professor_Hao_Ying_at_Wayne_State_University
___Imgur__The_most_awesome_images_on_the_Internet
Proj3_ECE5650_Winter2017_multithreaded_downloader.pdf
project2.py
project3.py
screenshots.docx
~Sreenshots.docx
External Libraries
project2.py
project3.py
Run:
imagur p3
apandya p2
NSahran p2
4: Run
6: TODO
Terminal
Python Console
Event Log
11:1 LF UTF-8
file:///C:/Users/Sunny/Google%20Drive/Acadamics/Computer%20Networ
Atlassian
Primasolv
edu
Social
Other
Device Manager Adm
Other b
Sundeep
No bond is stronger
Mopar_logo.png
wiTECH2-dev_0.5.1......
wt2 (1).exe
Show
Address
6:05 PM

```
        protocal = htmlReq.split("//")
        host = protocal[1].split("/")[0]
        request = protocal[1][protocal[1].index('/
        protocal = protocal[0]
        return protocal, host, request


    def main():
        print "downloading site with out Multi thr
        parser = MyHTMLParser()
        try:
            start_time = datetime.now()
            print "start_time : " + str(start_time
            #storing the web site request in to a
            protocal, host, indexReq = extractDeta
            headers,body = sendRecvHTTPReq(host,
            websiteHome = indexReq if indexReq ==
            #to parse the HTML files
            parser.feed(body)

            if os.path.exists(parser.title):
                for root, dirs, files in os.walk(p
                    for name in files:
                        os.remove(os.path.join(roo
                    for name in dirs:
                        os.rmdir(os.path.join(roo
            else:  os.mkdir(parser.title)
```

```
C:\Python27\python.exe "C:/Users/Sunny/Google Drive/Acadamics/Computer Network Programming/Project
downloading site with Multi threading:http://imgur.com/
start_time : 2017-03-28 18:05:13.540000
Thread Joined for Object: 198
end_time : 2017-03-28 18:05:19.142000
totaltime : 5.602

please check ____Imgur__The_most_awesome_images_on_the_Internet folder for downloaded html website

Process finished with exit code 0
```

