AWK

Q1.
Given a directory which contains sub-directories and C files, write an AWK script along with a bash script to print the total number of lines of comments and total number of strings in the whole directory. The bash script should get all .c files recursively and the AWK script should parse them. You can assume all the C files are error-free and pre-processed.

Ex:-

```
dir\
  |-inc\
    |-solution.h
  |-src\
    |-main.c
    |-solution.c
```

solution.c

```
// This is a single line comment
/* This is a multi-line comment */
/* This is another
   multi-line “comment”
   considered as 4 lines
*/
#include <solution.h>
int solution(void)
{
    char *s = "try combinations of strings and /*comments*/ too";
    return 0;
}
```

main.c

```
#include <stdio.h>
int main(void)
{
    // The following output is the actual solution !!!
    printf("7 lines of comments\n");
    printf("3 quoted strings\n");

    return 0;
}
```

Explanation:-
The file *solution.c* contains 1 string and 6 lines of comments. The *main.c* file contains 1 comment and 2 strings and thus the output for the whole directory will be 3 strings and 7 lines of comments.

Q2.

Write an AWK script which accepts packet capture log as input. The packet capture log contains information about packets transferred across the network. In this assignment, we are interested to find out packets transmitted through TCP connection. A TCP connection is identified by a four tuple <source IP, source port, destination IP and destination port>. The data is transferred in both directions and identified by the sequence number. Consider the following example packet capture log.

00:58:45.452688 IP 74.125.130.189.443 > 172.25.92.26.48288: Flags [P.], seq 1095791942:1095792002, ack 2430684857, win 258, options [nop,nop,TS val 3399350415 ecr 26537982], length 60

In this packet, the relevant portions for this assignment are given below.

timestamp=00:58:45.452688
Sender IP = 74.125.130.189, Sender Port = 443
Receiver IP = 172.25.92.26, Receiver PORT = 48288

Data sequence start = 1095791942, end = 1095792002, where every data byte is assigned a unique sequence number.

For a given packet capture log file print the following output for each connections.

Connection (A = IP:PORT, B=IP:PORT)

A-->B
#of packets (all) #of data packets #Data (bytes) #Retransmitted (bytes) Throughput (bytes/sec)
B-->A
#of packets (all) #of data packets #Data (bytes) #Retransmitted (bytes) Throughput (bytes/sec)

Note:

1. All packets transferred may not carry data (some of them are acknowledgements).
2. Retransmission events can be captured by comparing sequence numbers of packet with previously transmitted packets.
3. The *throughput* is the unique number of bytes delivered per second (bytes/sec), known as "goodput" in Computer Networks.
4. Attached archive contains packet capture logs (in the above format in .txt files) and the raw packets (files with .pcap extension) which can be opened using Wireshark (GUI) or tcpdump (command line).