Northeastern University Faculty Jobs Database

Suhas Balanagar, Ankita Brahmankar, Ankita Moharir

Abstract

This project is aimed to build, gather and model the data of jobs in an educational domain for teaching staff and faculty. The major objective is to help job seekers find jobs matching/suiting their profiles easily from a list of departments in that educational domain. The educational institute we are focusing on is Northeastern University, where we have developed a job recommendation system which returns a list of jobs suitable for the candidate based on the most demanding skills of the particular college.

Introduction

Searching for a job can be a full-time commitment. Job seekers need to find a position that suits them and their abilities, which will pay them commensurate with their skills and experience, for an employer that they find acceptable. This is a tall order. Not only that, the process of searching and applying for such a position is riddled with additional challenges that make job hunting an often unpleasant, frustrating experience. Here are the five top challenges of talented job seekers in today's market: too many places to look, uninformative job description, lengthy and confusing hiring processes and not enough information.

Candidates want to know about their potential employers before they apply for a position. Most job seekers spend a couple of hours researching a college before they apply. Unfortunately, universities often don't have information about their colleges, its services, culture, values. Potential candidates have to play an online game of hide-and-go-seek for college information that involves combing through information on Better Business Bureau, Glassdoor, Google, Yahoo Finance, and other resources in order to learn about prospective employers.

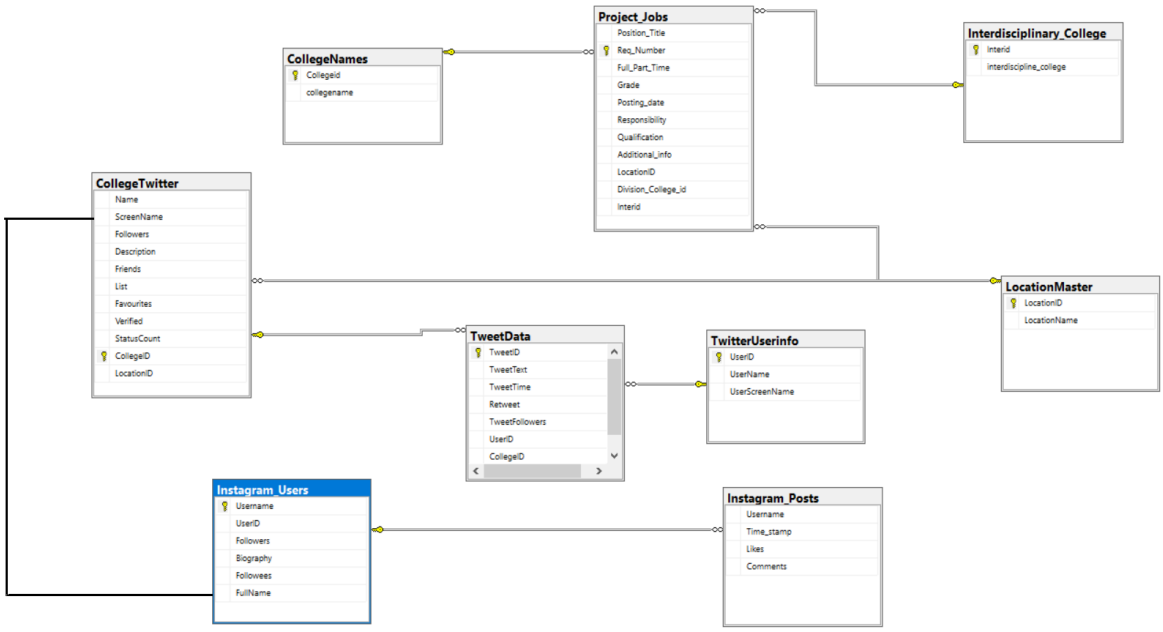
Proposed Solution

With help of the existing websites, we provide a place where job seekers can match with the right employers based on their qualified titles and the recommended jobs. For doing so, data is first collected from the Northeastern University website and then stored into the tables where it is stored and modeled based on the requirements and most famous skills. Data collection is done by scraping the data from website using Python Beautiful soup, a package for parsing html and xml documents, which is useful for scraping. The data obtained is further saved into panda data frames, which represents a tabular representation of data. We

establish a database connection with python using Microsoft server management studio to store our data and then help employers find their best job which suits their qualification.

