

AI Skunkworks Project

Jobs Database

Project Manager: Anusha Jain - jain.anus@husky.neu.edu

Co-Project Manager: Mahalakshmi Arunachalam - arunachalam.m@husky.neu.edu
Anurag Singh – singh.anura@husky.neu.edu

Expected Start Date: March 20th, 2019

Expected Project Duration: TBD (For Spring 2019 students who need to submit a project, the deadline for tasks will be course project deadline)

Abstract:

Looking for jobs or internships seems a task of its own and the search is no longer based on sole fulfillment of the required job skills, but a lot of networking and recommendations is involved around too. The amount of work involved in finding the correct job builds a great amount of anxiety among the job seekers and the recruiters who want the right talent for their company.

There are two concerns that are to be addressed here. First, matching the job seekers with the right employers and second, provide guidance to aspiring job seekers on the skills that are in demand so that they can build them to stay relevant in the job market.

The job providers and job seekers form a large amount of data which provides for many interesting trends for analysis and interpretation to make the most of data available.

With the data currently available from the seekers and providers, these pitfalls can be fixed. The presence of information on job skills, salaries and user tendencies in many existing websites such as Indeed, LinkedIn, Glassdoor etc can be utilized to match people to positions which may seem simply impossible without using AI to analyze data.

The jobs database would be a one stop solution to reduce the job search and talent acquisition stress levels. Artificial intelligence (AI) and machine learning can be utilized for complex task of matching work to talent so that it is efficient and less resume spamming.

Domains for Job Database:

1. Finance
2. Health Care
3. Digital Marketing
4. Engineering – Computer vision
5. E-commerce

Team Size:

It should not be more than 3. Each team will be assigned with one domain.
Total -> 10-15 Students can work on this project

Prerequisites:

Basic knowledge of SQL and database management system

Tasks and Deadlines:

1. Create a list of 300/450 companies with their website and media page links (2 days)
2. Create a robot which extracts data from their website and media pages for one company. (5 days)
3. Generalize the code which can run for all the companies or for most of the companies (3 days)
4. Manually change the links for remaining companies to extract the data (4 days)
5. Create a word cloud for the most relevant skills for a particular domain (3 days)

Outcome:

By the end of this semester we will have a list of at least 1500 companies with their website and media pages links for 5 different domains. Apart from that, we will also have the list of relevant skills required for a particular domain.

Time allocation:

*Assumptions: Student will work for at least 4 hours/day on this project.

2 members team will have $\Rightarrow 2 \times 4 \times 17 = 136$ hours to complete this project

3 members team will have $\Rightarrow 3 \times 4 \times 17 = 204$ hours to complete this project

Note for Spring 2019 Students:

Students are welcome to come up with their own task proposals related to the project, but they are expected to complete all the 5 tasks mentioned above. Additional work will add more points to their project during evaluation.