San Jose State University
SJSU ScholarWorks

Master's Projects   Master's Theses and Graduate Research

Spring 2014

OneCareer-A Visualization approach to job search process

Avinash Anantharamu
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_projects

Part of the Computer Sciences Commons

OneCareer-A Visualization approach to job search process

A Writing Project Report

Presented to

The Faculty of the Department of Computer Science

San Jose State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

By

Avinash Anantharamu

2014

ABSTRACT

One Career - A Visualization Approach to job search

Social media is a platform where job seekers share their portfolios. It enables them in actively networking with prospective employers and peers. Some of the popular social media sites like LinkedIn, Facebook, Pinterest, Google+, GitHub and Indeed encompass a wide array of features aiding recruiters and job-seekers. LinkedIn is the most popular among them. It makes use of the Connection Suggestion algorithm. However, It does not validate job seekers' profile data with existing online data sources. It neither offers skill and salary based insights nor visualizes job postings graphically. The focus of this paper is on building a viable platform which facilitates effective networking between job seeker's and the recruiters. Using the existing version control systems like GitHub, the veracity of jobseekers' mentioned data can be validated. This paper provides a detailed algorithmic approach to efficiently parse and visualize data through semi-automation.

SAN JOSÉ STATE UNIVERSITY

The Undersigned Writing Project Committee Approves the Writing Project Titled

One Career - A Visualization Approach to job search

By

Avinash Anantharamu

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE

Dr. Teng Moh	Department of Computer Science
Dr. Chris Pollett	Department of Computer Science
Dr. Ronald Mak	Department of Computer Science

Table of Contents

LIST OF FIGURES

Chapter 1

Introduction

LinkedIn is a widely accepted professional networking website. In LinkedIn, job seekers fill in and share information about academic institutions, professional experience and skills they possess. Job seekers build their professional network with recruiters thus enabling them to explore career opportunities with their preferred organizations.

Job seekers' skills and expertise are a large part of how they establish their identity. The LinkedIn automatic resume generator cannot recognize a job seeker's proficiency in any of the listed skills, fails to provide insight about variations in his/her skill based job listings grouped by region and their corresponding salaries. From a job seeker's pool of projects, the application cannot estimate his/her relative contribution towards each of them.

There is no professional network application that can fulfill all the requirements of every user. It lacks certain features like the skill graph, user contribution graph in team projects, geographical visualization of job listings, skill based salary and job trends.

The primary objective of this paper is to embed these features by extracting data from existing version control systems like GitHub, Perforce, Accurev and Tortoise SVN. This data is parsed to retrieve relevant information that is visualized to aid jobseekers and recruiters to arrive at more informed decisions.

Chapter 2

Literature Survey

Researchers have been applying various data mining techniques to existing version control systems for a wide of applications. The prominent among them being to view job seekers' coding samples, how job seeker documents his/her code , the number of times a particular piece of code was referenced, the problem type the job seeker prefers to solve, whether or not the job seeker a good squad player, etc.

Parsing data from the job seeker's existing version control systems to create his/her portfolio has been a topic of research for a long period. "Mining GitHub: Why Commit Stop - Exploring the Relationship between Developer's Commit Pattern and File Version Evolution," 2013 20th Asia-Pacific Software Engineering Conference by Yang Weicheng, Shen Beijing and Xu Ben [13] has been the frontrunner.

The paper is a study and analysis of developers' commit patterns in version control systems arrived at by inference from job seekers' commit behavior using dependency graphs, commit patterns and estimating project completion date based on the previously completed project timelines.

The results and inferences drawn from [13] serve as the focal point of this paper. The data utilized for analysis was compiled from eight random GitHub projects selected manually. A set of visualizations was performed on the different commit practices from the past projects. Some of the commit practices that were considered are the extent to which the code was modified, commit intervals between modifications, job seekers who made the modifications to an existing code base, commit behavior and dependencies of code modification within that repository.

The observations inferred that any minor change in the code would require a significant amount of modification in subsequent commit actions. For developers who commit frequently the average gap length between huge commits is three times the normal commits and for developers with a less frequent commit history, the commit pattern could not be arrived at and is treated as void after a sequence of commits. It indicates that while the application is being tested, the developers were waiting for feedback to fix any reported bugs before the next commit.

The prominent professional networking applications that were built using version control systems are:

My Git Hub Resume [16] is an application where a job seeker's data are extracted from his/her GitHub repositories. It generates a resume from existing repositories; list of skills based on projects that were worked on and maps them to a traditional A4 size resume.

The Git Inspector [17] is a version control based visualization application that visualizes information such as author name, the total number of commits, number of insertions, number of deletions and percentage of code that was modified by each author.

Glassdoor [19] is a crowdsourcing platform where job seekers gain insights about the organization's rating, salary information and questions based on the position title that were asked in previous interviews.

Indeed.com [3] is an employment-related Meta search engine for job listings. As a single-topic search engine, it employs vertical search for job listings. This site aggregates job listings from thousands of organizations' websites, including job boards, newspapers, associations, and company career pages.

These applications parse, filter and visualize a job seeker's existing work portfolio shared with version control systems. This paper focuses on a comparison of features available on

professional networking sites such as Glassdoor, Git-Inspector, Indeed.com, My Git Resume and LinkedIn with corresponding features that are built using OneCareer.

Chapter 3

Problem Formulation

Problem formulation is the creative and probably the most important step towards overcoming a problematic state than problem-solving. A good definition of what the problem is believed to be more than half of the way towards its eventual elimination.

This chapter explores the limitations of existing professional networking websites that facilitate communication between recruiters and job seekers.

1) Traditional resume is broken

OneCareer's Automated GitHub resume generator addresses the skill graph and hyperlinks to version control systems issues that are not addressed in traditional A-4 and infographic resume generation tools.

We are now in the digital age, yet when it comes to applying for jobs we still have to submit an archaic representation of ourselves, an A4 piece of paper. Our resume and cover letter is the only view a potential employer gets to see of us. So it has to be exceptional to stand out from the rest. Sadly, the traditional resume is broken. Its purpose has changed and the medium has evolved. We no longer need to limit ourselves to paper based resumes. Of late, there has been a trend of infographics and data visualization moving into resumes. It's definitely worth mentioning that there is a whole bunch of new online sites that capitalize on this growing trend. The service they offer is to create an automatic infographic resume that is usually based on your LinkedIn profile.

Applications generating infographic resume from LinkedIn have their own limitations:

- Job seeker's skill recommendation authenticity cannot be determined as it may or may not be endorsed by people who have worked with him.
- Access to the source code of projects listed on LinkedIn may or may not be available.
- There is no empirical method to classify the skills possessed by the job seeker in the hierarchical order of expertise.

The infographic resume automation process can be extended beyond LinkedIn by making use of existing version control systems like GitHub, tortoise SVN or Perforce to overcome the limitations mentioned above.

2) Did everyone equally contribute in a team project?

