SPRING 2024 CS5334 ADVANCED INTERNET INFORMATION PROCESSING PROJECT REPORT

SUBMITTED BY:

KRISHNA VAMSI SOMU (A05211570) – wcy14 SAI PRANAV DATRIKA (A05187372) – ckw91

Contents

SPRING 2	SPRING 2024 CS53341							
	D INTERNET INFORMATION PROCESSING							
PROJ	ECT REPORT	1						
	INTRODUCTION							
	EXISTING SYSTEM BEHAVIOUR							
	EEATI DES INDI EMENTED							

1. INTRODUCTION:

This is a standard job search for a fresh job application. By giving the client the chance to choose appropriate alternatives based on the job title, job type, specialization field, geophysical location, expected minimum salary, and preferred company, the system delivers correct matching results. This project was implemented with Python and Oracle.

The report's remaining sections are arranged as follows:

- Description of the project's existing system behavior.
- Description of the project's implemented features.
- 1. Use job title field as part of the ratings
- 2. Use company field as part of the ratings
- 3. Keyword feature

2. EXISTING SYSTEM BEHAVIOUR:

The key features of the demo program's current job search are explained in this section using Python and Oracle.

JOB Table

For this project, a JOB table has been developed to manage and store data related to job searches. When the web server receives a request for a job search, it initiates the creation of a basic SQL query within the JOB table.

The job search's web interface provides search drop down menus are: -

- a. Job Type
- b. Job Title
- c. Specialized field
- d. Keyword Search
- e. Search by region OR
- f. Search by state OR
- g. Search by city

- h. Expected minimum salary.
- i. Preferred Company

Job Type

Based on the existing behavior when the user searches for jobs with job type as 'regular', the search will **only** return jobs with job type '*regular*'. As part of the project, those jobs with other types also should be returned with some assigned penalty on ratings. There is a default value of 100 displayed for rating for all the matched rows returned.

Company Field

When the user searches for jobs with company field as 'Google', then **only** those jobs with Company's name as 'Google' company are returned. As part of the project, along with Google's jobs, all other related companies' jobs are also to be returned with some penalty assigned on the ratings. There is a default value of 100 displayed for rating for all the matched rows returned.

Keyword Feature

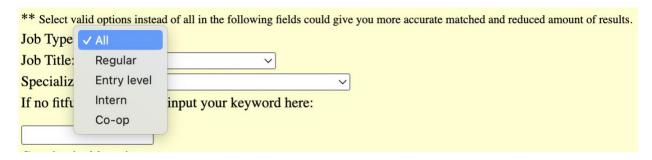
Based on the existing behavior the Web UI has *keyword* field to input a value, but it's not implemented. The keyword feature is basically checking the *description* column field in the JOB table.

3. FEATURES IMPLEMENTED:

1. Use the job type field as part of the ratings

To implement this feature, we made the following changes to the code in the file ematch.py

➤ We introduced a new variable **self.askedJobType**. This variable accepts the input from the UI where we select the required job type from the available 4 options under Job Type.



➤ We updated the code in the function **build_sql_query_stmt()**. The existing code would build the SQL query in such a way that it would retrieve only those records

from the JOB table which had the job type name exactly same as the one selected by the user. As part of this project, our aim was to display all other jobs of remaining job types as well but with some penalty. We have given the penalty as 10 for unmatched job type records.

We introduced a new function called **compute_job_type_penalty.** This function takes in a parameter – "jobType". If the askedJobType equals with the jobType that is coming from the database, then the penalty is 0 otherwise the penalty is set to 10. This function returns penalty to the **compute_ratings** function where this function is called to update the **jobRating** as per the condition.

2. Use company field as part of the ratings

To implement this feature, we made the following changes to the code in the file ematch.py.

