Software Design - CSC 312

Status Report 2

LinkedIn Profile Analyzer

Group 4

Delario Nance, Jr., Sky Luo, Donald Lin, Alp Niksarli

April 7, 2024

1 Introduction

1.1 Highlights

• What was the plan for this iteration?
- Development
    - Build a LinkedIn scraper that scrapes LinkedIn profiles or LinkedIn profile PDFs
    - Build a Job scraper that scrapes job postings on platforms like LinkedIn and Indeed
    - Build a Front-end dashboard
    - Build an NLP Analyzer that extracts keywords/skills from a LinkedIn profile and recommends skills to learn for a given job role
    - Build a Course scraper that scrapes courses on platforms like YouTube and Codeacademy
- Customers
    - Create user information release terms

• Highlight what the team accomplished.
- Scope
    - We changed our project scope from an SEO tool for websites to a LinkedIn profile optimization tool
- Development
    - We built a module for extracting text from a PDF, given the PDF's file path or Python object representation
    - We built a module for gathering LinkedIn data on job openings which match given keywords and locations
    - We built a module for asking GPT-4 (1) if a given job title matches keywords in a given text, and (2) what skills to learn so that the text better matches the job title
    - We built a module for gathering Coursera data on courses which teach a given list of skills
    - We integrated the 4 modules above into a Flask app, allowing the user to input their LinkedIn profile PDF and desired job role to receive:
        - Profile feedback
        - Recommend skills to learn, with links to related Coursera courses
        - Links to related LinkedIn job postings
- Customers
    - Interviewed a Philosophy student to analyze their current knowledge on their skills, what is represented in their LinkedIn profile and how it can be improved.
    - Learned that Coursera and YouTube are used as primary sources for skill acquisition, and geared the product towards these venues.
    - We want to interview a STEM oriented student as well in order to understand their perspective and goals so that we can shape the product accordingly

1.2 Changes

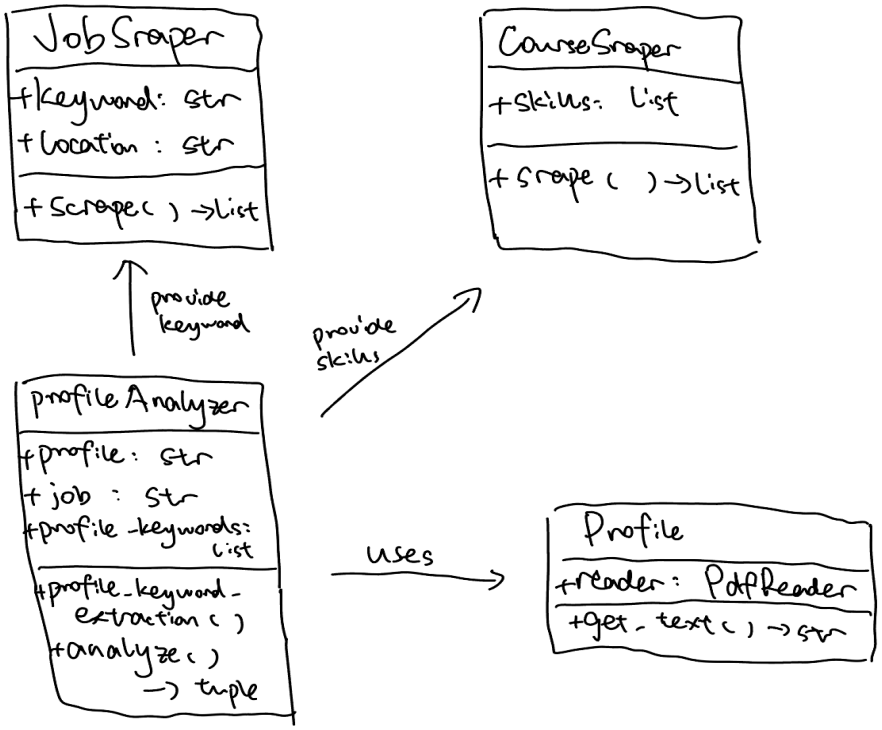
• Summarize any major changes since the proposal.
    - We decided to shift our project idea from an SEO tool for startups in the HurtHub to a LinkedIn profile optimization tool
• Include each change's date, motivation, description, and implications.
    - Change: Shift project idea to a LinkedIn profile optimization tool
        - Date: March 22
        - Motivation: When meeting with our customer Liz Brigham (Hurt Hub), she told us that our software lacked novelty because WordPress offers many SEO-related services which our software would offer
        - Description: Our new project idea involved building a tool which allows a user to input a PDF/hyperlink of their LinkedIn profile along with a desired job role to gain feedback on courses to take or skills to learn so that their profile is more aligned with the job role
        - Implication: We will likely not have enough time to shift our project idea again. Therefore, we should regularly make sure that our project scope is wide enough to offer a novel product but narrow enough so that a MVP will be completed by the final presentation
• If there were none, note that there were no changes.

