**Task 3 Documentation**

**1. Overview**

**During my internship, I worked on a project to automate the search and data gathering process for grants specifically aimed at startups in the pre-seed, ideation, or MVP stage in Jordan or internationally. This tool automates grant searches using Selenium in Python, saving the results in a text file along with the date of the search.**

**Key Responsibilities:**

* **Developed an automated search tool to find startup grants in Jordan or internationally.**
* **Implemented optional filters to limit search results to a specific website (e.g., grants.gov).**
* **Extracted relevant grant links from Google search results and saved them in a text file.**
* **Logged the date of each search for record-keeping and traceability.**

**2. Tools and Technologies Used**

* **Python: Used for writing automation scripts to perform web searches and gather information.**
* **Selenium: Browser automation tool used to simulate searches and retrieve results from Google.**
* **ChromeDriver: WebDriver used to control Google Chrome in headless mode for efficient search execution.**
* **BeautifulSoup: Used for parsing the HTML structure of the retrieved search pages.**
* **Jupyter Notebooks: To structure the code and experiment with various search and filter configurations.**

**3. Project Details**

**3.1 Grant Search Automation**

* **Objective: Automate the process of searching for startup grants on Google, filtering results by specific websites, and saving relevant information in a structured format.**
* **Technologies Used: Python, Selenium, ChromeDriver, BeautifulSoup.**

**Key Features:**

1. **Search Automation:**
   * **The tool automates Google searches for grants using keywords relevant to startups in their pre-seed/ideation/MVP stages in Jordan and internationally.**
   * **The user is prompted to specify if they want to restrict search results to a specific website (e.g., grants.gov or europa.eu).**
2. **Filtering and Customization:**
   * **Users can enter a website filter (e.g., site:grants.gov) to refine the search results to only show links from the specified domain.**
   * **The query is dynamically adjusted based on user input to prioritize results from the target domain.**
3. **Link Extraction and Information Logging:**
   * **The search results are processed, and relevant links (excluding unnecessary or internal links) are extracted.**
   * **Grant details such as the title and URL are gathered, and the results are stored in a text file.**
   * **Each search is logged with the date and time of the query for auditing purposes.**
4. **Headless Browser Execution:**
   * **The search and link extraction is done using a headless Chrome browser, which means the browser runs in the background without a visible window, optimizing performance.**

**4. Challenges and Solutions**

**4.1 Handling Web Pagination**

* **Challenge: Extracting data from multiple Google search result pages.**
* **Solution: The script was designed to handle pagination by clicking the "Next" button to continue extracting links from subsequent pages until no more pages are available.**

**4.2 Filtering Irrelevant Links**

* **Challenge: Filtering out links that are not relevant to the grant search (e.g., internal Google links or unrelated ads).**
* **Solution: Used filtering techniques to exclude links that did not match the desired criteria (e.g., non-grant-related links or unrelated domains).**

**5. Outcomes and Learnings**

**Outcomes:**

* **Automated the process of searching for grants in specific geographic locations (Jordan or international).**
* **Reduced manual search time by filtering results from specific websites.**
* **Created a log of search results and dates for future reference and reporting.**

**Learnings:**

* **Gained experience in browser automation and dynamic content extraction using Selenium.**
* **Improved skills in handling search result pagination and applying filtering techniques to optimize search results.**
* **Enhanced knowledge of Python scripting for data extraction and logging.**
* **Learned about headless browsing and its benefits for faster, more efficient data gathering.**