6,522 points

lol

image · 17,299 views

4,750 points

Dammit will smith

album · 4,452 views

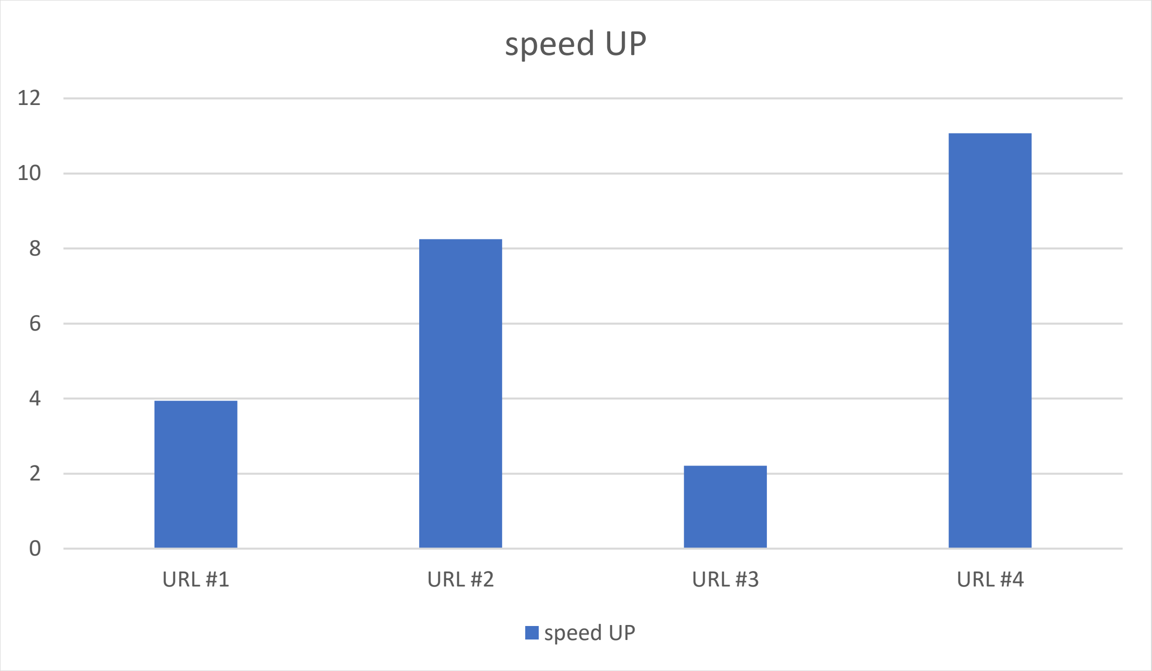
10,656 points

Girl beats cop at rock, paper, scissors to get out of a ticket

animated · 151,632 views

Analysis

URL #	No. of objects / No. of threads	downlaod time with out multi-threading(sec)	download time with multi-threading(sec)	scaleup
1	17	1.86	7.334	3.943
2	17	8.739	1.065	8.252
3	29	119.733	53.981	2.21
4	200	55.99	5.062	11.068



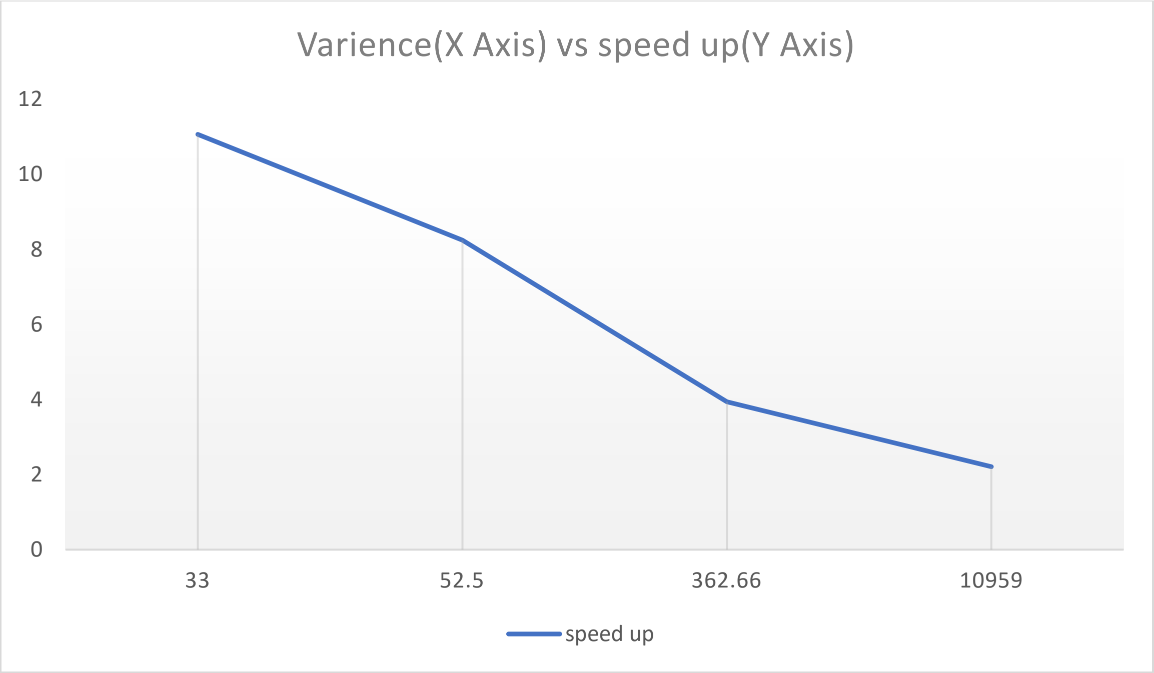
As, the webpages are different we get different speedups, we cannot derive a relationship between speedups, but we can try to understand why the speed ups differ.

We know that we can achieve best speedups for a parallel program when the task sizes are equal. Here we will try to compare the variance in task sizes with speedups i.e. the we will compare the variance in sizes objects downloaded with speedups

Variance will be given by

(max object size - min object size)/ min object size

URL #	Max object size (KB)	Min object size(KB)	variance	Speedup obtained in above table
1	1091	3	362.66	3.943
2	107	2	52.5	8.252
3	10960	1	10959	2.21
4	136	4	33	11.068



As we can see that the speed up decreases as the variance in object size increases.

Conclusion:

We have developed a multi-threading program to download web pages and calculated the performance, how and why different speedups are achieve due to the variance in object sizes