2 UML: Class Diagram

Draw a class diagram that *only* includes the important classes your group mainly worked on during this iteration. Identify a single owner on the team for each class, even if multiple team members contribute.

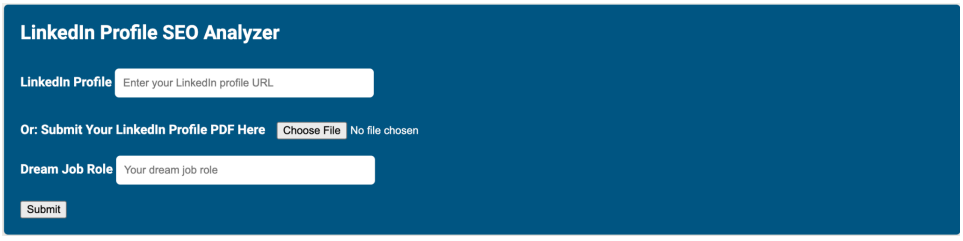
All of our attributes/methods are public, so we put a + sign before every attribute/method.

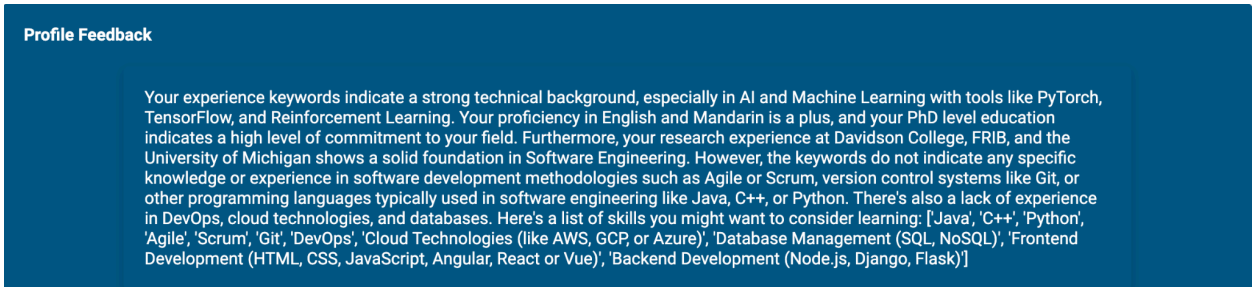
- Job scraper: Donald Lin
- Course scraper: Alp Niksarli
- Profile scraper: Delario Nance
- Profile Analyzer: Sky Luo