The OneCareer's GitHub repository analyzer provides a graphical representation to estimate the individual contribution in team projects that is not provided by existing professional networking sites.

It is widely acknowledged that group or team projects are a staple of graduate coursework. Such projects provide students with opportunities that better prepare them for their professional careers, so teamwork is often required. While there is a multitude of educational benefits of group projects, they can also pose considerable challenges in fairly and accurately discerning individual contribution for evaluation purposes.

Some of the metrics used to evaluate individual contribution are:-

- The group mark is everybody's mark,
- Everybody reports what they personally did,
- Peer grading system among group members,
- Cross-validating with the results of individual work

These metrics which are used by academia would not be available for the organizations. Organizations are always on the lookout for outstanding performers, which leaves them with limited options but to make a hiring decision based on the job seeker's resume and cover letter. Hence there is a need for a tool which helps recruiters visualize the job seeker's relative contribution in group projects to arrive at a more informed decision.

3) Jobs around me

OneCareer's Zip Jobs enables a geographical visualization of job postings for a specified zip code that is not available in the existing professional networking sites.

A job seeker would prefer to work for an organization within a 50 mile radius from where he/she resides. Existing job listing sites support zip code based job search within a specified boundary. However, a job seeker would not be able to determine the exact distance to a workplace within the zip code. Hence there is a need for a tool which aids the job seeker to exactly determine the distance to the workplace.

4) Which are the hot skills and how much are they paid in the market?

OneCareer's Jobs and Salary Trends application allow job seekers to get insight into postings based on skills along with region and skill-based salary trends in comparison with nationwide averages that are not available in existing professional networking sites which limit this information to each organization.

Whether you are about to graduate from college or are a working professional looking for a career change, choosing an appropriate domain can be a Herculean task. It's challenging to select a career path that matches up with your personal passions and still puts you at the top. Understanding some of the trends that shape the job market can help narrow your search. If you are looking for job security and growth, you will want to focus on the industries that are most likely to add the greatest number of new jobs in the coming decade.

How to research salary and do salary comparisons is one of the most frequent requests for information received by the Society for Human Resource Management (SHRM) [21]. This makes sense when you consider the importance of salary to place you in the right career path.

Hence there is a need for a tool which provides job seeker insights on salary and skill trends. Once job seekers are able to gain sufficient insight into trends, they would also like to

get insight into what would be the average pay for that particular skill in the region in comparison with nationwide trends so that they could strike a fair deal from their employers**.**

5) Crowdsourcing and localization jobs

OneCareer's Jobsource application enables job seekers to gain access to organizational and freelancing job postings within their city of residence and organizations to exploit the services of skilled people outside of the workforce to meet deadlines which are restricted to freelancing jobs in existing professional networking sites.

According to Wikipedia [1], crowdsourcing is the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people, and especially from an online community, rather than from traditional employees or suppliers. Localization is the process of adapting a product or service to a particular language, culture, and desired local "look-and-feel".

Crowdsourcing and localization are emerging as the new on-line distributed problem solving and production model in which networked people collaborate to complete a task. Enterprises are increasingly making use of crowdsourcing and localization to access a scalable workforce online in addition to their traditional workforce in order to meet tight project deadlines and overcome the language barrier across the world.

The success of a crowdsourcing job portal is dependent largely on the number of people whom you can deploy to contribute to your cause online. However, the outreach of this crowd sourcing endeavor is dependent largely on actively engaging skilled people across the globe to overcome the language barrier. By harnessing skills of jobseekers and freelancer effectively, companies can save on operational costs and meet shortened deadlines. Hence there is a need for a tool which achieves this.

Chapter 4

Solution

The current chapter provides a brief overview of the arrived solution to the problems formulated in the previous chapter.

4.1 GitHub Resume Generators

Writing a resume for every job listing with different skill sets is not only tedious, but it is also an inefficient use of available jobseeker time. The GitHub resume generator tool was built with the assumption that the jobseeker has a routine of sharing his/her work on version control systems like GitHub. GitHub Resume Generator creates an infographic resume populated with data ranging from the project title, description and hyperlink to the repository and estimating skills based on the current files in all repositories thus saving time and effort.

GitHub Resume Generator checks if the GitHub username entered by job seeker is valid or invalid. Upon validation it uses the GitHub REST API to fetch the user's avatar, email id, employer name, blog address, number of repositories, the total number of views, languages of each file present in all repositories. In case of an invalid entry it prompts the job seeker to re-enter the GitHub username.

Resume Generator initially sets the weight to four and it starts by checking if user repositories have at least ten forks up to the current date. If there are fewer than ten forks, then the weight reduces by one. Similarly the number of views, repository count and repository languages are retrieved and validated and weights are decremented in case they are under the threshold value.

Based on the weight, a hire or not to hire value is computed. Then we parse and keep count of each skill a job seeker has in his repository based on file extension and divide it by the total number of files in the repository to get an understanding of how well the job seeker has

fared in each skill set. Then we sort the job seekers' repositories by the number of forks and display the repository name, description and hyperlink for each repository.

Algorithm: GitHub Resume Generator
Input: GitHub User Name
Output: Generated Resume
1. Retrieve a job seeker's information, including profile image, first and last name, current location, email address, and hyperlink to his blog
2. Initialize the weight to 4.
3. If any of his repositories are forked
4. Check if the fork count is greater than 10
5. Decrement weight by 1.
6. If profile views are more than 1000
7. Decrement the weight by 1.
8. If the repository count is more than five.
9. Decrement the weight by 1.
10. If the user has files updated in at least one programming language.
11. Decrement the weight by 1.
12. Set the weight to -1.
13: If the weight is not zero then go to 14 else go to 15.
14. Hire <- True
15. Hire<- False
16. Count and keep track of all the file extensions in repository.
17. Sort the count in descending order.
18. Estimate the percentage of every skill using the formula (skill type count/total count)*100
19. Display repository name, description and hyperlink sorted by number of forks.

4.2 GitHub Repository Analyzer

Recruiters are always on the lookout for outstanding college graduates and are most likely to make less informed decisions in the process. This is primarily due to a lack of a tool to validate a candidate's contribution in projects, conformance to deadline and estimating project completion time based on previous revision history.

While hiring programmers, if recruiters have a tool which enables them to find the right pool of people in a region and then check how active are they in terms of highlighting their portfolio, then the margin of error is likely to decrease.

GitHub Repository Analyzer validates if job seekers' GitHub usernames and repository names are valid after which it stores the names of contributors in a hash map and assigns a unique color to each author. Then the commit dates of each user are stored in enumerators which are passed as arguments to D3 JS which in turn visualizes data. However, if either the GitHub username or repository name entry is invalid an error message is displayed on the screen.

