Meeting 6 Summary

Discussed what I have put in my January Deliverable document to make sure that nothing is left out:

* Introduction
* Current progress – more in terms of a narrative for what’s been done so far with some technical details rather than providing evidence for the work. Need to add a section for the architecture of the project – talk about how components interact and how to organise components to get the best out of resources
* Literature review – mostly completed but may have more additions as the project progresses
* Remainder of work to do

Talked about references:

* All points made need to be cited
* Citations need to be consistent – ie same structure each time
* No need to cite general knowledge but be careful

Considered ideas for testing modules:

* Measure the performance increase/decrease when running AI libraries with the web crawler compared to with the browser – main idea is that its better to do as much of the computation as possible in areas where the user isn’t interacting with the system so that the user experience is as fast as possible. But don’t want crawler to be incredibly slow, and not all computation can be done there because current user information is often needed
* Unit tests for crawler – can show that the crawler works and explain its methodology but will be beneficial to create documentation to test the individual components/how it works with different parameters
* Can also help to show that code is clean and well structured
* Will be good to ensure program works even if AI libraries are excluded in case of some error with import or change in specification from the library.

Talked about making sure the crawler is ethical and doesn’t violate any privacy or access laws

Talked about the scalability of the system:

* How will it work with more than one user – will data overlap or is it better to keep separate. Obviously easy with each user having an account in the system but what if crawlers ran at the same time or lots of users used the system without being signed in
* This could lead into analysing group behaviour as well to profile what a certain demographic are likely to want to search at the current time
* Need to test how the program responds when lots of users are interacting with the system – isn’t a big issue for the purposes of the project but good to know and add to the further work section
* Similarly, need to test for when the website database is very large to make sure that retrieval times are still acceptable and doesn’t cause too large a delay