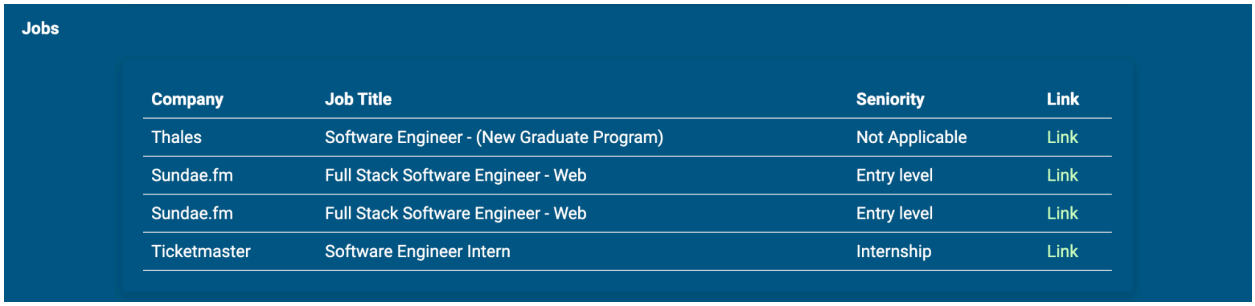


3 Current Status

- Add the screenshots of the parts that are working

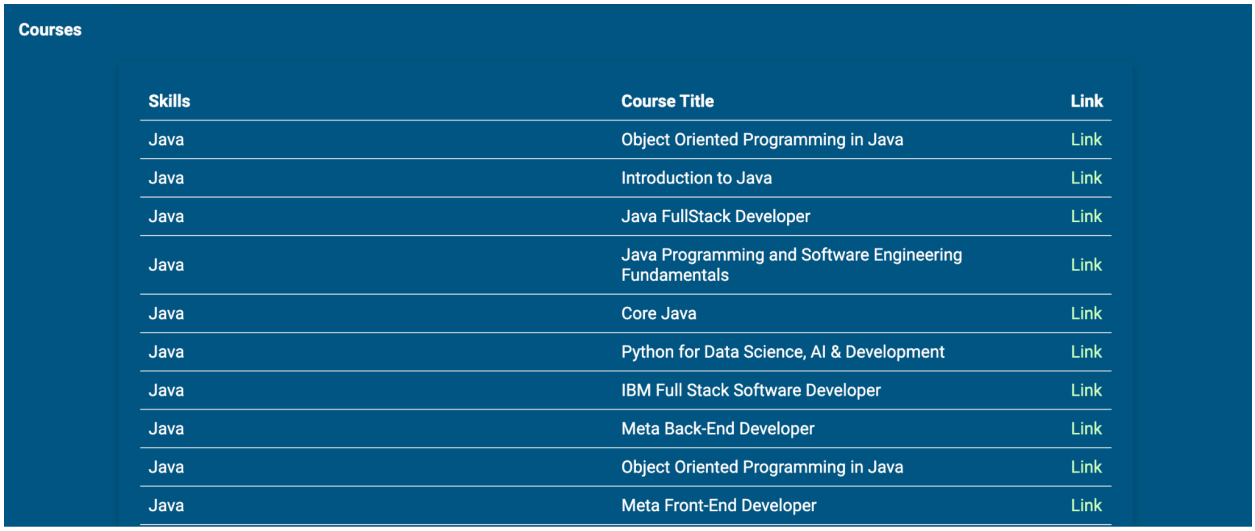






Jobs

Company	Job Title	Seniority	Link
Thales	Software Engineer - (New Graduate Program)	Not Applicable	Link
Sundae.fm	Full Stack Software Engineer - Web	Entry level	Link
Sundae.fm	Full Stack Software Engineer - Web	Entry level	Link
Ticketmaster	Software Engineer Intern	Internship	Link



Courses

Skills	Course Title	Link
Java	Object Oriented Programming in Java	Link
Java	Introduction to Java	Link
Java	Java FullStack Developer	Link
Java	Java Programming and Software Engineering Fundamentals	Link
Java	Core Java	Link
Java	Python for Data Science, AI & Development	Link
Java	IBM Full Stack Software Developer	Link
Java	Meta Back-End Developer	Link
Java	Object Oriented Programming in Java	Link
Java	Meta Front-End Developer	Link

- Map the screenshots to the class diagram. In other words, which class does the feature you show in the screenshot belong to? Multiple classes may belong to a single feature. Clearly describe what those are.
    - The first screenshot shows the profile scraper and profile analyzer input section
    - The second screenshot shows the profile scraper and the profile analyzer output section
    - The third screenshot shows the job scraper output section
    - The fourth screenshot shows the course scraper output section

4 Project Management

Continue to maintain the Change Log. Add any new changes to the project, tracking the date and description of each change. Use the table below:

Date	Description
March 22nd	Met with with Liz Brigham (Hurt Hub); changed project idea
March 25th	Formalized project scope; established Profile Scraper,

	Job Scraper, Profile Analyzer, and Course Scraper work items
March 27th	Wrote initial Job Scraper for extracting job posting data from LinkedIn
March 31th	Pushed initial Profile Analyer for using NLP to give feedback on a LinkedIn Profile PDF
April 2nd	Wrote initial Profile Scraper for extracting text from LinkedIn Profile PDF
April 3rd	Wrote Course Scraper for extracting course data from Coursera
April 3rd	Integrated Profile Scraper, Job Scraper, Job Analyzer, and Course Scraper into Flask app
April 3rd	Changed the styling of web page to look more LinkedIn-esque

5 Review and Retrospective

• What went well?
- Our dedication to the new idea helped us build the product very efficiently through the use of clean code and collaborative development.
- Sky's and Donald's project from earlier on and the integration of OpenAI API facilitated the building blocks of our projects so that we were more confident in our product.

• What didn't go well?
- The beginning of the development process was not as coordinated since we were moving into a completely new idea which caused us a bit of time deficiency.

• For the goals that were not met, what were the issues?
- At the beginning of the sprint, our goals were much different due to the different project plan.
- However, as soon as we shifted to the new idea, the goals we had were pretty much met during the sprint.
- Some tasks such as hosting the site or putting private keys on the public repo were delayed to administrative restrictions, but they can all be handled in the upcoming sprint.