Algorithm: GitHub Repository Analyzer
Input: GitHub username and Repository name
Output: Repository Visualization
1. Check if job seekers repository name and username are valid
2. Get Commit History of the Repository along with contributor's information
3. Screen out the commit dates based on authors commit history
4. Assign different colors to each individual author so no two authors have the same color
5. Compute and store number of days between each commit for each source in the depository
 6. Pass author names, color code and commit history as parameters to D3.Js for visual representation

4.3 Zip Jobs

Job seekers tend to look for jobs around the place where they live to reduce commute time to work and employers prefer people who are close to their workplace to ensure better productivity and availability. Zip Jobs uses Google Lat and Long API's to check if a jobseeker is in the U.S. and his zip code is passed as a parameter to the Indeed API to fetch job information in an XML DOM format to populate and represent available jobs information using pins on a Google map.

Algorithm: Zip Jobs
Input: Zip code of a place in USA
Output: Visualization of Job listings on a Google Map
1. Use Lat and Longitude to validate if the job seeker is from the USA, else exit.
2. Use the Indeed REST API to fetch all the jobs within the mentioned zip code.
3. Use the Google Maps API to populate the pins with job information.

4.4 Skill and Salary Trends

Skills and salary trends are based on JSON data returned from the Indeed REST API which are processed and grouped by skill. It is graphically represented using the D3JS library.

Algorithm: Skill Trend
Input: Job-Seeker click on the skill
Output: Visualization for that skill is displayed
1. Invoke the Indeed jobs trend REST call API with skill name as parameter.
2. The REST call returns data in JSON format which is visualized using D3JS.

Algorithm: Salary Trend
Input: Job Seeker click on the skill
Output: Visualization for that salary for skill is displayed
1. Invoke the Indeed salary trend estimator REST call with skills for which data is required along with position type as parameters.
2. The REST call returns data in JSON format which is visualized using D3JS.

4.5 Job Source

Crowdsourcing and localization enable employers to list jobs and reach active jobseekers and freelancers outside of their traditional workforce around the globe to meet tight deadlines and avoid the background checks and logistics involved in the hiring process. It enables organizations to save money and time. The algorithm is listed below.

Algorithm: Job Source
Actor: Recruiter
Input: Job Title, Job Description, Job Location and hyperlink
Output: Job posting successful or duplicate Job Posting
1. Check if any of mandatory fields are null.

2. Upon validation, insert the entries in a database and upon return, display alert messages
   Which indicate successful insert or prompt to repost.

Algorithm: Job Source
Actor: Job Seeker
Input: Job Title, City, Skill
Output: Job Listings
1. Check if any of the input parameters are null and constitute the query.
2. Query the MySQL database and display all the lookup results.

Chapter 5

Design and Implementation

The aim of this chapter is to elucidate features implemented in OneCareer. These features can be categorized into fully automated and semi automated features.

A fully automated feature requires minimal end-user interaction. GitHub resume generator, zip jobs, chartroom, Salary and job trends, coding challenges and job source fall under this category.

A semi automated feature requires considerable end-user interaction. GitHub repository analyzer falls under this category.

This chapter outlines architecture, implementation details and third party application programming interfaces used.

5.1 Architecture

OneCareer makes use of three tier architecture. It is a client–server architecture in which the user interface, business logic, computer data storage and data access are developed and maintained as independent modules. It is intended to allow any of the three tiers to be upgraded or replaced independently in response to future enhancements.

Each tier's functionality is described below:

- **Presentation tier:** This is the topmost level of the application. The presentation tier displays information related to such services as the exchange of messages chat room, job listing and search, browsing zip code based jobs, viewing the coding challenge, analyzing GitHub repository and viewing skill graph in a resume. It communicates with other tiers by which it puts out the results to the client tier and all other tiers in the network.
- **Business tier:** The logical tier is pulled out from the presentation tier and, as its own layer, it controls an application's functionality by performing detailed processing

- **Data tier:** The data tier includes the data persistence mechanisms (database servers, cloud servers) and the data access layer that encapsulates the persistence mechanisms and exposes the data. The data access layer should provide a channel to the application tier that exposes methods of managing the stored data without exposing or creating dependencies on the data storage mechanisms.
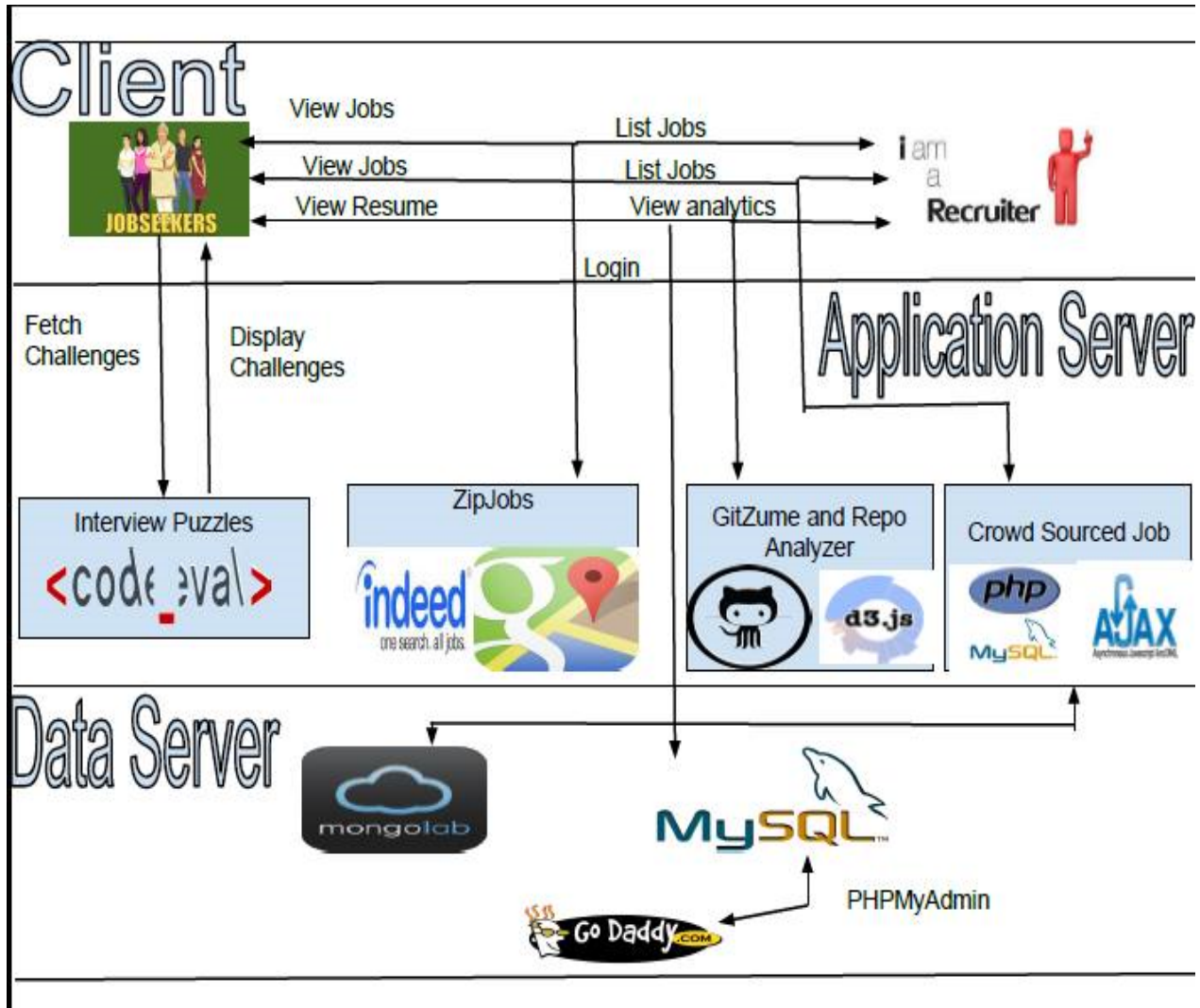

