Lab6 Meeting Notes

Day 1 (2/18/2025)
• Discussed key points of the project.
• Split and divide project tasks.
• Finished the initial setups.

Day 2 (2/19/2025)
• Shared progress of each member's tasks.
• Learned and discussed the tools needed to convert PDF to text.
• Try to convert a PDF file and check the conversion effect. Find that a lot of information cannot be converted. Discuss how to solve this problem.

Day 3 (2/20/2025)
• Shared progress of each member's tasks.
• Process the text extracted from PDF, remove irrelevant information, and unify the information format.
• Finished the PDF extraction part.

Day 4 (2/21/2025)
• Shared progress of each member's tasks.
• Observed the structure of the website where we need to crawl data and tried to write a crawler.
• Observed the crawled data and processed them

Day 5 (2/22/2025)
• Shared progress of each member's tasks.
• Combine the 2 scripts and run them together, then debug them.
• Completed the writing of various documents and the uploading of codes.