Reviews related to the college is gathered from the social media platforms like twitter and instagram , making easier for job seekers to make a job application choice. A word cloud for the most occurring jobs description is created mainly to help and suggest suitable jobs for the seeker. We make a basic user interactive interface for the job seeker to enter and obtain the kind of jobs suited and required for him.

Entity - Relationship Diagram



Implementation

To implement the mentioned solution, we need to gather data from the job postings from the university's website by scraping all the details. For scraping, the university's job portal website was scraped with the help of HTML nd CSS tags. Using these tags, the required data can be scraped and then stored either into a dataframe or into a local file. To obtain the jobs for faculties, we need to establish functional dependencies on the details. Therefore, we create dataframes from the data collected by scraping from website which contain all the details related to the job. Also jobs are found on Social networks, so the jobs and the details related to the

company/college/user posting it are scraped from social networking sites like Twitter, Instagram with the help of APIs. Using these APIs and their parameters most of the data from social networks are obtained. Data of all the jobs posted by the colleges of that university along with other details are stored in a data frame or written into a local file. Also the details related to the college, the jobs posted, their location, college names and their social media like number of followers, number of posts posted, total number of likes etc., information are stored in a dataframe. This data is then imported into the SQL database by creating the appropriate tables.

Data Normalization is performed on the data to

USE CASES

1. Fetch jobs based on the job skill provided by the job seeking person
2. Fetch jobs based on the Job position provided as an input
3. Get most posted job title based on the location where it is available
4. Get the tweets by a particular college
5. fetch the jobs based on college name
6. Based on the input provided for position type, get the relevant
7. Get relevant Jobs based on the Location
8. Get the maximum followers based on the college
9. Fetch the job posted during last number of days
10. Get the job type based on the college
11. Get the twitter user data
12. Get the maximum tweets posted by a college
13. Fetch the most recent tweet by a college
14. Fetch the tweet which is very popular
15. Get the tweets based on the location provided

Functions:

1. A function that is used to get the most active user who is verified when the name is provided
2. Get the part time jobs available for that particular position
3. A function to getch the maximum retweets done by a person
4. To get the followers count in twitter when a name is provided
5. Based on the job type i.e full time get the full time jobs for that position
6. Get the part time jobs count based on the college name

Views:

1. Gets the job positions available
2. Gets the tweets count based on the colleges
3. Fetches the twitter data

4. Gets the tweet data with user
5. Fetch the instagram post data
6. Get Tweet count based on the user

Contributions

Suhas Balanagar, 001356872 - Performed scraping of the websites with the help of html tags and stored it in the database. Then using this data, performed sql queries by creating tables, inserting the data, creating views, functions and stored procedures. Then connecting the database to the python notebook and creating switch cases. Then writing an in-detail report of the project explaining all the functionalities.

Ankita Brahmankar, 001389907 - Performed scraping of the Twitter data with the help of API stored it in the database. Then using this data, performed sql queries by creating tables, inserting the data, creating views, functions and stored procedures. Then connecting the database to the python notebook and creating switch cases. Then writing an in-detail report of the project explaining all the functionalities.

Ankita Moharir, 001356837 - Performed scraping of the Instagram data with the help of API and stored it in the database. Then using this data, performed sql queries by creating tables, inserting the data, creating views, functions and stored procedures. Then connecting the database to the python notebook and creating switch cases. Then writing an in-detail report of the project explaining all the functionalities.

Citations

Learnt and gathered knowledge from the below sources:

https://developer.twitter.com/en/docs.html

http://socialmedia-class.org/twittertutorial.html

https://www.programiz.com/python-programming/anonymous-function

https://github.com/nikbearbrown/INFO_6210/

https://www.datacamp.com/community/tutorials/wordcloud-python

https://pypi.org/project/instaloader/

https://towardsdatascience.com/web-scraping-html-tables-with-python-c9baba21059

Text License

Below is the license attached.