

Final Project Report

Group 6: Ankita Naikdalal, Heet Palod, Kulraj Singh Kohli, Varun Kandukuri, Vaidehi Patil

Project Name: Job Recommendation Engine - Hunt4Jobs

Problem Statement:

To design an online job portal that will provide a platform for job-seekers to get job recommendations based on the job-seekers skillset.

Motivation:

One of the primary functions of websites like LinkedIn, Glassdoor, etc. is to try and match job-seekers to appropriate postings based on their interests and skill sets. This is an exercise that many of us students go through on a frequent basis in our job hunt.

Given that there are thousands of postings every single day, we wanted to create a platform that would help job-seekers to navigate through these listings and find appropriate jobs for them in an efficient manner.

Core Functionalities:

- Candidates will be able to sign-up and register on the website.
- Candidates will be able to login and enter their academic experience, work experience and skillset on the website.
- Candidates will be able to view a list of job recommendations matching their own skill set.
- The website administrator will be able to perform basic CRUD operations on candidate profile/personal details.

Technologies used:

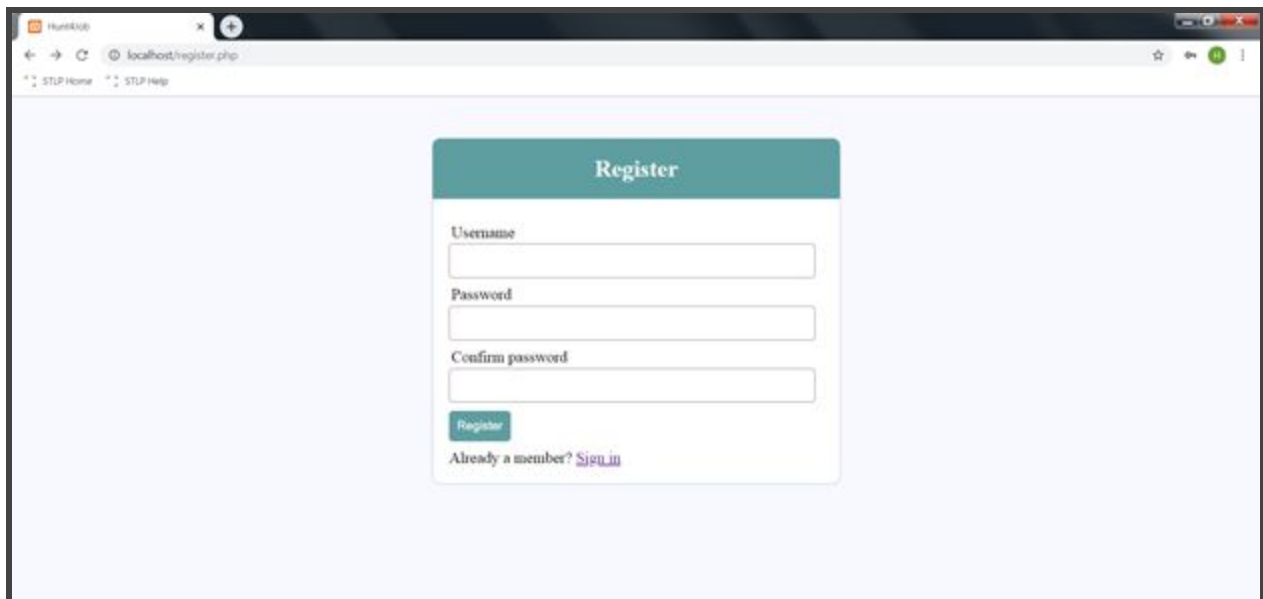
- The job portal website is created using HTML5, CSS3, PHP, JavaScript and XAMPP
- phpMyAdmin (MySQL) is used for managing and storing the data
- Python Selenium is used for web scraping job posting data from job portals

Methods:

I. Website Creation:

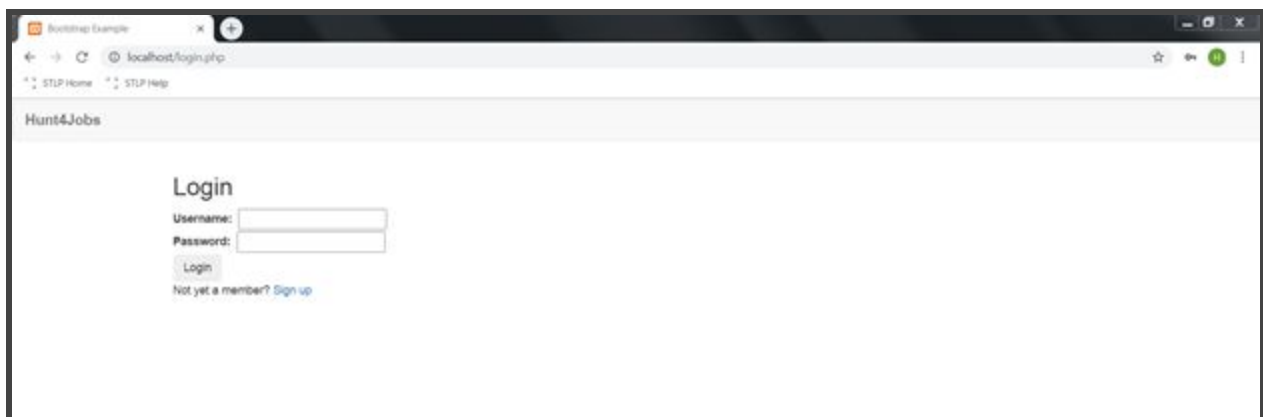
Our website “Hunt4Jobs” has the below pages through which the candidate can navigate to search for relevant job postings matching his skill set. We have enforced data validation checks on all the data entry pages to ensure that the data entered by the candidate is non-empty, clean and in the correct format before storing it into our database.

A] Registration/Sign up Page



The screenshot shows a web browser window with the address bar displaying 'localhost/register.php'. The page has a light blue background. In the center, there is a white rectangular box with a teal header that says 'Register'. Inside this box, there are three input fields labeled 'Username', 'Password', and 'Confirm password'. Below these fields is a teal 'Register' button. At the bottom of the box, it says 'Already a member? [Sign in](#)'.

B] Login Page



The screenshot shows a web browser window with the address bar displaying 'localhost/login.php'. The page has a light blue background. At the top left, there is a header 'Hunt4Jobs'. Below it, there is a 'Login' section. It contains two input fields labeled 'Username:' and 'Password:'. Below these fields is a 'Login' button. At the bottom of the section, it says 'Not yet a member? [Sign up](#)'.

Server: 127.0.0.1 » Database: test » Table: user_login

Showing rows 0 - 5 (6 total, Query took 0.0000 seconds.)

```
SELECT * FROM `user_login`
```

☐ Show all | Number of rows: 25 | Filter rows: Search this table | Sort by key: None

