

## Team: Luyao Wang, Di Jin, Nitin Perumandla

- **Project title**

Knowledge Graph Web App

- **Project description**

Freshmen who have just entered the university do not know very well about the course, skill and occupation. So we want to develop an app to help them. If they input their favorite occupations, the app can automatically show the course resources they need to learn, the order of course learning, and course-related skills. If they input the skills they want to learn, the app can display relevant course resources and jobs that require these skills. By entering the courses they like, the app can display the required technical traits that can be learned in the course and what job needs these skills can meet. Finally, the app can display all courses and their studying orders.

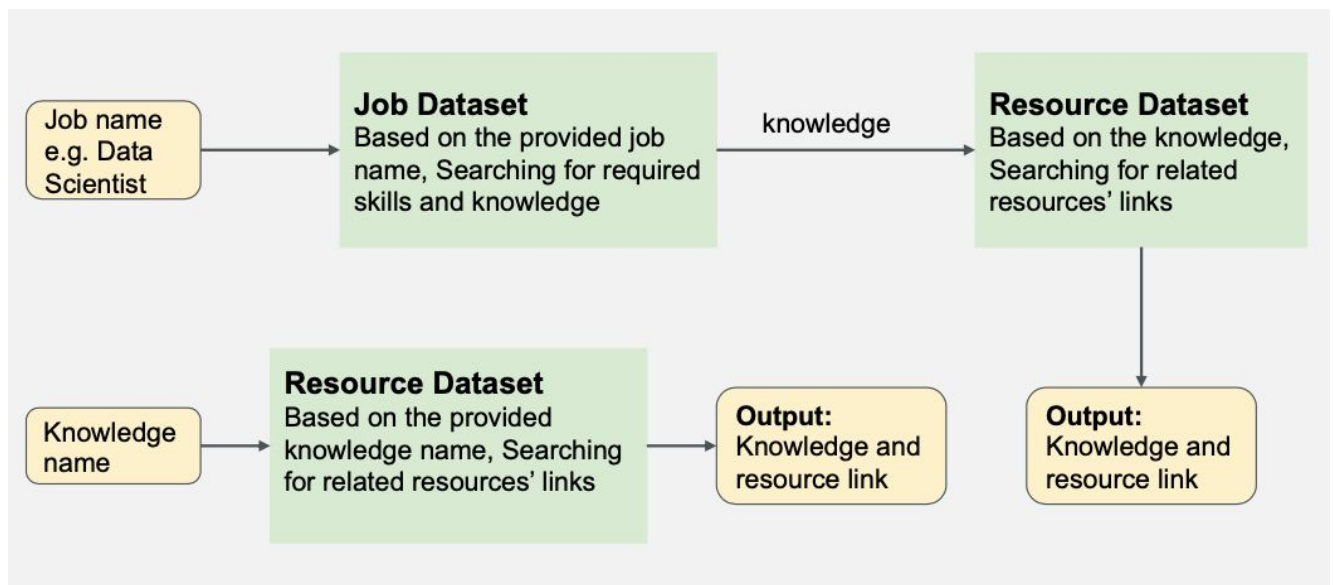
1. **Search function :**

Input : job name

Output 1: Skills from the job requirements, and the resource of courses from which users can learn these skills

