PROJECT REPORT-TEAM 11

CARE: Career Analytics Recommendation Engine

Shriram Chandra Anbalagan
Kowsik Krishnan
Parama Dutta
Janakiram Pulipati

MAY 12, 2014
MIS ELLER
**Introduction - high level overview**

Forbes says that more than 18 million job opportunities are posted across Twitter, Facebook and LinkedIn. Apart from jobs opportunities the social media platform is a great source for latest news and updates about various technologies, organizations, career and interview tips submitted by various recruiters and experts. None of the current job portals are leveraging this enormous information to present it to the job seekers and help them prepare to get their dream career. Also a sense of what the employees feel about their employers is missing in most of the job search portals. This information shall give the job seekers a glimpse into the organizations work environment and help decide them to choose the right company for the job they seek.

**System objectives**

- **Market / User needs**
  There is still a gap in the present market between the number of jobs posted and number of jobs reached to the appropriate candidates. This is mainly due to the diversity of information across several social media platforms. Also the user faces hard time navigating through many Facebook pages, tweets and LinkedIn job postings. Current market needs a website which brings the jobs posted across various social media sites such as twitter, Facebook and LinkedIn, under a single roof and make user experience easy.

- **System / functionalities**
  Upon user request, the system searches for the job postings across various social media sites and presents them to the user along with some add-on analytics features such as recommending users with upcoming technologies, hottest companies and tailor-made jobs.

- **Target Audience**
  - Job Seekers
  - Students
  - Corporate professionals
  - Corporate HRs

**Market landscape**

**Competitors: functionalities / features they provide / market presence and who are the major players**

Our major competitors would be well established job portals such as monster.com, dice.com and indeed.com etc. Each of these websites provides basic job search functionalities along with features such as hot jobs, discussion forums, and interview tips etc. The hot jobs prediction by monster.com is done based upon the location. The analytics component in the above mentioned websites is not so significant to provide advanced functionalities/features to the users.
Competitor Analysis

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Monster</th>
<th>LinkedIn</th>
<th>Glassdoor</th>
<th>Indeed</th>
<th>CARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Media Job Search</td>
<td>❌</td>
<td>✔️</td>
<td>❌</td>
<td>❌</td>
<td>✔️</td>
</tr>
<tr>
<td>Government Jobs</td>
<td>✔️</td>
<td>❌</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Candidate Reputation Score</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>✔️</td>
</tr>
<tr>
<td>Video tutorials</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>✔️</td>
</tr>
<tr>
<td>Company Analytics</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>✔️</td>
</tr>
<tr>
<td>Career Analytics</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

As can be seen from the table above, our website goes above and beyond the conventional criteria offered by most websites and provides in depth analytical features as well as leveraging job and career information from social websites like Facebook, Twitter and LinkedIn.

Unique features and Barriers to Entry (don't talk about APIs)

Unique features

The major unique feature is that our website retrieves job postings from social media sites such as Facebook, twitter, LinkedIn, Indeed, dice etc. Along with this, we provide unique features such as help user choose the best company, trend analysis of the technologies and recommendation engine for the candidate according to his skillset.

Suppose when a candidate has job offers from two companies and is in a dilemma to choose between one of them, we help me him choose a company by performing sentiment analysis of the companies.
Technologies are ever changing. We perform trend analysis of the technologies and show it to the user so that he/she understands the demand of workforce in that particular technology in the coming future.

When the user enters the website and registers, job recommendations will be made according to his skillset.

The unique features of our website include
- Visually attractive company comparison statistics
- Salary Analysis of different positions within each specified company
- Monthly sentiment analysis of each company with most positive and negative tweet.
- Integrate job searches from social media websites (Facebook, Twitter and LinkedIn)
- Provide clear description of Government jobs targeted at US residents
- Profile based job recommendations
- Provide the top 5 trending jobs to the user by getting market statistics
- Provide technological tips to the user
- Provide Search Engine optimizations (SEO) options for revenue generation
- Crows Sourcing option in the website for HR’s to put polls.

**Barriers to entry**

Though we have lot of features unique to our website, it takes time for the users to start visiting and using the website. Till then sustainability would be a major barrier for us. Some of the other barriers of entry are
- Need an established user base to generate revenue
- Need to partner up with educational institutes that offer online training courses
- Need to provide features not available in our competitors
- Need to generate revue to maintain sustainability

**Novelty details (algorithms, mining methods)**

**SoLoMo / Analytics functionalities**

The API such as Twitter, Facebook and LinkedIn helps us to incorporate the Social aspect of Web 3.0. The google maps API helps us to incorporate the Location aspect of Web 3.0. Since the team does not have mobile developers we would not be able to incorporate the mobile aspect of Web 3.0. However with proper support this application can be made supported on mobile platforms. The following analytics would be implemented

- **Social Component**: Our website gathers data and analyzes trends from Social Media websites.
- **Location Based Component**: Our website provides jobs depending on the users specification of the desired location.
- **Mobile Component**: Our website is multiple device compatible.
- **Sentiment analysis**: To analyze the user sentiments on a company
✓ **Trend analysis**: To analyze the trends in a given technology and its jobs and the cities which provide that jobs

✓ **Recommendation engine**: To analyze the users preference and suggests jobs that suits his profile / interest

✓ **Company Comparison Analysis**: To analyze and provide details of the companies having different functional units

### Mining Methods and Algorithms

In our website we used the following web mining methods and algorithms for analyzing different components of our website.