Figure 1: Architecture diagram of OneCareer

5.2 GitHub Resume Generators

OneCareer's GitHub Resume Generator feature is built for both recruiters and job seekers. From the job seekers point of view he/she registers with One Career with GitHub

username and selects type of job they are actively pursuing. From the recruiter point of view they need not have to store all the resumes, review them manually or parse for keywords using application and peek into each resume to figure out whether the job seeker is looking for an internship or full time position. This application saves job seeker time of writing resumes, recruiters are spared from the burden of storing and manually reading resumes. It is space efficient as resumes are generated on the fly and no disk space is utilized.

The figure below indicates the four stages involved in a GitHub Resume Generator application. In the input phase application accepts GitHub username. In parsing phase fetches all the information required for specified GitHub user and stores them in cache, In Visualization phase we retain required information and generate visualizations for the same and finally in display stage the visualized data is shared with the end user.
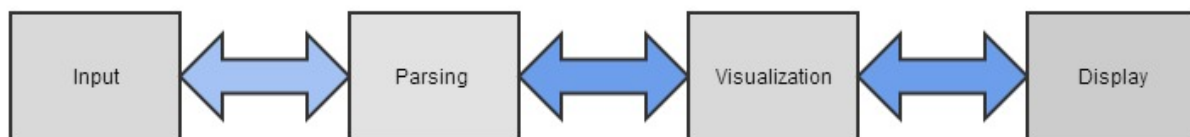
Figure 2: Overview of GitHub Resume Generator

5.3 GitHub Repository Analyzer

OneCareer's Analyzer is built for recruiters. This tool provides recruiters insights about job seeker coding habits, conformance to deadline, contribution in team projects apart from traditional data they have access to in the form of cover letter, resume and recommendation letters.

This application requires recruiters to login into OneCareer and manually enter GitHub user names from the available pool and the name of the GitHub repository following which they click on the Get Data button to view job seekers commit history timeline and share of contribution in team projects. The stacked option checkbox enables the overlap of each team

member's contribution thereby provides a broader insight into the relative contribution of each job seeker.

The figure below indicates the four stages involved in a GitHub Repository Analyzer. In the input phase application accepts GitHub username and repository name. In parsing phase fetches all the information required for specified GitHub users repository and stores them in Cache, In Visualization phase we retain required information and generate visualizations for the same and finally in display stage the visualized data is shared with the end user.
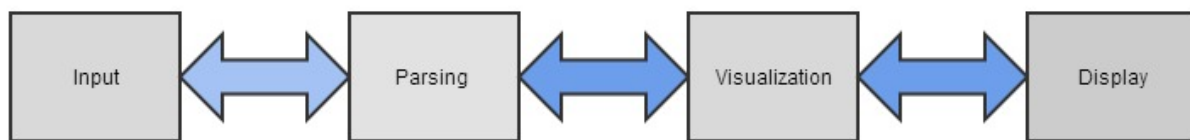

Figure 3: Overview of GitHub Repository Analyzer

5.4 Zip Jobs

OneCareer's Zip Jobs is a tool built for job seekers'. Zip Jobs accepts a zip code from a job seeker and validates if it is a valid zip code using Smarty Streets zip code API. Upon validation it invokes Indeed's REST API with zip code to obtain the active job listings in XML format. This XML data is visually pinned on Google Maps using Google Map's API. The default radius is 50 miles around the entered zip code. Job Maps also support zoom in, zoom out and satellite view options available in Google maps.

The figure below indicates the four stages involved in a Zip Jobs. In the input phase application accepts zip code from job seeker. In fetching phase it retrieves all the job listings with the specified zip code in XML format. In Visualization phase the fetched data is passed to Google maps and in display stage the pinned jobs information is shared with the end user.

17
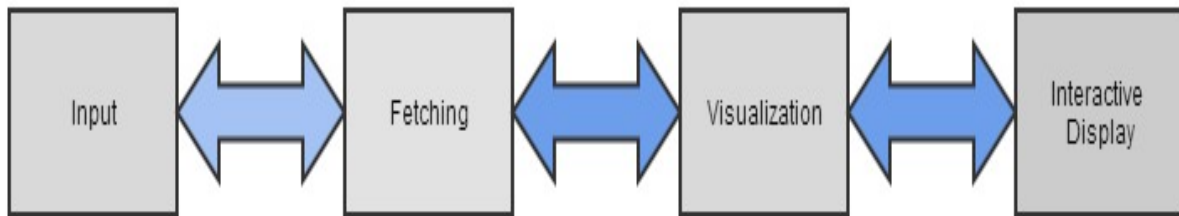
Figure 4: Overview of Zip Jobs

5.5 Skill Trends and Salary Trends

OneCareer's skills and salary trends are tools built for job seekers. Skill Trends graphically represents the variation of percentage of matching job postings over a timeline for entry level software engineers. It takes into account nationwide data and is represented using a series of line graphs.

Currently the application supports HTML-5, jQuery, Android, iOS, Hadoop, Mobile Application, MongoDB and Social Media. It is build using the data provided by Indeed.com REST API and passed to the Data Driven Document JS library for graphical representation.

Salary Trends application graphically represents the mean average of salaries for entry level software engineers and Interns for a range of skills. The average salaries of California state are compared against a nationwide average using a series of bar and line graphs. Currently the application supports Java, .Net, HTML-5, Android, iOS, Big Data, C++, Perl, Python, Hardware Engineer, QA Engineer and C Programmers. Salary Trends are built using the data obtained

from Indeed.com REST API and passed to the Data Driven Documents JS library for graphical representation

The figure below indicates the three stages involved in a Skill and Salary Trends. In the input phase, the application job seeker clicks on the desired skill. In the REST API phase it retrieves skill based job listings and salary based on skill data over a period. In the Analytics phase the fetched data is represented as graphs using D3.JS.

Figure 5: Skill and Salary Trends Overview

5.6 Job Source

OneCareer's job source application is built for both job seekers and recruiters. Job source enables organizations to exploit crowdsourcing and localization to reach skilled workforce, reduce hiring risks and meet project deadlines.

The figure below indicates the three stages involved in job source. In the input phase recruiters list jobs using title, skills required location and link to the job posting. In REST API phase this data is stored in cloud mongo lab database. In Display phase the confirmation would be sent to the recruiter.