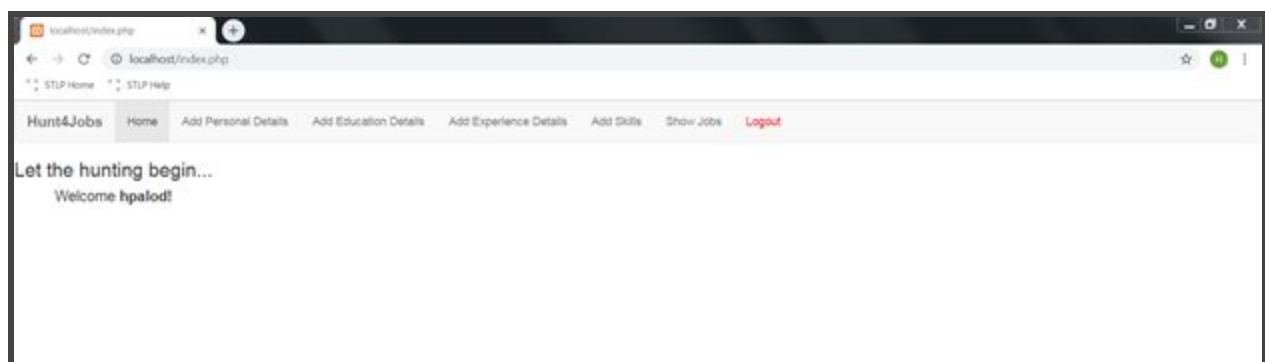
+ Options

				id	username	password	date_of_creation
<input type="checkbox"/>	Edit	Copy	Delete	1	hpalod	9cdfb439c7876e703e307864c9167a15	2020-06-04 00:19:32
<input type="checkbox"/>	Edit	Copy	Delete	2	varunk	149e200961ea52955bd55cb49a127935	2020-06-04 11:59:40
<input type="checkbox"/>	Edit	Copy	Delete	3	ankita007	de5b5bf65ba66517f9b70b1443d2102d	2020-06-04 12:54:41
<input type="checkbox"/>	Edit	Copy	Delete	4	coolraj	1525640fa90656c3fb4d44e9f020d64b	2020-06-04 13:05:28
<input type="checkbox"/>	Edit	Copy	Delete	5	vaidehip	5772f5dd7afb90241fe9265becd65703	2020-06-04 13:24:59
<input type="checkbox"/>	Edit	Copy	Delete	6	david	5f532a3fc411ea403f37070f59a7a53a	2020-06-04 15:57:47

☐ Check all | With selected: Edit Copy Delete Export

To store the **password** after filling up the 'Sign Up' form, we encrypt the user-input string using MD-5 algorithm before storing it in the database. For example, we created a username 'david' with password 'microsoft'. But, it got stored in the database with its MD5 encrypted password, as highlighted yellow in the above figure.

C] Home Page



D] Add Personal Details Page

The screenshot shows a web browser window with the address bar displaying 'localhost/add.html'. The page title is 'Add Personal Details'. The navigation bar includes 'Hunt4Jobs' and links for 'Home', 'Add Education Details', 'Add Experience Details', 'Add Skills', 'Show Jobs', and 'Logout'. The main content area features a form titled 'Add Personal Details' with the following fields:

- First Name:
- Last Name:
- Date of Birth: with a calendar icon
- Gender:
☐ Male
☐ Female
- Email:
-

E] Add Education Details Page

The screenshot shows a web browser window with the address bar displaying 'localhost/add_education.html'. The page title is 'Add Education Details'. The navigation bar includes 'Hunt4Jobs' and links for 'Home', 'Add Education Details', 'Add Experience Details', 'Add Skills', 'Show Jobs', and 'Logout'. The main content area features a form titled 'Add Education Details' with the following fields:

- Highest Degree Achieved:
- Major:
- University:
- Graduation Date: with a calendar icon
- GPA:
-

F] Add Skill Page

The screenshot shows a web browser window with the address bar displaying 'localhost/add_skill.php'. The page has a navigation bar with 'Hunt4Jobs' and links for 'Home', 'Add Personal Details', 'Add Education Details', 'Add Experience Details', and 'Logout'. The main content area is titled 'Add Skills' and contains a list of skills with checkboxes:

- ☐ Python
- ☐ SQL
- ☐ Java
- ☐ JavaScript
- ☐ Django

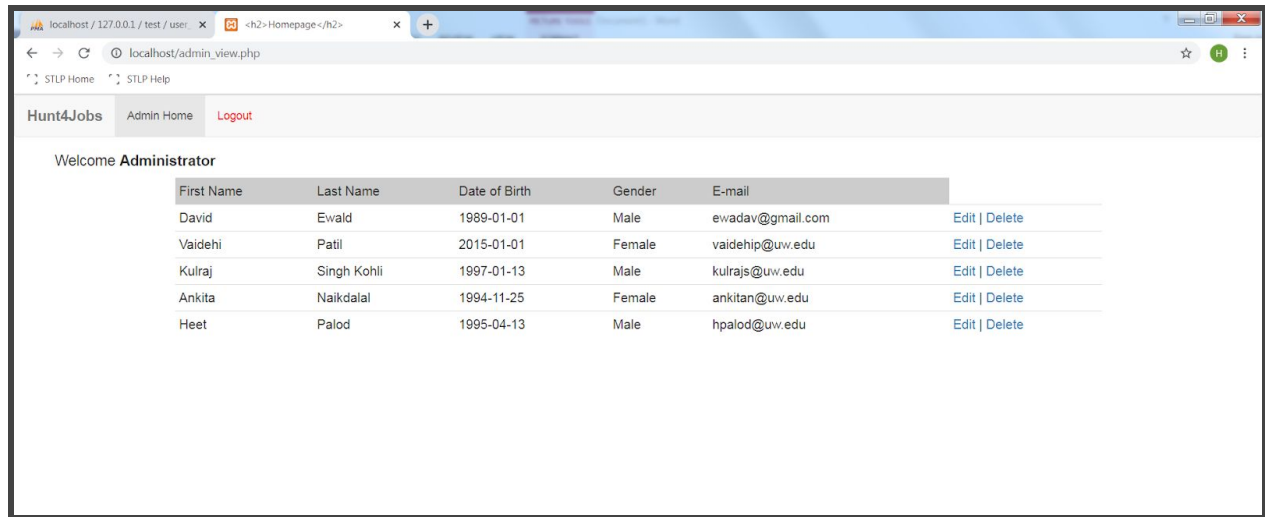
G] Show Jobs Page

The screenshot shows a web browser window with the address bar displaying 'localhost/show_jobs.html'. The page has the same navigation bar as the previous page. The main content area is titled 'Relevant Job Postings (5)' and displays a table of job postings:

Job ID	Employer Name	Job Title	City	Job Description	Skills
1	Amazon	Data Scientist	Seattle	Knowledge regarding building and testing ML Models in supervised and unsupervised machine learning.	Python, SQL, TensorFlow
2	Microsoft	Data Engineer	Seattle	As a member of the Engineering Services and Data Infrastructure team, you will be using your software development skills to architect systems and services to modernize our engineering infrastructure to make the product development experience at Tableau the best in the industry.	Python, SQL, Tableau
3	Spark Cognition	Data Engineer	Seattle	Provide Business Intelligence (BI) and Data Warehousing (DW) solutions and support by leveraging project standards and leading analytics platforms.	Python, NoSQL, Tableau
4	SalesForce	Software Developer	Seattle	Salesforce has immediate opportunities for extremely talented software developers who want their lines of code to have significant and measurable positive impact for users, the company's bottom line and the industry.	Java, SQL, NoSQL
5	Apple	Software Developer	Seattle	Apple Inc. is an American multinational technology company that designs, develops, and sells consumer electronics, computer software, and online services. It is considered one of the Big Tech technology companies, alongside Amazon, Google, Microsoft and Facebook.	Java, Python, SQL, NoSQL

The 'Show Jobs' page populates relevant job postings based on the maximum Jaccard similarity of the user's/candidate's selected skill set on the 'Add Skills' page and the job posting's required skill set.

H] Admin View Page



Using the 'Admin View' page, the administrator can read/edit/delete personal details of the registered candidates.

Source Code Reference: https://github.com/heetpalod/imt563_project

II. Job Matching Algorithm:

To match all candidates to the most relevant jobs based on their skillset, we utilized a metric known as Jaccard Similarity.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

The idea is simple — allocate a similarity score between two sets based on the proportion of elements between them that overlap.

For our purposes, set A in the above formula would be a set of skills needed for one particular job and set B would be a set of skills that a candidate possesses.

For Example:

Skills needed by job (A) = {Python, SQL, Tableau, SSIS, Power BI}

Skills possessed by candidate (B) = {R, SQL, Tableau}

Jaccard Similarity between A and B = $2 / 6 = 0.333$

We repeat the above process for each candidate against every job in our repository, and sort the results in descending order of similarity. The final outcome is that we have the most relevant jobs being displayed for each candidate.

Python code: https://github.com/heetpalod/imt563_project/blob/master/job_skillset.ipynb

III. Data Collection using Web-scraping:

To collect the job listings, we utilized python's selenium package that helped us to scrape the data dynamically off the web.

