How to start COMP30023 Assignment 1

Lachlan Andrew

March 24, 2020

1 Why is this being written?

Several people have said that they do not know how to start Assignment 1. Although not everything required for this project has been taught yet, there is much that you can do already.

2 General guidelines

For any project, the first thing to consider is not how to do the low-level tasks, like creating a socket, but what the big picture is like.

- What tasks have to be performed? These will form functions or modules.
- What is the sequence of these tasks? That is, what is the overall algorithm?
- What data structures are required to support the algorithm? What abstract services will be required (random access? sequential access? insert? delete?), and what concrete data structures provide these? It is often useful to start with a simple data structure that may be inefficient, and only upgrade to a sophisticated one if it is shown to be a performance bottleneck. One such starting point is storing everything in an array, and using exhaustive search for retrieval. But how many arrays will there be, and what has to be stored in each?

Examples of tasks that you may have are

- Fetch a page
- Parse an HTML document
- Identify duplicate URLs

You can think of many more.

Each of these taks can be broken down further. For example, "Fetch a page" can be broken into "Construct a query", "Send the query to the server" and other tasks.

If you don't yet know how to send a query to the server, you can write a "stub" for "Fetch a page". It could simply read a page from a text file. This allows you to write and the overall structure before you know how to contact the server.

3 Specific suggestions

Read.

You know that you will be using sockets. The lecture notes tell you that this requires the function socket(). The first step in writing this code could then be: "man socket". The lecture notes also tell you that the side initiating a connection needs the function connect(). Guess what the second step is to understand how to create a socket. (Hint: it starts with "man".)

The spec also said that "real programmers use stackexchange". The first hit on DuckDuckGo for the query: "how to open a socket in C" is `https://www.geeksforgeeks.org/socket-programming-cc` which gives example code to open a socket, both for the server and the client. You can go through that example, delete lines that seem unnecessary, and modify lines to suit your particular case. Try to cut it down to the smallest set of lines that still work. Look up the manual to find the meaning of each of the constants, like SOCK_STREAM.

There is also a document kindly written by Colton Carner with hints on PCRE.

4 Conclusion

Nothing in this document should be a surprise. If you feel that you don't know how to start, ask yourself if you really mean that you don't know how to finish. Programming involves trying things, making mistakes, and rewriting. This is a programming assignment, not simply a coding assignment or a lab class. It is challenging, but it is only by challenging yourself that you learn.

You know several ways you can start the project. Hop in and enjoy the adventure, despite not knowing where you will end.