Figure 6: Job Source Recruiter Overview

The figure below indicates the three stages involved in job source. In the input phase job seekers search criteria for jobs using title, skills required. In REST API phase this query is performed on mongo lab. In Display phase search results are displayed.

Figure 7: Job Source Job Seeker Overview

Chapter 6

Experiments

A comparative study of OneCareer's features with equivalent features in existing applications is being carried out.

The GitHub resume generator was compared with a tool titled My GITHUB RESUME which, was intended to a generate resume from GitHub repositories. However, My GITHUB-RESUME failed to generate resume even upon manually entering a different set of valid GitHub usernames.

GitHub Resume Analyzer is compared with Maxim Chernyak [16]. The comparison of features between both the applications is represented in a table:

Feature Name	GitHub Resume	Maxim Chernyak
User Information	Username City email id Employer blog address	Username LinkedIn profile GitHub Link Twitter Blog Email Phone
Skills	Yes with skill estimation	Skill set list
Job Experience	No	Yes
Project with hyperlinks	Yes	No
Followers	Yes	No
Repository Count	Yes	No
Dynamic	Yes	No

The GitHub repository analyzer's features were compared to an application titled GitInspector [17]. The comparison of features between both the applications is represented in a table:-

Feature	GitHub Repository Analyzer	GitInspector
Commit Date	Yes	No
Type of Commits	No	Yes
View others GitHub	Yes	No
Stacked Views	Yes	No
Export Data	No	Yes

Zip Jobs was compared with indeed zip code based job search. The comparison of features

between both the applications is represented in a table:

Feature	Indeed	ZIP JOBS
Zip code search	Yes	Yes
Visualization on Maps	No	Yes
Custom Queries	Yes	No

Comparison of LinkedIn, Indeed and OneCareer based on features is represented in a table:

Feature	LinkedIn	Indeed	OneCareer
Graphical zip code job listing	No	No	Yes
Automating Resume Generation	Yes	No	Yes
Analyzing Repositories	No	No	Yes
Salary Trends	No	Yes	Yes

From the above observations, certain features are available in different standalone

applications which are incorporated in a single standalone application. OneCareer effectively

generates resume for any valid GitHub user while some stand alone applications do not offer

certain features mentioned in the table. GitHub repositories analyzed by Gitinspector do not

provide dates of commit or pull requests for individual users. Hence as project gets older it

becomes complicated to keep track of the user updates. Whereas, in OneCareer is performed on

individual repository thereby handling complexity. Several job listings sites like Indeed, Monster

support zip code based job search. However, none of them provides graphical visualization for

filtered job listings.

Chapter-7

Conclusion

OneCareer has automated the process of info graphic resume generation with skill list. It represents visualization of job seeker skill set along with profile image, current work location and project information with hyperlinks. Thereby, the drawbacks of limited information and dependence on LinkedIn as the single source of data are addressed.

GitHub repository Analyzer has successfully aided recruiters to gain insights about job seekers relative contribution in team projects. Thereby, it enables them to arrive at more informed decision.

Job room   provides a platform for job seekers to interact among peers anonymously thereby; ensuring exchange of information takes place anonymously.

Zip Jobs enables job seekers to perform geographical visualizations of job listings by zip code. This feature assists the job seeker to arrive at more informed choice with commuting as criteria.

Skill trends enable job seekers to periodically gain insights on the number of job postings based on skill.

Job trends enable job seekers to gain insights into skill based salary trends in that region along with nationwide averages.

From the above observations, it can be inferred that jobseekers and recruiters would gain additional insights from features provided in OneCareer in addition to existing professional networking websites.

Chapter 8

Future Works

The primary objective of this paper was to build a platform that facilitates effective interaction among job seekers and recruiters. However, the application does not bestow the job seekers the option of selecting the version control platform of their choice. Job seekers need not use GitHub as a version control system.

Job seekers may prefer to hide or edit certain fields generated by GitHub resume generator which is currently not supported by the application.

The GitHub repository analyzer cannot retrieve repository names of specified GitHub user and requires recruiter to manually enter the data. This approach is cumbersome and needs to be automated.

The current system requires jobseeker to search for the jobs manually either by entering zip code or by filling by custom search criteria, further research can ensure data proceed from GitHub resume be used for auto suggesting jobs to achieve better skill to job mapping.

The existing job search based on job listings retrieved from Indeed, GitHub and LinkedIn jobs. However, API's do not support job listings for all geographies.

The salary trends data obtained from Indeed's API. It is not able to categorize data based on university name and job title. The data obtained from survey tools such as Google Forms or Survey Monkey could be made use of build salary information based on job title and university.

APPENDIX A: KEY FEATURES

Third party tools have been used in developing and testing this application such as Indeed API, GitHub API, Google Maps API and LinkedIn API whose detailed description is provided below:-

Indeed

Indeed is the world's number one job site, with over 140 million unique visitors per month. Indeed is currently available in more than 50 countries [3]. OneCareer incorporates an Indeed's REST API to extract job listings, skill based salary information and skill based demand over a time line [3].

The REST API offered by Indeed is described in the table below:-

Following is HTTP-REST service: http://www.Indeed.com
Return Value: It returns the search results of indeed in XML or JSON format
Syntax of Indeed Restful service: http://api.indeed.com/ads/apisearch?publisher="publisherid"&q="query1&query2" &OptionalParameters
A user with publisher key can request the API service 30 times a day and after than the API is a paid service and user has to either pay up front or need to enter credit card details for further access to API within the day else the API does not return any data

The optional parameters for the Indeed's REST API are described in the table below:

Optional Parameter representation	Description / Possible Values
Format	X.M.L or J.S.O.N
L	Location (zip code, City, State, Country)
Sort	Relevance or Date (Default value is relevance)
Radius	Radius from reference location (Default value is 25 miles)
St	Site Type (job board, employer) (Default value is employer)
Limit	Number of return search values (Default is 10)

Latlong	Latitude and longitude

GitHub

GitHub is the best place to share code with friends, co-workers, classmates, and complete strangers. Over four million people use GitHub to build amazing things together [6]. OneCareer incorporates GitHub's REST API to mine job seekers repositories to extract information about his Full name, email id, current location, current employer, display image, repositories name and description, number of forks for each repositories and total number of followers following job seeker.

The REST API offered by GITHUB is described in the table below:-

Following is HTTP-REST service: http://www. github.com
Return Value: It returns the user information ranging from number of repositories, number of followers, code repositories description, URL if any present, current city, employer name, number of files uploaded in each repository with format of the file uploaded in XML or JSON format
Syntax of Indeed Restful service: https://api.github.com/users/&OptionalParamters

The optional parameters of the application are described in the table below:-

Optional Parameter representation	Description / Possible Values
Format	X.M.L or J.S.O.N
Username	GitHub Username
Repos	Valid username
Avatar	Image Size
Name	Valid username
Location	Valid username
Email	Valid username

Company	Valid username
Blog	Valid username
Followers	Valid username
Language	Valid repository name

LinkedIn

LinkedIn operates the world's largest professional network on the Internet with more than 300 million members in over 200 countries and territories [20]. OneCareer incorporates LinkedIn's REST API to auto suggest jobs to the job seeker located in California. To enable this feature in the web application, it is mandatory to have a registered LinkedIn developers account coupled with registering the domain of the website.

