National University of Singapore
School of Computing

CS2105 **Assignment 0** (2 Marks) Sem 2 AY22/23

Submission Deadline

27th Jan 2023 (Friday) **23:59**. 1 point penalty will be imposed on late submission (Late submission refers to submission or re-submission after the deadline). The submission folder will be closed on 3rd Feb 2023 (Friday) **23:59** and no late submission will be accepted afterwards.

Objectives

- This is a warm up assignment to familiarize you with programming skills that will be useful for later assignments.
- This programming assignment is worth 2 marks. All the work in this assignment shall be completed **individually**.

Grading

- We accept submission only of Python 3 (3.8.10 in particular) programs.
- Programming languages other than Python 3 are not allowed.
- We will test and grade your programs on the `stu` server. Please make sure that your programs run properly on `stu` and not only on your own system.
- By default, we use the `python3` program installed in folder `/usr/bin/python3` on `stu` for grading. Unless stated otherwise for individual tasks, you are allowed to use libraries installed in public folders of `stu` (e.g. `/usr/lib`) only.
- Your programs will be graded automatically using scripts.
  - Please make sure that your programs behave exactly the same as described in this document, because the grading scripts are unable to award partial marks.
- A set of testing scripts is released to you in the assignment package.
  - These test scripts cover common cases that your programs are expected to handle.
  - During actual grading, we will use extra cases to test your programs. Hence, passing all the released test cases does not guarantee that you will get full marks.
  - In addition, we will detect fraudulent cases such as hard-coding answers within programs and deduct marks accordingly.

We will deduct 1 mark for failure to follow instructions.

Common Issues

- Incorrect File name causing evaluation script failure: Make sure that your programs have correct names.
- Incorrect output format: Make sure that every line of your output is properly ended with a line break (i.e. "\n" character). In particular, your program should end the output with a line break, unless otherwise stated.
- Output mismatch: Do not output irrelevant messages that are not shown in sample runs (for example, debugging messages).

Accessing the `stu` Server

To test your programs, please use your SoC UNIX ID and password to log on to `stu` first:

- If you don't have your SoC UNIX account, please create it here:

  https://mysoc.nus.edu.sg/~newacct
- If you forget your SoC password, please reset it here using your NUSNET ID and password:

  https://mysoc.nus.edu.sg/~myacct/resetpass.cgi
- If you are using a UNIX-like system (e.g. Linux, Mac, or Cygwin on Windows), SSH should be available from the command line and you may simply type:

  ```
  ssh <SoC-UNIX-ID>@stu.comp.nus.edu.sg
  ```
- To copy files to `stu` on command line, use `scp`. The usage is as following:

  ```
  scp -r <source-folder> <SoC-UNIX-ID>@stu.comp.nus.edu.sg:<target-folder>
  ```
- If you are using a Windows machine, you may need to install an SSH client (e.g. "SSH Secure Shell Client"

Testing Your Programs

- Upload your programs along with the `test` folder from the package to `stu`.
- Make sure that your programs and the `test` folder are in the same directory.
- Run `bash test/<Exercise Name>.sh` for testing.
  - For example, to test your program for Exercise 1, run the following command:

    ```
    bash test/Checksum.sh
    ```
  - By default, the script runs through all test cases. You can also choose to run a certain test case by specifying the case number in the command:

    ```
    bash test/Checksum.sh 3
    ```
- To stop a test, press and optionally hold `Ctrl-c` if pressing once does not exit the test.

- You may assume that all input data to your programs are valid. Hence, there is no need to perform input data validation in your programs.

If you have any question or encounter any problem with the steps above, please post your questions on `piazza` or consult the teaching team.

Program Submission

- Please create a zip file containing your source files (only) and submit it to `Assignment_0` canvas.
- The file name should be `<Matric Number>.zip` where `<Matric Number>` is your matriculation number which starts with letter `A`. An example file name would be `A0165432X.zip`.
- All file names, including the zip file and all source files within the zip, are **case-sensitive**.
- In addition, your zip file should not contain any folders or subfolders or irrelevant files such as test.jpg.
- Although we always try to find your actual source files during grading. You are encouraged to submit only once, or otherwise only your latest submission will be graded.

You are not allowed to post your solutions to any publicly accessible site on the Internet.

Plagiarism Warning

You are free to discuss this assignment with your friends. But, ultimately, you should write your own program. We employ zero-tolerance policy against plagiarism. If a suspicious case is found, student would be asked to explain his/her code to the evaluator in face. Confirmed breach may result in zero mark for the assignment and further disciplinary action from the school.

Question & Answer

If you have any doubts on this assignment, please post your questions on `piazza`. We are not supposed to debug programs for you, and we provide support for programming queries on a best-effort basis only. The intention of Q&A is to help clarify misconceptions and give you necessary directions.

Exercise 1 – Checksum (1 mark)

Checksum can be used to detect if data is corrupted during network transmission (e.g. a bit flips from 0 to 1). Write a program `Checksum.py` to calculate the CRC-32 checksum for a file `<src>` entered as command-line argument. File `<src>` should be placed in the same folder as `Checksum.py`.