- ➤ We declared a new class variable **self.askedCompany** in the class ematch_class. This variable accepts data from the Web UI where user selects preferred company.
- ➤ We updated **ematch_data_struct.py** to include 2 new dictionaries. They are **company_array** and **relatedCompanies.** In the company_array, we kept the company names as keys and have given same value for related companies. In relatedCompanies, we used the above unique values as keys and the related companies together in an array as value for each key.
- ➤ We declared a new function **find_matched_companies** which takes a parameter **searchedcompany.** In this method, the dictionary company_array from ematch_data_struct is called, so that it gives us the required value. We used this value to find all the related companies from relatedCompanies dictionary. This find matched companies returns the list of all related companies.
- We updated the **sql_stmt** to include only related companies. A penalty has been set for the related companies so the jobRating will be reduced. Previously, We have organized related companies details in company_array. In this dictionary, we have set a common integer value for related companies. With this common integer, we have assigned penalty as the sum of this value and 10.

3. Keyword feature

To implement this feature, we made the following changes to the code in the file ematch.py

➤ We introduced a new variable **askedKeyword**. This variable takes the input from the keyword field in the UI.

** Select valid options instead of all in the following fields could give you more accurate matched and reduced amount of results.						
Job Type: All						
Job Title: All						
Specialized field: All						
If no fitful field is found, input your keyword here:						
develop						

- ➤ We created a new dictionary data structure with name company_dict where it will store company name as key and its respective description as value from the database.
- ➤ We introduced new function **find_keywords**. This function searches the above dictionary's values to find whether the given keyword exists or not. If the given keyword exists as part of description for any company, then those companies were added to a new dictionary named **self.filteredcompanies**.
- In this dictionary, if the keyword matches exactly then a value '1' is assigned, if the keyword matches only partially then the value of '2' is assigned.
- We created a new function called **keyword_penalty** to assign jobratings based on the matched keyword. If the keyword is matched exactly, then the penalty is 0. If it matched partially, then a penalty of 10 is assigned.
- ➤ In the **process_results**, we used the above described keyword_penalty and computed the jobRating. Here, we have the filtered records with updated job ratings.

To implement this feature, we made the following changes to the code in the file jobsearch.py

> Here , we updated the code to call the **find_keywords** function before process_results and sort_results functions.

4. URLs TO ACCESS THE PROJECT

http://newfirebird.cs.txstate.edu/~ckw91/demo/proc/unix-version/html/index-py.html http://newfirebird.cs.txstate.edu/~wcy14/demo/proc/unix-version/html/index-py.html

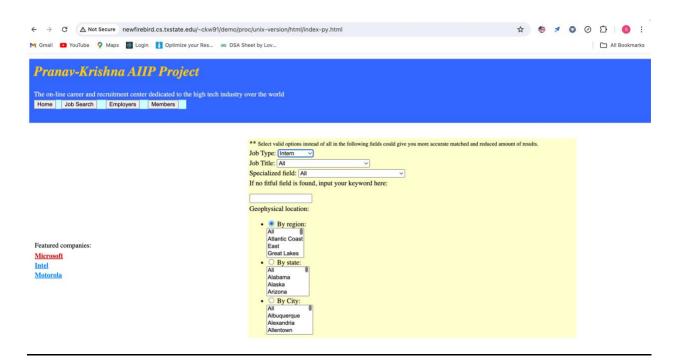
5. SCREENSHOTS:



Index-py.html



JobType: We updated job type field for some records as "intern". As we searched for "intern" job type, we get those records with job ratings as 100 and others with penalty of 10.





(Total 104 matches)