The REST API offered by LinkedIn Developers is described in the table below:

Following is HTTP-REST service: http://www. developer.LinkedIn.com/apis
Return Value: It returns the search results of indeed in XML or JSON format
Syntax of LinkedIn Restful service is   http://api.linkedin.com/v1/jobs/%7Bid%7D:(id,customer-job-code,active,posting-date,expiration-date,posting-timestamp,company:(id,name),position:(title,location,job-functions,industries,job-type,experience-level),skills-and-experience,description-snippet,description,salary,job-poster:(id,first-name,last-name,headline),referral-bonus,site-job-url,location)

The optional parameters of the application are explained in the table listed below

Optional Parameter representation	Description / Possible Values
Format	X.M.L or J.S.O.N
Active	Indicates whether the job is active or not
Posting-time stamp	Time is milliseconds
Skills and experience	Skill type and Experience valid parameters can be Internship, Entry level, Associate, Mid Senior Level, Executive, Director or Not Applicable

Salary	The salary listed for posting jobs
Location	Zip code, City, State or Country

Google Maps

Google Maps is a desktop and mobile web mapping service application and technology provided by Google, offering satellite imagery, street maps, and Street View perspectives, as well as functions such as a route planner for traveling by foot, car, bicycle or with public transportation. Also supported are maps embedded on third-party websites via the Google Maps API [6], OneCareer incorporates Google maps API to visually represent job postings from indeed based on zip code.

The REST API offered by Google Developers is described in the table below:-

Following is HTTP-REST service: http://www.developers.google.com/maps
Syntax of Google Maps API   https://maps.googleapis.com/maps/api="api-key"/js?sensor="value"&optional-paramters

The optional parameters of the application are explained in table listed below

Optional Parameter representation	Description / Possible Values
API key	Publisher Key Id
Sensor	True or False (Default is False)
Name	The name of this map. This name will identify this map to users. It must be unique across all maps and layers.
Description	A text description for this map. This is an important field, because the text is shown to the end user in some clients, such as Google Earth clients. Descriptions can use standard HTML hyperlink syntax (*linkname*). Descriptions can use to insert line breaks. Additional HTML tags,

	such as image source, are not supported in descriptions.
Tags	The text strings that should return this map in response to a search. Separate multiple values with commas.
Shared with	The access list whose members can work with this map in your authoring environment.
Viewable by:	The access lists whose members can view the published map. This list can include external viewers, in addition to people in your organization.
Layer names	By default, the layer name is shown to users, so it's useful to replace the system-generated layer names with user-friendly names. For example, you could create display names such as "Most populous cities," "Fueling stations," "Water depth of 100m," "Endangered wildlife," "Recommended hotels," and so on.

Google Custom Search Engine

Google Custom Search is a platform provided by Google that allows web developers to feature specialized information in web searches, refine and categorize queries and create customized search engines, based on Google Search's Page Rank algorithm [6]. OneCareer incorporates Google Custom Search Engine API to enable both job seekers and recruiters are able to perform custom searches.

The REST API offered by Google Developers is described in the table below:-

Following is HTTP-REST service: http://www.developers.google.com/cse
Syntax of Google Custom Search API http://www.google.com/cse/api/<user_id>/cse/<cse_id>

APPENDIX B: System Requirements

To solve the problems formulated in the previous chapter we have identified functional, non functional requirements, scenarios and sub system decomposition.

Functional Requirements

Functional Requirements are used to describe the interactions between a system and environment independent of its implementation. The environment includes the user and any other external system with which the system interacts.

Functional Requirements for OneCareer are:

1. **GitHub Resume Generator**: Extracting the data from job seekers version control systems, processing it and representing it visually is the prime features missing in existing resume parsers. The following issues can be addressed using:

   - A set of optimized algorithms to efficiently process the data.
   - A set of visualization techniques and libraries to represent the processed data.

2. **GitHub Repository Analyzer**: Traditionally recruiters would short list job-seekers using their resumes and cover letters, which would give them bird's eye view about their skills. However, recruiters do not get additional insight into job seekers coding habits. This is efficiently addressed by GitHub Repository Analyzer using:

   - Job seeker's commit pattern over a series of repositories.
   - Job seeker's contribution in team projects.

3. **Zip Jobs:** Existing job listing websites support, zip code based job search without geographical visualization. Zip Jobs are a useful tool which addresses this limitation using:

   - Crawl active jobs from job listing websites.

- Pinning the results on Google maps.

4. **Salary Trends:** In existing job listing sites like Glassdoor they provide estimates of salary based on the position title and organization. However, they do not provide salary trends based on region and skill. OneCareer solves this issue with:

   - Fetching salary information based on skills and regions.
   - Visualizing then using D3 Js

5. **Skill Trends:** In existing job listing sites they do provide trends based on skill. OneCareer addresses this issue using:

   - Fetching number of skills based job openings.
   - Visualizing then using D3 Js

6. **Jobsource:** In existing job listing sites use of Crowdsourcing and localization is limited to freelancer job postings. However, this trend is catching up in organizations OneCareer attempts help organizations exploit this trend using:

   - Locality based job search

Non Functional Requirements

Non Functional requirements describe aspects of the system that are not directly related to the functional behavior of the system. Non Functional requirements include a broad variety of requirements that apply to many different aspects of the system, from usability to performance.

GitHub Resume Generator:

Reliability

- **Fault tolerance:** The application will not crash under any instance.
- **Recoverability:** In case of a crash due any unavoidable circumstances. The application should be up and running within 10 seconds of downtime.

Performance

- **Response time:** Upon job seeker's entering input the resume will be generated within 4 seconds for 99% of the cases.

- **Scalability:** Application can support up to 50000 job seekers at any given instance.

- **Availability:** The application is designed to be available 24/7/365.

Operations Requirements

The application is compatible with Web browsers such as Safari, Internet Explorer, Firefox and Chrome.

GitHub Repository Analyzer:

Reliability

- **Fault tolerance:** The application will not crash under any instance.

- **Recoverability:** In case of a crash due any unavoidable circumstances. The application should be up and running within 10 seconds of downtime.

Performance

- **Response time:** Upon recruiter entering input result will be returned within 4 seconds for 99% of cases.

- **Scalability:** Application can support up to 5000 GitHub repositories at any given instance.

- **Availability:** The application is designed to be available 24/7/365.

Operations Requirements

The application is compatible with Web browsers such as Safari, Internet Explorer, Firefox and Chrome.

ZIP JOBS:

Reliability

- **Fault tolerance:** The application will not crash under any instance.