1) Create a web driver for chrome

```
from selenium.webdriver import Chrome
import time
#connect to chrome

browser = Chrome()

#open web page

browser.get('https://www.indeed.com/l-Seattle-jobs.html')
time.sleep(10)
```

2) Using Xpath we locate the data in HTML code and store it in the appropriate form

```
# Extracting the names of jobs from the page

job_names = browser.find_elements_by_xpath('//*[@class="jobtitle turnstileLink " ]')
for names in job_names:
    JobNames.append(names.text)

# print(len(JobNames))

# Extracting the Address of hospitals from the page

job_address = browser.find_elements_by_xpath('//*[@class="location accessible-contrast-color-lo
for address in job_address:
    JobAddress.append(address.text)

# # Extracting the Ratings of hospitals from the page

summary = browser.find_elements_by_xpath('//*[@class="summary"]')
for summ in summary:
    Summary.append(summ.text)

# Extracting the numbers of hospitals from the page

company_names = browser.find_elements_by_xpath('//*[@class="company"]')
for cnames in company_names:
    CompanyNames.append(cnames.text)
```

This 'collected data' is then exported as an excel file and stored on a local machine, after then transferred to our phpMyAdmin (MySQL) database.

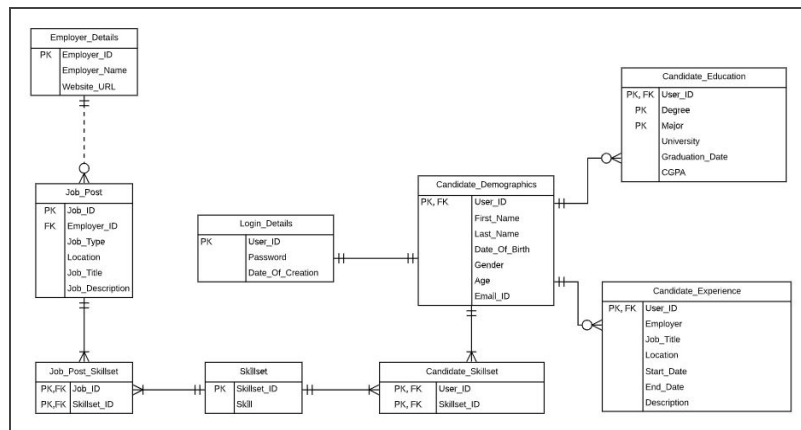
3) We managed to scrape 1400 jobs in the Seattle area with the position name, Address, Company Name, Requirements, City, State and Pincode from the website 'www.indeed.com'