Job Number	Rating	Job Type	Job Title	Special field	Company	City	State	Region	Salary Range
1	100	intern	Engineer	Accounting	Motorola	Austin	Texas	N/A	32000 - 40000
2	100	intern	Engineer	Device	Motorola	Austin	Texas	N/A	32000 - 40000
3	100	intern	Senior Engineer	Mac Development	Apple Computers	Austin	Texas	N/A	35000 - 50000
4	100	intern	Engineer	Java Development	Apple Computers	Oakland	N/A	Pacific	56000 - 64000
5	100	intern	Engineer	System Testing	Micron	Boise	N/A	Northwest	49000 - 65000
6	100	intern	Engineer	Windows Development	Intel	Portland	N/A	N/A	32000 - 40000
7	100	intern	Engineer	Process	Intel	Santa Fe	N/A	N/A	32000 - 40000
8	90	regular	Engineer	Analog Design	Cirrus Logic	Austin	Texas	N/A	50000 - 58000
9	90	regular	Engineer	System	Cirrus Logic	Austin	Texas	N/A	55000 - 63000
10	90	regular	Engineer	System	Lockheed	Albuquerque	N/A	Southwest	60000 - 68000
11	90	regular	Senior Engineer	UNIX System Programming	BMC	Austin	Texas	Southwest	45000 - 53000
12	90	regular	Engineer	Digital Design	Cirrus Logic	Austin	Texas	N/A	30000 - 38000
13	90	regular	Manager	Manufacturing	Intel	San Jose	California	West	80000 - 88000
14	90	regular	Engineer	Digital Design	Dell	Austin	Texas	N/A	30000 - 55000
15	90	regular	Senior Engineer	UNIX Administration	GE	Green Bay	Wisconsin	N/A	35000 - 45000
16	90	regular	Engineer	UNIX System Programming	Chase	Columbia	South Carolina	N/A	40000 - 60000

(Page 1 of 7 pages) Goto page: 1 <u>2</u> <u>3</u> <u>4</u> <u>5</u> <u>6</u> <u>7</u> ≥



qstr=job_type=intern&job_title=all&specialty=all&field_keyword=&location_type=region&min_salary=10000&company_name=all====

Database version: 11.20.1.0 Beginning of building sql_query_stmt job_type=intern&job_title=all&specialty=all&field_keyword=&location_type=region&min_salary=10000&company_name=all ['job_type=intern', 'job_title=all&specialty=all&field_keyword=&location_type=region&min_salary=10000&company_name=all ['job_type=intern', 'job_title=all&specialty=all&field_keyword=&location_type

Job Search Results

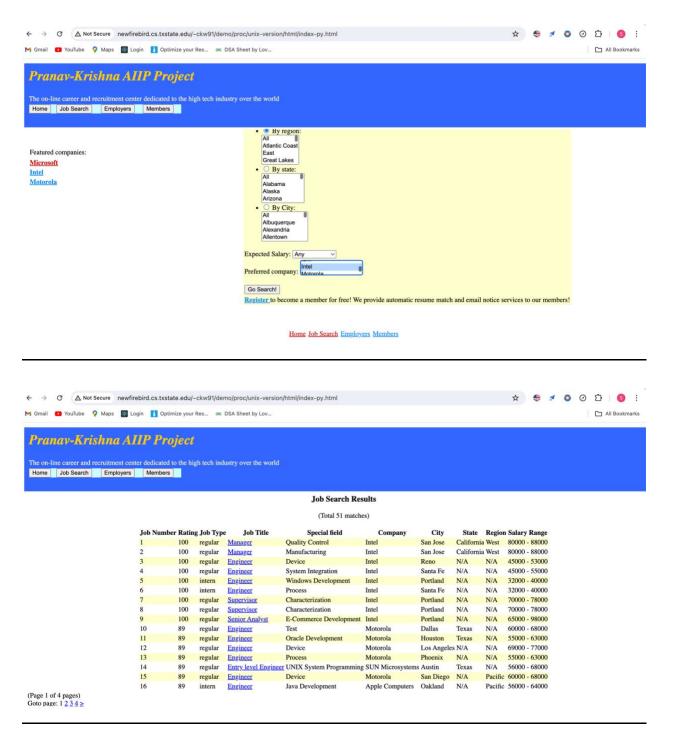
(Total 104 matches)

Job Number	Rating	Job Type	Job Title	Special field	Company	City	State	Region	Salary Range
97	90	regular	Staff Engineer	R&D	Apple Computers	N/A	N/A	Middle-west	56000 - 80000
98	90	regular	Engineer	Device Driver Development	GE	Allentown	N/A	N/A	50000 - 60000
99	90	regular	Senior Engineer	Device Driver Development	EDS	Little Rock	Arkansas	Middle	60000 - 70000
100	90	entry level	Analyst	Digital Design	Microsoft	Seattle	Washington	Northwest	60000 - 75000
101	90	regular	Staff Engineer	Embedded System	Apple Computers	N/A	N/A	Mountain	61000 - 61000
102	90	regular	Programmer Analyst	Marketing	Ford	Atlanta	N/A	Atlantic Coast	59000 - 79000
103	90	regular	Staff Engineer	Multimedia Application Development	Micron	N/A	N/A	Mountain	55000 - 61000
104	90	regular	Manager	Database Development	Micron	Des Moines	Iowa	Middle	45000 - 51000

(Page 7 of 7 pages) ≤Goto page: <u>1 2 3 4 5 6</u> 7 >

This is the end of the result.