- **Recoverability:** In case of a crash due any unavoidable circumstances. The application should be up and running within 10 seconds of downtime.

Performance

- **Response time:** Upon job seekers entering input results will be returned within 6 seconds for 99% of cases.
- **Scalability:** Application can support up to 4000 job seekers at any given instance.
- **Availability:** The application is designed to be available 24/7/365.

Operations Requirements

The application is compatible with Web browsers such as Safari, Internet Explorer, Firefox and Chrome.

Salary Trends:

Reliability

- **Fault tolerance:** The application will not crash under any instance.
- **Recoverability:** In case of a crash due any unavoidable circumstances. The application should be up and running within 10 seconds of downtime.

Performance

- **Response time:** Upon job-seeker clicking on input results will be returned within 1 second for 99% of cases.
- **Scalability:** Application can support up to 15000 job seekers at any given instance.
- **Availability:** The application is designed to be available 24/7/365.

Operations Requirements

The application is compatible with Web browsers such as Safari, Internet Explorer, Firefox and Chrome.

Scenario

A scenario is an instance of, use case describing a concrete set of action.

Scenario ID	1
Scenario Title	Job Seeker Registration

Description	Job Seeker Registration with One Career
User or Actor	Job Seeker
Entry Condition	Job Seeker has a GitHub profile.
Exit Condition	Job Seeker has successfully registered
Exception	For Duplicate credential system will prompt the job seeker to re register
<<Include>>	NA
<<Extend>>	GitHub
Main Flow	→ Job Seeker populates the registration fields and clicks register ← The system displays an alert message registration successfully
Special requirements	Job Seeker should have a valid GitHub account

Scenario ID	2
Scenario Title	Job Seeker Login
Description	Job Seeker logs into with OneCareer to view his GitHub Resume Generator
User or Actor	Job Seeker
Entry Condition	Job Seeker has a GitHub profile and job seeker has registered with OneCareer.
Exit Condition	Job seeker has successfully viewed his/her resume
Exception	For Invalid credential system will prompt to job seeker to again authenticate his credentials
<<Include>>	NA
<<Extend>>	MySql
Main Flow	→ Job seeker populates the login name and password and clicks login ← The system displays an alert message login successfully
Special requirements	Job seeker should have a valid GitHub account and registered OneCareer account

Scenario ID	3
Scenario Title	Job Seekers Salary Trends
Description	Job Seeker views what his peers are earning across his state and country
User or Actor	Job Seeker
Entry Condition	Job Seeker has a technology background.

Exit Condition	Job Seeker has successfully viewed salary trends
Exception	NA
<<Include>>	NA
<<Extend>>	Indeed
Main Flow	→ Job Seeker clicks on the skill for which he wants to view salary trends ← The system displays salary trends to Job Seeker
Special requirements	NA

Scenario ID	4
Scenario Title	Job Seekers Skill Trends
Description	Job Seeker viewers what how are demand for skills, changing over the decade
User or Actor	Job Seeker
Entry Condition	Job Seeker has a technology background.
Exit Condition	Job Seeker has successfully viewed skill trends
Exception	NA
<<Include>>	NA
<<Extend>>	Indeed
Main Flow	→ Job Seeker clicks on the skill for which he wants to view salary trends ← The system displays skill trends to Job Seeker
Special requirements	NA

Scenario ID	5
Scenario Title	Job Seekers Job listing based on Zip code
Description	Job Seeker views the active job listings in the zip code of his choice on a Google map
User or Actor	Job Seeker
Entry Condition	Job Seeker is actively looking for a job.
Exit Condition	Job Seeker has viewed all the listed jobs
Exception	For Invalid zip code application will not display any data
<<Include>>	NA

<<Extend>>	Indeed

Main Flow	→ Job Seeker enters zip code where they want a job
	               ← The system displays all open positions on a
	                   Google map

Special requirements	Indeed API has to support the country in which the job seeker is looking for jobs

Scenario ID	6
Scenario Title	Recruiter Registration
Description	Recruiter registration with OneCareer
User or Actor	Recruiter
Entry Condition	The recruiter must be looking to hire talent.
Exit Condition	The recruiter has successfully registered
Exception	For Duplicate credentials system will prompt the recruiter to reregister
<<Include>>	NA
<<Extend>>	MySql

Main Flow	→ Recruiter populates the registration fields and clicks register
	               ← System displays a alert message registration
	                   successful

Special requirements	N/A

Scenario ID	7
Scenario Title	Recruiter Login
Description	Recruiter logs into OneCareer
User or Actor	Recruiter
Entry Condition	Recruiter should have registered with OneCareer.
Exit Condition	Recruiter has logged into the application
Exception	For Invalid credentials system will prompt the recruiter to again authenticate his credentials
<<Include>>	NA
<<Extend>>	MySql
Main Flow	→ Recruiter enters their login name and password and clicks login ← System displays a alert message login successful
Special	Recruiter should have Registered OneCareer account

requirements	

Scenario ID	8
Scenario Title	Recruiter's GitHub Repository Analyzer
Description	Recruiter logs into GitHub Repository Analyzer in OneCareer
User or Actor	Recruiter
Entry Condition	Recruiter should have job-seekers GitHub id and repository name who contribution they want to evaluate.
Exit Condition	Recruiter is able to view job seekers contribution and commit history.
Exception	For Invalid GitHub username or repository names It will display a not found error
<<Include>>	NA
<<Extend>>	GitHub
Main Flow	→ Recruiter enters their job-seekers GitHub username and repository and clicks Get Data button ← System displays a job-seekers contribution in comparison with his team-mates
Special requirements	Recruiter should have list of GitHub username and their repository names

Scenario ID	9
Scenario Title	Job Seekers GitHub Resume Generator
Description	Job Seekers views his/her resume in OneCareer
User or Actor	Job Seeker
Entry Condition	Job Seeker should have a valid GitHub id and should have registered with one career
Exit Condition	Job Seeker has viewed his resume generated from GitHub.
Exception	N/A
<<Include>>	NA
<<Extend>>	GitHub
Main Flow	→ Job seekers enters their GitHub username clicks get login ← System displays a resume generated out of GitHub
Special requirements	N/A

