Software Design and Engineering

Lab Document

High Level Purpose Statement:	My goal is to get some exposure to Maven and understand why it is so widely applied in the software engineering industry. I believe this is a good opportunity to explore Java libraries.
Experimental Design:	I am foreseeing this project to have four major phases: 1) Maven installation / configuration 2) Researching Java libraries and making selections 3) Creating the project 4) Testing the project Ideally, I would like to find something to get some exposure to web scraping.
Resources Available:	**Maven Installation:** https://www.youtube.com/watch?v=Xatr8AZLOsE&t=321s **Jsoup Tutorial:** https://www.youtube.com/watch?v=riZ2GAaMDGM **Forum that helped me solve a 4-hour frustration of Maven not working for IntelliJ:** https://intellij-support.jetbrains.com/hc/en-us/community/posts/11333307374994-Maven-Dependency-not-recognized-by-IntelliJ **Web Crawler Tutorial (from when I was still experimenting):** https://www.youtube.com/watch?v=wrFXBV4MwvI **JUnit Tutorial:** https://www.petrikainulainen.net/programming/testing/junit-5-tutorial-running-unit-tests-with-maven/ **My Go-To for Java Swing:** https://web.mit.edu/6.005/www/sp14/psets/ps4/java-6-tutorial/components.html
Time Estimate:	I am predicting to spend at least 10-12 hours on this project. (Haha, I tripled that)
Experiment Notes:	1) <u>Maven installation / configuration:</u> Wow, this was incredibly frustrating. I spent one night and part of an afternoon trying to get Maven to work with IntelliJ, along with the configuration of the dependencies. The solution was not very complicated, but I think the lack of sleep and building frustration/confusion of installing Maven was fogging my brain. I would say I spent about 4-5 hours total in this phase.

	2) Researching Java libraries and making selections: This phase also took a long time. I spent multiple hours brainstorming and experimenting with multiple ideas. The project that I wanted to do most was to create a job-scraping application. Unfortunately, after about 4 hours of failing to scrape data from Indeed, Glassdoor, and LinkedIn, I realized that it was impossible to create this project within this time frame. These websites acknowledged my web scraper as a "robot" and refused to provide access to their data. I discovered another dependency would have better luck, but I already spent so much time learning Jsoup that I decided to give up the idea. Instead, I am creating a simple weather scraping application. I also decided to use JUnit for testing my software. I know that JUnit is popular in industry, so I wanted to get exposure to it. 3) Creating the project: After about 24-26 hours, I finally completed the project. I started out with a very basic main function that performed scraping, data configuration, and data compilation. After gaining a proficient understanding of Jsoup, I decomposed the main function into multiple classes, aiming to utilize the Model-View-Controller pattern. I spent a lot of time designing the architecture of the program. Once the program and its GUI were functional, I spent a big chunk of the day creating weather icons. 4) Testing the project: This was not difficult. I created the WeatherScraperTest and configured Maven to run this test to ensure that no unchecked errors would occur.
Results:	Overall, I spent roughly 34+ hours on this project. I think that it turned out okay and has the potential to continue growing. Regardless of the outcome, I learned a lot of incredibly valuable things about Maven and web scraping, and I'm looking forward to learning more about them in the future.
Consequences for the Future:	I learned a lot about Maven and web scraping. This project was challenging in the beginning due to the lack of knowledge that I had with Maven and Jsoup. After spending a lot of time on it, my understanding improved significantly. After completing this project, I think that I have grown as a developer because I gained a lot of knowledge and applied that knowledge in such a short time. My interest in web scraping only grew after working on this project.