1		Position Name	Work_Address	Company	Requirements	City	State_Pin
1360	1358	Produce Clerk	Seattle, WA 98109 (East Qu	Metropolitan Market	Prepares products for sale; prices and stocks merchandise according to guidelines.Job S	Seattle	WA 98109 (East Queen Anne area)
1361	1359	Production Worker	Kent, WA 98032	Modular Arts, Inc.	4 day per week positions available (10 hour days) Mon-Thurs 7:00am to 5:30pm.We prc	Kent	WA 98032
1362	1360	Machine Operator	Renton, WA 98057	Distant Lands Coffee	The key elements necessary for this position are DLC product knowledge, specific mach	Renton	WA 98057
1363	1361	Universal Worker	Seattle, WA 98101 (First Hill	Merrill Gardens at First H	Demonstrated knowledge of building maintenance, electrical, mechanical, painting, plu	Seattle	WA 98101 (First Hill area)
1364	1362	Security Program/Projec	Seattle, WA	King County	Communicate project risks, develop project risk mitigation strategies and ensure projec	Seattle	WA
1365	1363	20-21 Asian Counseling	Seattle, WA	AmeriCorps	Members will also build upon existing curriculum and develop new teaching materials f	Seattle	WA
1366	1364	Light Weight Foam Oper	Renton, WA 98057	Sealed Air Corporation	N Manufactures sufficient quantities of quality products at or below budgeted producti	Renton	WA 98057
1367	1365	Scribe	Bellevue, WA 98004 (Overl	Columbia Asthma and Al	The majority of the work will be assisting the provider by entering information into the	Bellevue	WA 98004 (Overlake area)
1368	1366	Personal Banker	Seattle, WA 98118 (Seward WaFd Bank		The Personal Banker also participates in making out of office calls to businesses and ref	Seattle	WA 98118 (Seward Park area)
1369	1367	UPS Redmond Twilight P	Redmond, WA 98052 (SE R	United Parcel Service	If you're a student at an approved college, university, trade or technical school, UPS off	Redmond	WA 98052 (SE Redmond area)
1370	1368	Sandwich Artist	Bellevue, WA 98007 (Lake	Subway (Bellevue Franch	Crew member exhibits a cheerful and helpful manner while greeting guests and preparin	Bellevue	WA 98007 (Lake Hills area)
1371	1369	Lead Technician, Biologi	Seattle, WA	Lumen Bioscience	BA/BS degree in biochemistry, cell biology, biotechnology or related field of study.Enga	Seattle	WA
1372	1370	IT Manager	Seattle, WA 98103 (Fremont	Mentor Creative Group	Manage the entire organization's technology needs, including develop, maintain, and up	Seattle	WA 98103 (Fremont area)
1373	1371	Parcel Assistant- Wareh	Kent, WA 98032	FedEx Ground PH US	1am Monday through Friday-You must be willing and able to be trained & certified for	Kent	WA 98032
1374	1372	Dispatcher	Everett, WA 98204 (Holly	ai Comcast	Support field technician activations and troubleshooting by utilizing department tools ai	Everett	WA 98204 (Holly area)
1375	1373	Customer Service Repre	Seattle, WA	Trupanion	Emotional resilience: be aware of one's own emotional reactions and adapt to stressfu	Seattle	WA
1376	1374	Bank Teller	Seattle, WA 98136 (West S	WaFd Bank	The position of Customer Service Representative is responsible for providing courteous	Seattle	WA 98136 (West Seattle area)
1377	1375	Appointment Setter - M	Seattle, WA	Blue Flame Heating, Air,	This high-volume, popular brand location offers the perfect setting to engage with cust	Seattle	WA
1378	1376	Customer Care Represe	Seattle, WA	Embrace Pet Insurance	Being a member of the Contact Center means responding to 70-90 phones calls per day	Seattle	WA
1379	1377	SOCIAL WORKER 1 OR	Seattle, WA	University of Washingto	A Master of Social Work Degree from a program accredited by the Council on Social W	Seattle	WA
1380	1378	Scheduling Coordinator	Duvall, WA 98019	Bruner Orthodontics	Dental and/or sales experience a plus!This is a full-time position Monday-Thursday and	Duvall	WA 98019
1381	1379	Leasing Agent [Zella]	Seattle, WA 98119 (Queen	Blanton Turner	You will be the face of our organization, and have a chance to engage with prospective	Seattle	WA 98119 (Queen Anne area)
1382	1380	Processor - Commercial	Seattle, WA 98121 (Downt	American Seafoods Com	Able to be on-board ship for long periods, without suffering from motion sickness (that	Seattle	WA 98121 (Downtown area)
1383	1381	Administrative Specialist	Seattle, WA 98104 (First Hill	P.I.E. Management, L.L.C	Answer telephone calls to the designated disposition authorization lines and respond	Seattle	WA 98104 (First Hill area)
1384	1382	Merchandiser -	Renton, WA 98057	Columbia Distributing	Valid driver's license and auto insurance required.Accurately clock in and out, adhering	Renton	WA 98057
1385	1383	Warehouse Associate/Fr	Kent, WA 98032	H&M Bay Inc.	Perform proper forklift inspections and follow forklift operating guidelines and procedu	Kent	WA 98032

Python code: https://github.com/heetpalod/imt563_project/blob/master/Job-Scrape.ipynb

Database Implementation:

Entity-Relationship Diagram



We used phpMyAdmin SQL version 5.0.2 to implement our database. We have hosted it on localhost (127.0.0.1) with Server version: 10.4.11-MariaDB

SQL Database Creation/Queries Code:

https://github.com/heetpalod/imt563_project/blob/master/database.sql

Limitations:

- Due to time constraints, we have currently web-scraped only about 1000 job postings located in the Seattle area.
- We have created a portal only to be used by candidates and not the employers. Our employer data is auto-generated.
- We have created the job portal website on local host and plan to host the same on cloud in due course of time.

Future Scope:

- We plan to migrate our solution to AWS and make the web portal centralized in the future.
- We also plan to include employer registration and job posting management features from the employer's end into our project in the future.