• How do you plan to overcome the issues?
- We are planning to create .ENV files for each user and put these files in .gitignore so that they don't appear in the public repo.
- We also all plan to fork the repo so that we have administrative authority over our version of the code which is useful for hosting or using other SaaS tools on the repo.

• What do you plan to do differently in the next iteration?
- In the next iteration, we will have more user interviews in order to explore the user needs more in-depth and shape the product more towards these needs

- We also plan to create more sprint tasks in-detail so that everybody can get done with these micro tasks quickly and efficiently allowing others to complete their tasks in chronological order (when needed).

6 Team Management

• What were the team roles for this iteration?
- **Developers -** All team members
- **Product Owner -** Alp

• What did each team member contribute?
- **Delario Nance** **Profile Scraper -** A Python class with a *reader* attribute which is a *PdfReader* object that essentially lets the backend extract text from a given PDF. The class has a *get_text()* method which returns a string of all the text in the PdfReader's corresponding PDF.
- **Donald Lin** **Job Scraper -** A Python class with *keywords* and *location* attributes, which respectively are a string representing a job title and a string representing a city/state/country. The class has a *scrape()* method which searches LinkedIn for jobs matching the *keywords* and *location* attributes and returns a list of dictionaries which each store information about the found jobs
- **Sky Luo** **Profile Analyzer -** A class with *profile*, *job*, and *profile_keywords* attributes, which respectively are the text obtained from the Profile Scraper, a string representing a job title, and a variable which essentially is *None*. The class has an *analyze()* method which uses NLP to extract keywords from *profile* and then asks GPT-4 to give feedback on if the keywords match the job title and to give advice on skills to learn to better match the job title
- **Alp Niksarli** **Course Scraper -** A Python class with a *skills* attribute which is a list of skills that the user should gain. The class has a *scrape()* method which searches Coursera for courses which teach the skills stored in the *skills* attribute and returns a list of dictionaries which each store information about the found Coursera courses

• What were the challenges regarding team management, e.g., regular meeting, etc.?
- This sprint is relatively smooth and we have overcome most of the problems we encountered in the previous sprints. For instance, despite changing the direction of our project, all of us finished our tasks on time and we were able to produce a working software already. However, we still have the challenge of scheduling an in-person meeting with the career center advisor soon.

• What are the plans to overcome the challenges?
- We are quite on track of scheduling meetings with more customers to figure out more specific needs. This includes meeting with more students in STEM subjects, as well as meeting with a career advisor from Matthew's career center to seek a different perspective.

• If you were the third party who knows very well about your team, what suggestions would you give to your team?

- For each team member to put some amount of work toward their work items each day, even if it is a small amount. This could help each team member complete their work items closer to the start of the sprint than the end of the sprint.

7 Goals for the Next Iteration

• Write the next iteration's product log.
- Need a website URL/domain
- Need a profile scraper that takes LinkedIn profile URL as input
- Need a better profile feedback response format
- Need a better user interface design
- Need to show website is loading when the backend is doing scraping and analysis
- Need fewer and well-selected courses displayed
- Need a concise terms of service
- Need test cases

• Write the next iteration's sprint log.
- Infrastructure
  - Set up AWS EB Server Service for Hosting Website
  - Set up AWS Database for storing courses for different job titles & skills
- Profile Scraper
  - Build a class which can scrape a LinkedIn profile given the profile's URL
- Profile Analyzer
  - Refine the prompt template for ChatGPT 4. Define a structured answer format.
- Frontend
  - Decide the number of courses to display to the user, and then limit the number of displayed courses to that number
  - Divide the courses into essential courses and recommended extra courses
  - Display a pop-out icon showing work in process after the user inputs their LinkedIn profile (PDF) and desired job role
- Customers
  - Write Terms of service
  - Increase software value to people seeking soft skills in next sprint
  - Meet with at least 1 STEM major
  - Meet with at least 1 Matthews Center career advisor
- Testing
  - Plan test cases

• Other than the issues discussed in Section 6, i.e., Review and Retrospective, what potential challenges do you see in the next iteration?
- One potential challenge is making our software have value to people looking for ways to learn specific soft skills. The Philosophy major which we referred to in Section 1.1 stated that many skills which are important for his career are soft skills. However, soft skills are often less "teachable" than hard skills like Python and JavaScript.

• Briefly explain how your team would overcome each of the mentioned Challenges.

- We would ask the Matthews Center career advisor who we plan to interview for their advice on how to make our software have value for people seeking soft skills. For example, the career advisor might share with us different websites which offer courses specialized for teaching soft skills and then recommend us to add those in our software