Input : Skill name

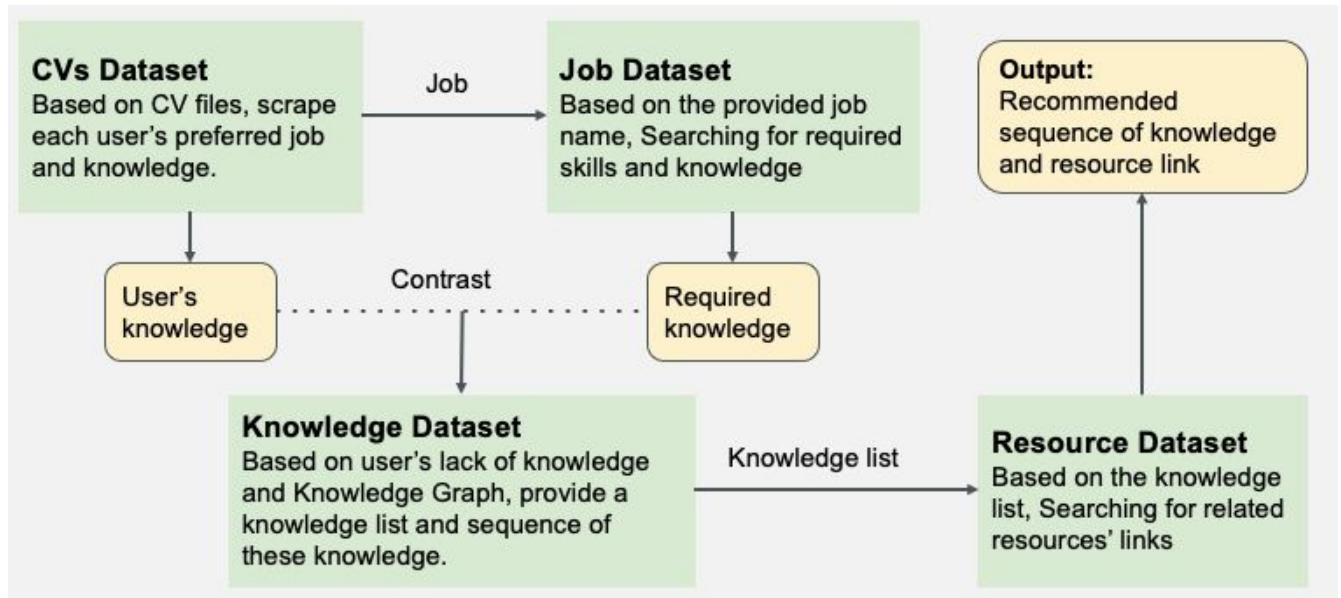
Output : The course's resources about these skills



## 2. Recommend function:

“Input” : User’s CVs, wanted jobs

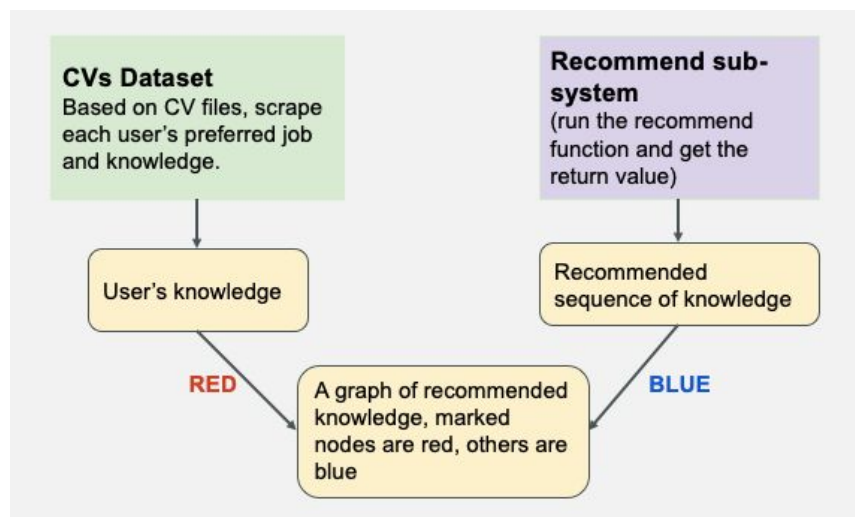
Output: For each user, we provide courses and prerequisites (relationships) that are related to all knowledge or skills he/she needs.



## 3. Mark function:

“Input” : User’s CVs, required knowledge from job posting dataset

Output: a graph including all required knowledge. The knowledge that the user has learned is marked as red while others are blue.



## • Data sets

1. **Job Posting Dataset:** (by using python, we can scrape useful information from the following website, we plan to collect 20-30 various job postings.)

E.g. <https://www.ziprecruiter.com/Jobs/Data-Scientist>

2. **Knowledge Dataset:** (this dataset includes all the knowledge, skills and concept nouns from our job postings, syllabus and user's CVs, the following link is an example of syllabus.)

