CS 421 Computer Networks

Programming Assignment 1 Report

Berdan Akyürek 21600904

Asım Güneş Üstünalp 21602271

Section 2

In the assignment, FileDownloader program is implemented in Python3.9.6 and tested. During the tests, no errors were found.

The program uses sockets to send HTTP requests and get responses. A function called send_request creates an HTTP request using the parameters URL, type(GET/HEAD), and lower and upper range values. Also, the same function sends this request to the website and returns the response after receiving it. The implementation is as follows:

```
def send_request(url, typ, lower = -1, upper = -1):
    host = url.split("/")[0]

    socket_to_index = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
    socket_to_index.connect((host, 80))

    index_path = url.split("/", 1)[1]

    typstr = "GET"

    if not typ:
        typstr = "HEAD"
    q = "%s /%s HTTP/1.1\r\nHost:%s\r\n" % (typstr, index_path, host)

    if lower != -1 and upper != -1:
        q += "Range: bytes=%d-%d\r\n" % (lower, upper)

    q += "\r\n"

    socket_to_index.send(q.encode())

    respon = socket_to_index.recv(4096).decode("utf-8")
    return respon
```

Other functions named head and download call this function to send a HEAD or GET request to the server. HEAD request is used in order to validate the files before downloading and GET requests are used to download the content of the webpage. Also in order to download files partially, a Range header is added to some GET requests. Some GET and HEAD request and response examples are as follows:

```
GET request when no range value is given:
GET /~cs421/fall21/project1/files/decrypted_file_1.txt HTTP/1.1
Host:www.cs.bilkent.edu.tr

GET request when some range values are given:
GET /100/captmidn.txt HTTP/1.1
Host:www.textfiles.com
Range: bytes=0-999

HEAD request in order to check validity before download:
HEAD /games/arcana.txt HTTP/1.1
Host:www.textfiles.com

HEAD response of an existing file:
HTTP/1.1 200 OK
Date: Thu, 11 Nov 2021 16:55:56 GMT
Server: Apache/2.4.37 (FreeBSD) OpenSSL/1.0.2o-freebsd
Last-Modified: Thu, 15 Feb 2001 00:17:07 GMT
ETag: "5b7f-37d4d6135c6c0"
Accept-Ranges: bytes
Content-Length: 23423

HEAD response of a non-existing file:
HTTP/1.1 404 Not Found
Date: Thu, 11 Nov 2021 16:55:54 GMT
Server: Apache/2.4.25 (FreeBSD) OpenSSL/1.0.2u-freebsd PHP/7.4.15
Content-Type: text/html; charset=iso-8859-1

Header of GET response of a file that has a lower bound smaller than
file length:
HTTP/1.1 206 Partial Content
Date: Thu, 11 Nov 2021 16:57:48 GMT
Server: Apache/2.4.37 (FreeBSD) OpenSSL/1.0.2o-freebsd
Last-Modified: Sun, 01 Aug 1999 17:20:56 GMT
ETag: "5f4a-35109f02a6e00"
Accept-Ranges: bytes
Content-Length: 1000
Content-Range: bytes 0-999/24394
Content-Type: text/plain
```

GET responses are parsed using the blank line between header part and the content part. This way, the HTTP header is not included in the saved files.

According to HEAD responses, answers are validated. If the returned head is not 200 OK, it means the file does not exist, so a GET request did not sent. Otherwise, file length is checked using . If the file length is smaller than the lower bound of the range, a GET request did not sent and the file is not downloaded. Because, if the lower bound is n, it is not possible to download the nth byte of a file if it contains less than n bytes. Otherwise, if there is no bounds or lower bound is smaller then file length, file downloaded with GET request.

The program first gets the index file and if file exists, iterates over all URLs, sends requests and receives responses using functions explained and does required operations.