Proposal: Career-Skill Knowledge Graph

Ruijie Rao, Linkai Shao

Project Domain & Goals

In recent years, the career market has been changing dramatically. More and more employers are shifting from unidimensional Degree-based Hiring to a much more diverse talent acquisition mode: Skill-based Hiring (Harvard Business Review, 2022). This means that instead of focusing solely on the employee's college degree, the compatibility between the skillset required by the job and the skillset acquired by the employee matters more now. In the meantime, due to the pandemic's impact, self-learning platforms like Coursera and Udemy have been soaring for the last 2 years (Forbes, 2020). With access to the internet, anyone can learn any kind of skills (like Python, Java, SQL, UI/UX Design) they want online and get certified. A graduating student can prepare for his first career much easier using self-learning tools, while adult learners can also benefit greatly by upgrading their skillset and hence receiving a promotion or even changing a career. The rising of Skill-based hiring and self-certifying platforms seems to present a bright future for all employees, but actually, there is still a gap to be filled. Firstly, graduating students have no knowledge of what career they want to pursue. Secondly, people have no knowledge of what skills help them pursue their career paths.

These are the questions we want to answer using our **Career-Skill Knowledge Graph**. The knowledge graph consists of Job Titles and Skills. Job Title entities include attributes such as annual salary, number of postings(demand), job descriptions, job category, and related skills. Skill entities include attributes such as skill category(hard, soft, language), and skill description. By connecting job titles and skills and enriching their attributes, we aim to give any employee a better understanding of the current job market, while more importantly helping them to see their career path in a skill-based way.

Datasets

We will use structured data downloaded from Kaggle(Job Prediction, Job Description, Job Posting, Dice dataset) and job skills API, and hybrid data mined from websites such as Glassdoor, Indeed, Buildin, and Linkedin Jobs. These data sets and websites include information like job titles, descriptions, required skill sets, salary expectation, industry type, business, job postings, locations, etc. We use some existing ontologies (eg. Schema.org, Skills and Recruitment Ontology). From Schema.org, we will mainly use JobPosting and JobTitle. And from Skills and Recruitment Ontology, every ontology class is related to job careers. So we will use any of them when necessary. Currently, some potential candidates include subskill(used to indicate what basic skills are required for a complex skill), hasSkillGap(used to indicate the gap between users' current skills and target position), hasEssentialSkill(describes the most important skill for a position), etc.

Technical Challenge

First of all, to introduce more relevant and suitable information about all jobs to job seekers, we plan to make some job recommendations based on users' existing skills. We will build machine learning models to make recommendations and list all job positions with high skills matches. This could solve many graduating students' doubts about what kind of jobs match their skill sets learned at school. Secondly, for people who have targeting jobs, we will provide all relevant job information about the position (eg. required skills, salaries) to let them know what extra skills they need to learn or does the salary meet their expectation. We will solve this problem using SPARQL/Neo4j queries on our knowledge graph to extract all relevant information. Finally, we will examine our knowledge graph by inviting students/job seekers to use our interface. Based on their user experience, we will use accuracy as a metric and see how accurate our knowledge graph is able to display relevant information and make recommendations. For example, look at the top 3 recommended jobs and see if they match the user's target jobs. One big challenge during the creation process is extracting target skills from every job description we collect. Although we have access to a wide range of soft and hard skills, recognizing them from JD requires some rule-based NLP entity linking and matching. We will need to manually label a number of test data to evaluate this process. Another challenge is training the ML model for job title recommendation based on a given skillset. We decide to vectorize the required skills of all job title nodes and train a KNN model based on this skillset vector. A score based on cosine similarity will be used to evaluate the performance of this model.