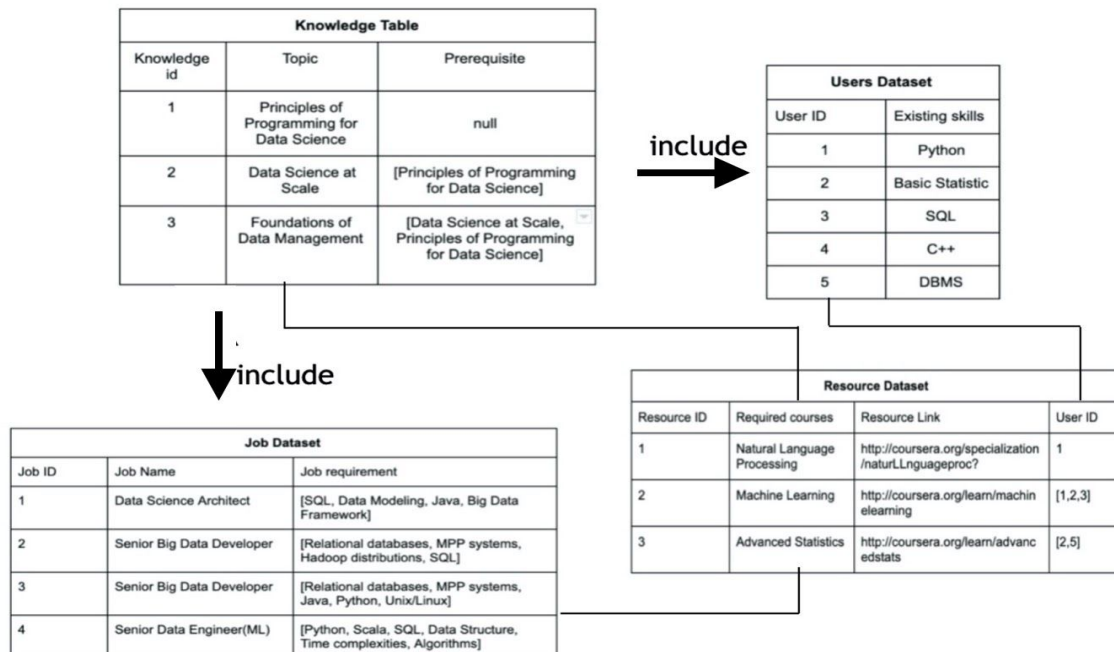
E.g. <https://web-app.usc.edu/soc/syllabus/20193/30259.pdf>

3. **Resources Dataset:** (almost all the courses can find related online resource on coursera, thus, we decide to collect links from coursera as the resource dataset)

E.g.

[https://www.coursera.org/search?query=python&utm\\_campaign=header-for-students&utm\\_content=corp-to-landing-for-students&utm\\_medium=coursera&utm\\_source=header-for-students-link](https://www.coursera.org/search?query=python&utm_campaign=header-for-students&utm_content=corp-to-landing-for-students&utm_medium=coursera&utm_source=header-for-students-link)

4. **User's CVs Dataset:** (Select some sample CVs, e.g. 10-20, and we can extract jobs' names and knowledge from each CV. Because CVs are private files, we plan to collect them in person, for example, we can ask our classmates for their CVs)



- **Data problems to be addressed**

1. **Data collecting:** we plan to learn how to use NLP to extract required knowledge from job postings and prerequisite classes from syllabus.
2. **Data integration:** For some same knowledge, they might be described in different ways, like “python” and “basic programming skill”, we need to identify them as the same knowledge and then only store “python” in our dataset.

- **Databases to be used and how to use them**

1. **Mysql:** to store the job posting dataset, knowledge dataset, resource dataset.  
**Firebase:** to store the users' CVs dataset.
2. We plan to load the data from Firebase to Mysql, and then show it to users, which could provide help to offline users.

- **Team members, background and skills.**

**Luyao Wang:**

1. Undergrad major: computer science
2. Skills: machine learning, python, relational database, software development.

**Di Jin:**

1. Undergraduate major : Management Information System
2. Skills: Python, Database management, Mysql, Tableau, SQL Server, Microsoft Project

**Nitin Perumandla:**

1. Undergraduate major: Computer Science and Engineering
2. Skills: Python, SQL Server, Machine Learning

- **Milestones and timelines.**

Milestones	Timelines
Project proposal	9/18/2020
Data collection	9/25/2020
Database design	9/30/2020

Web/App design	10/5/2020
Mid-term report	10/12/2020
Demo	11/1/2020
Project implementation	11/10/2020
Final report	11/23/2020