<u>Company Field:</u> Here, we search for all the related companies of preferred company.





qstr=job_type=all&job_title=all&specialty=all&field_keyword=&location_type=region&min_salary=10000&company_name=intel====
Database version: 11.20.1.0 Beginning of building sql_query_stmt job_type=all&job_title=all&specialty=all&field_keyword=&location_type=region&min_salary=10000&company_name=intel ['job_type=all', 'job_title=all', 'specialty=all', 'field_keyword=', 'location_type=region', 'min_salary=10000,' company_name=intel']
self_stmt = SELECT * FROM job WHERE lower(company_name) = :1 OR lower(company_name) IN ('intel','texas instruments','amd','emc','ibm','csc','broadcom','microsoft','apple computers','sun microsystems','dell','national semiconductor', 'micron', 'samsung','nec', 'motorola')
Built statment = SELECT * FROM job WHERE lower(company_name) = :1 OR lower(company_name) IN ('intel','texas instruments','amd','emc','ibm','csc','broadcom','microsoft','apple computers','sun microsystems','dell','national semiconductor', 'micron', 'samsung','nec', 'motorola')
SELECT * FROM job WHERE lower(company_name) = :1 OR lower(company_name) IN ('intel','texas instruments','amd','emc','ibm','csc','broadcom','microsoft','apple computers','sun microsystems','dell','national semiconductor', 'micron','samsung','nec', 'motorola')

Job Search Results

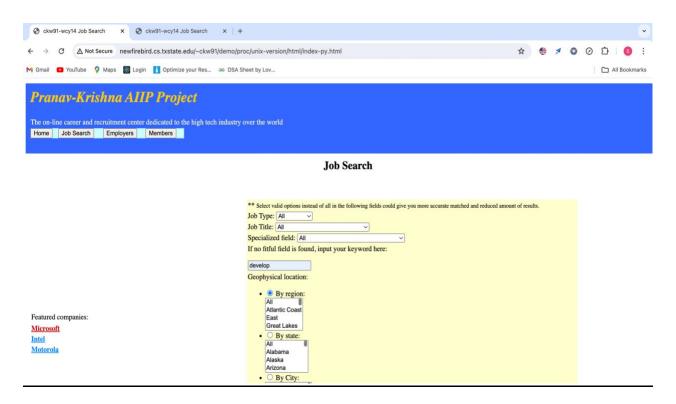
(Total 51 matches)

Job Nu	mber Rating	Job Type	Job Title		Special field	Company	City	State	Region	Salary	Range
49	89	regular	Staff Engineer	Embedded	System	Apple Computers	N/A	N/A	Mountain	61000	- 61000
50	89	regular	Staff Engineer	Multimedia	Application Development	Micron	N/A	N/A	Mountain	55000	- 61000
51	80	regular	Manager	Database D	evelopment	Micron	Des Moines	Lowe	Middle	45000	- 51000

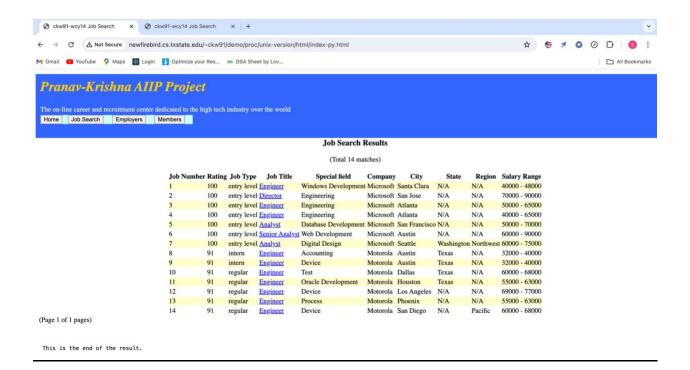
(Page 4 of 4 pages) ≤Goto page: 1 2 3 4 >