Sub-System Decomposition Diagram

The large software system is broken down into smaller system to identify functionality of each sub system this process is termed as subsystem decomposition.
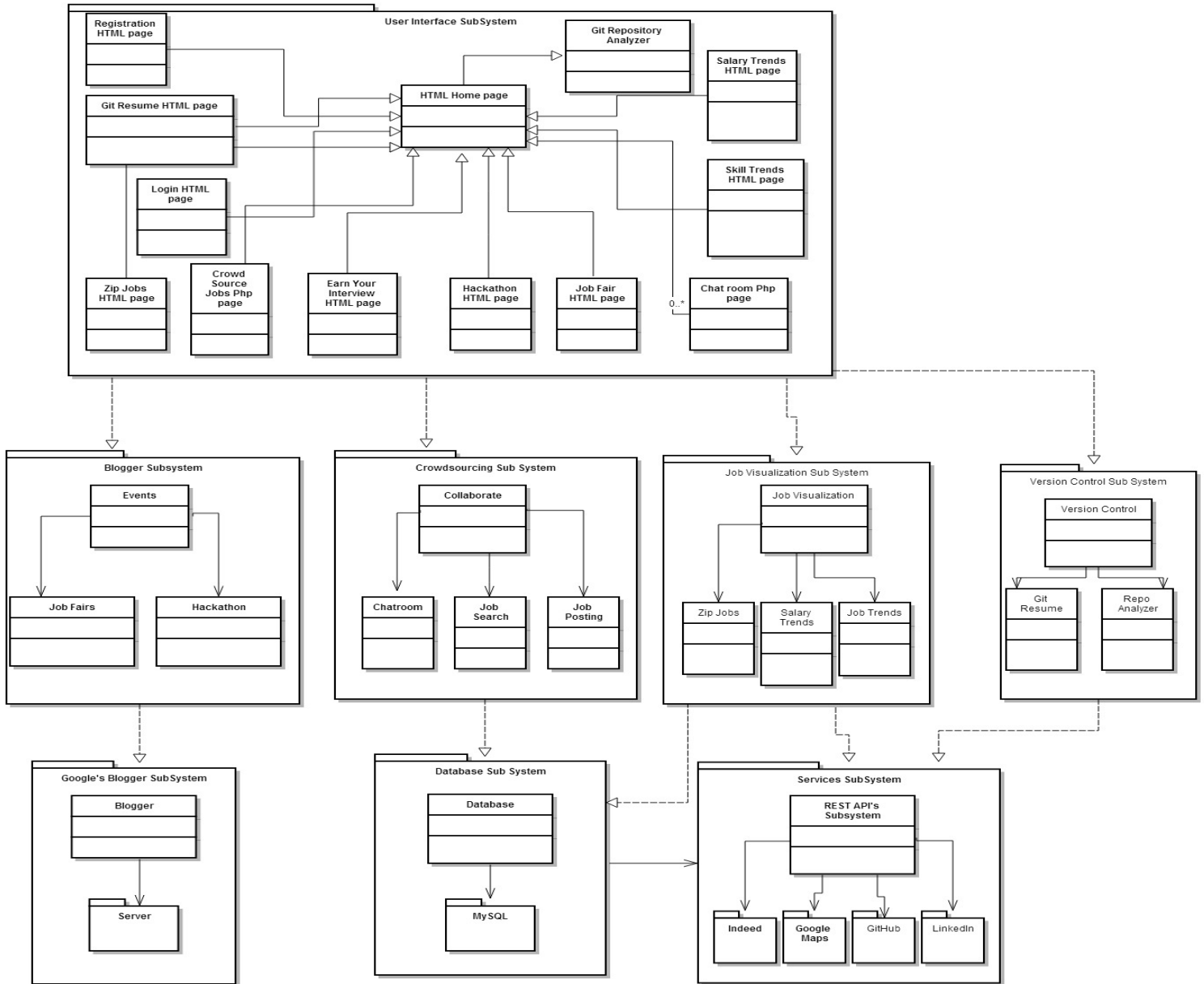

Figure 8: OneCareer's Sub System Decomposition

LIST OF REFERENCES

1. Wikipedia. (14th April 2014) Crowdsourcing, Available at: http://www.en.wikipedia.org/wiki/Crowdsourcing (Accessed: 25th April 2014).

2. Michael,W. and Patrick,S. (20th January 2014) Candy-a JavaScript-based multi-user chat client, Available at: http://www.candy-chat.github.io/candy/ (Accessed: 25th April 2014).

3. Indeed.XML Feed (2014) Indeed Publisher, Available at: https://www.ads.indeed.com/jobroll/ xmlfeed (Accessed: 25th April 2014).

4. LinkedIn. (2014) Job Look Up API's and Fields, Available at: https://www.developer.linkedin.com/documents/job-lookup-api-and-fields (Accessed: 25th April 2014).

5. Google,Inc. (2014) Google Developers, Available at: https://www.developers.google.com/maps/ (Accessed: 25th April 2014).

6. GitHub,Developers. (2014) GitHub Developer, Available at: https://www.developer.github.com/v3/ (Accessed: 25th April 2014).

7. Survey,Monkey. (2014) Survey Monkey, Available at: https://www.surveymonkey.com/ (Accessed: 25th April 2014).

8. Barchilon MG. Technology's impact on online resumes. 1998; 2: 183-187 o.2. 9. Wu B, Pei X, Tan J, and Wang Y. Resume mining of communities in social network. 2007; 435-440.

9. Krippendorff (24th April 2014) *PROBLEM-FORMULATION,* Available at:*http://pespmc1.vub.ac.be/ASC/PROBLEM-FOR.html* (Accessed: 24th April 2014).

10. Wikimedia Foundation, Inc (19th April 2014) *Localization,* Available at:*http://en.wikipedia.org/wiki/Localization* (Accessed: 24th April 2014).

11. Kopparapu SK. Automatic extraction of usable information from unstructured resumes to aid search. 2010; 1: 99-103.

12. Jiang ZX, Zhang C, Xiao B, and Lin Z. Research and implementation of intelligent chinese resume parsing. 2009; 3: 588-593.

13. Weicheng Y, Beijun S, and Ben X. Mining GitHub: Why commit stops -- exploring the relationship between developer's commit pattern and file version evolution. 2013; 165-169.

14. Daniel,B. (14th July 2011) Github is Your New Resume, Available at: http://www.code.dblock.org/github-is-your-new-resume (Accessed: 24th April 2014).

15. Gliffy. (2014) Online Diagram Software , Available at: http://www.gliffy.com/ (Accessed: 24th April 2014).

16. David,C. (2014) GitHub resume, Available at: http://www.resume.github.io/ (Accessed: 24th April 2014).

17. Git,Inspector. (2014) html_example.jpg, Available at: http://www.wiki.gitinspector.googlecode.com/git/images/html_example.jpg (Accessed: 24th April 2014).

18. Bernd Bruegge and Allen H. Dutoit. 2009. *Object-Oriented Software Engineering Using Uml, Patterns, and Java* (3rd ed.). Prentice Hall Press, Upper Saddle River, NJ, USA.

19. Glassdoor. (2014) Company Salaries | Glassdoor, Available at: http://www.glassdoor.com/Salaries/index.htm (Accessed: 24th April 2014).

20. LinkedIn. (2014) LinkedIn-About, Available at: http://www.press.linkedin.com/about (Accessed: 24th April 2014).

21. Mike, Bostock. (24th April 2014) *Data-Driven Documents,* Available at:*http://www.d3js.org/* (Accessed: 24th April 2014).

22. Mike, Bostock. (30th April 2014) *Application programming interface,* Available at:*http://en.wikipedia.org/wiki/Application_programming_interface* (Accessed: 30th April 2014).