You may use the `crc32()` function from the `zlib` library to calculate the CRC-32 checksum (see https://docs.python.org/3/library/zlib.html#zlib.crc32 for details. Remember to import this library before use!). Firstly, you will need to read all the bytes from a file and store them into a `bytes` object. You should open the file with **b**inary **r**eading mode by using **"rb"** as the mode argument of `open()`. Then, you can call the `read()` method of the file object and get the entire file content in a `bytes` object. Now you can get the checksum by directly calling `crc32()` with the file data, and finally print this unsigned 32-bit checksum.

An example code snippet is given below.

```
with open("test.jpg", "rb") as f:
  bytes = f.read()
checksum = zlib.crc32(bytes)
```

Note: in practice, you may encounter an out-of-memory error if you attempt to load a huge file entirely into the memory. In this exercise, however, we limit the maximum file size to 10MB so that you can do so safely without triggering such errors.

Sample run:

```
$ python3 Checksum.py test/test.jpg
3237218320
```

Exercise 2 – PacketExtr (1 mark)

In this exercise, you are going to write a program, `PacketExtr.py`, to read consecutive "packets" from the `stdin` stream, extract their data payloads in a *responsive* manner and output to `stdout`.

We define a custom format of packets for this exercise. A packet consists of a text-based header and a binary data payload following the header. The header is a string formatted as "`Size:␣<size>B`", where ␣ represents a white-space character, and `<size>` is a decimal integer representing the number of bytes of the binary data following the header. The header is ended with the "`B`" character, and the payload immediately follows the header without any byte (such as "\n") in between. For example, "`Size:␣2105B`" is a complete and valid header.

The binary payload can contain any byte, not limited to printable characters. Therefore, the payload should not be treated as string data in your program. This packet format clearly defines the boundary between header and binary data within a packet, and also the boundary between consecutive packets. It ensures that all information can be

parsed correctly over a data stream.

Packets are fed to your program sequentially through `stdin`, until End-Of-File (EOF) is encountered. It is guaranteed that all packets have correct formats and correct payload sizes. Your program should be **responsive** to the input in the sense that upon receiving a full packet, it outputs the payload of the packet to `stdout` without any extra characters including newline. Unlike Exercise 1, the program cannot read all data once and process them in batch.

It is recommended that your program reads and writes data in binary mode instead of the default text mode. For interactive I/O, `read1()` should be used instead of the ordinary `read()`, and `flush()` should be called after each `write()` to avoid delays caused by buffering. Both `read()` and `read1()` accept one argument which is the maximum number of bytes to read. The main difference is that, `read()` returns only when the specified number of bytes are read or End Of File (EOF) is encountered, while `read1()` returns immediately upon new data stream in and may return fewer bytes than specified. You can learn more about the details at https://docs.python.org/3/library/io.html#io.BufferedReader and https://docs.python.org/3/library/io.html#io.BufferedWriter. The following code shows how to do binary I/O on `stdin` and `stdout`:

```
import sys
# read **at most** 5 bytes from stdin
data = sys.stdin.buffer.read1(5)
# write data to stdout and flush immediately
sys.stdout.buffer.write(data)
sys.stdout.buffer.flush()
```

Here, the `data` object is of `bytes` class. Python 3 programs can operate on binary data using `bytes` objects. This class is very similar to `str`, the string class. For example, both classes have functions `find()`, `split()`, and also slice operators for range access (e.g. `a[0:10]`). Details about this class can be found at https://docs.python.org/3/library/stdtypes.html#bytes-and-bytearray-operations.

The following shows how to manipulate `bytes` objects:

```
# prepend b to the '...' expression to form a bytes object
# instead of str
pos = data.find(b'x')
if pos >= 0:
  # if byte 'x' is found in data
  part1 = data[0:pos+1]  #  this is similar to str slicing
  part2 = data[pos+1:]  # slice until the end of data
```

To convert a `bytes` object to `str`, your program can call the `decode()` method of `bytes`. In this exercise, there is no need to worry about text encoding, as we only use basic ASCII characters in headers.

Finally, to detect EOF on `stdin`, your program can check the length of the `bytes` object read from `stdin`. If your program expects to read more data from `stdin` but receives a zero-length `bytes` object, it means that there is no more data on `stdin` and EOF is encountered.

When testing your program on command line, you can feed the contents of a file to `stdin` of your program using file redirection (<), instead of typing into the terminal. You can also use another type of redirection (>) to save the output of your program to a file rather than let it print to the terminal. For example, the following line feeds the file `input.data` to the program and save its output to `output.data`:

```
python3 PacketExtr.py < input.data > output.data
```

You can then compare binary contents of `output.data` and the given reference output by running:

```
cmp output.data ref-output.data
```

By default, `cmp` outputs all differences found between the two files. Hence, no output means that the two files are identical.

Note that above command line testing can only test your program's correctness. During grading, in addition to testing correctness, we will test responsiveness by setting a **one-second timeout** after feeding your program a full packet. That is, if your program does not output the packet payload on time, we will deem your program as unresponsive. This timeout is sufficient for our test data sizes. We reiterate that your program should not do batch processing, as interactive processing is one of the key points of this exercise. We limit the size of each packet to 1MB and the size of all packets to 10MB.

Sample run:

```
$ python3 PacketExtr.py < test/packets-a.in > run-a.out
$ cmp test/packets-a.out run-a.out
```