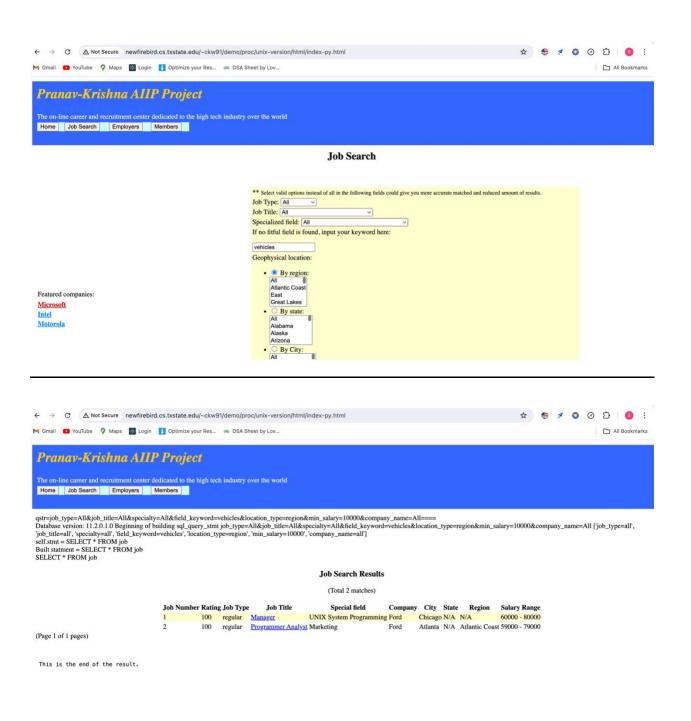
This is the end of the result.

KeyWord Search:

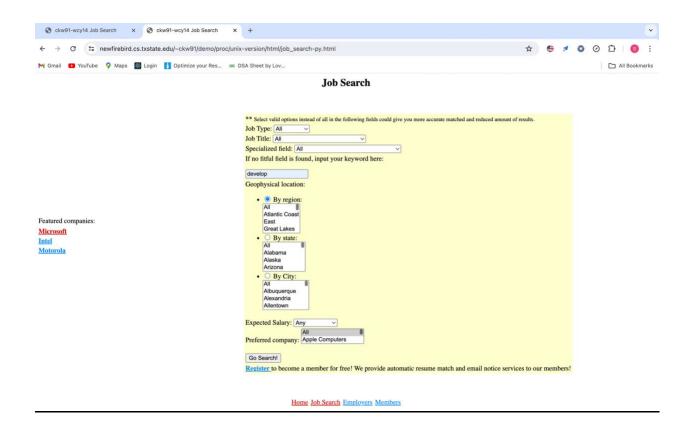


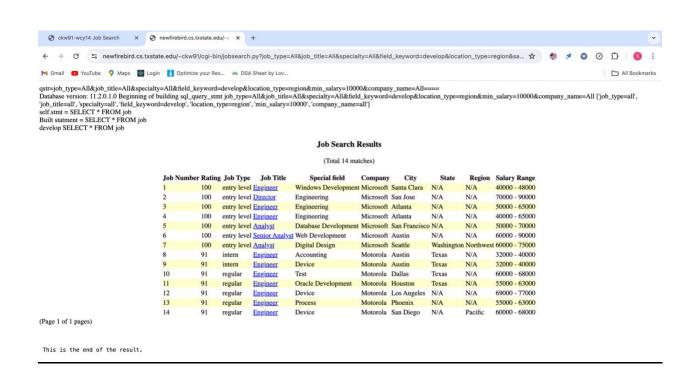
we updated the description column of company "Motorola" as "development of mobiles" to show the difference in jobratings between exact and partial match of keyword.





The word "vehicles" is part of description only for company ford. Hence, only ford records are displayed.





THANK YOU!