✓ **Sentiment Analysis – Alchemy API:**

Sentiment analysis combines natural language processing with artificial intelligence capability and text analytics to evaluate statements found across various social platforms to determine whether they are positive or negative with respect to a particular brand. Our website tracks sentiment analysis over time to:

- determine whether their actions improve or damage sentiment
- track brand reputation
- test how marketing efforts affect attitudes toward the brand
- attitudes toward new products

AlchemyAPI provides easy-to-use mechanisms to identify positive/negative sentiment within any document or web page. AlchemyAPI Sentiment Analysis APIs are capable of computing document-level sentiment, user-specified sentiment targeting, entity-level sentiment, and keyword-level sentiment. Multiple modes of sentiment analysis provide for a variety of use cases ranging from social media monitoring to trend analysis. API endpoints are provided for performing sentiment analysis on Internet-accessible URLs and posted HTML files.

✓ **Trending Analysis – Indeed API:**

One of the ways the website monitors what is going on in the web and mobile development industry is to look at job listings and the top trending jobs. While there are tons of sources to scour, a site called indeed.com conveniently aggregates multiple sources into a searchable trend database that allows you to look up and monitor evolving job titles, skill requirements, and technology trends over time. The website monitors evolving and growing job roles to help make strategy decisions on where to take software.

✓ **Prediction Analysis – WEKA API:**

To perform prediction analysis i.e. to predict the future trend for a specific technology we have used Weka data API. The process followed was to extract data using Weka API, store the data into MongoDB,
and perform regression analysis on this data to determine future trends. Weka is a java based software which interacts with data files and produces visual results. Data mined using Weka is stored in MongoDB which is an open source NoSql database.

Regression analysis was done to perform prediction i.e. determine future trends for a specific job. In regression analysis we generated a regression model using the existing data mined using WEKA and saved in MongoDB database. Based on the model we determined the trend of a specific technology. Weka has regression analysis feature in which we can supply data from a specific dataset and use this training dataset to create a regression model then predict the prediction variable or independent variable to using the regression model.

**System Design:**

**Revenue/Business model:**

**Business Plan:**
The purpose of this website is to establish content based jobs for the purpose of generating revenue through advertising and by user registration. This website aims to develop a community of users by offering value based content. Apart from jobs opportunities the social media platform is a great source for latest news and updates about various technologies, organizations, career and interview tips submitted by various recruiters and experts.

**Budget:**
The budget for creating our website can be divided into the following tasks and each task would approximately cost the following:

<table>
<thead>
<tr>
<th>Budget Line Item</th>
<th>Cost Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery &amp; Requirements Brief</td>
<td>N/C</td>
</tr>
<tr>
<td>Site Design &amp; Coding</td>
<td>$6,500 – $12,000+</td>
</tr>
<tr>
<td>Site Architecture, Wireframes &amp; Templates</td>
<td>$3,000 – $10,000+</td>
</tr>
<tr>
<td>Service</td>
<td>Cost Range</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Content Planning &amp; Development</td>
<td>$13,000 – $36,000+</td>
</tr>
<tr>
<td>Content Loading</td>
<td>$1,500 – $3,000+</td>
</tr>
<tr>
<td>Multilingual Site Versions</td>
<td>$10,000 – $25,000+</td>
</tr>
<tr>
<td>Technology Integration</td>
<td>$3,000 – $20,000+</td>
</tr>
<tr>
<td>Search Engine Optimization</td>
<td>N/C</td>
</tr>
</tbody>
</table>

The total estimated cost of the website is $37,000 – $106,000

**Promotion Plan- Search Engine Optimizations:**
The most important and inexpensive strategy is to rank high for our preferred keywords on the main search engines in “organic” or “natural” searches (as opposed to paid ads). Search engines send robot “spiders” to index the content of our webpage, so we should begin with steps to prepare our webpages for optimal indexing. We also need to ensure search engine robots to find all the pages in our website. We can also create video, audio and image content relevant to our website and optimize it to be ranked higher in searches.

**Revenue Plan:**
For a content-based website which does not sell products directly, the two main revenue models are charging customers for access to the site (or to a premium version of the site) or integrating advertising for other websites and businesses into the site. The premium version of the site might include features such as connecting with the HR, provide the users with a monthly report with various statistics. In future we can also plan to develop a module for HR too, where the HR gets a list of recommended candidates based on the user profile and how well their profile matches the job opportunities the HR has posted. And also we can provide the HR with the social media profile of the job seeker.
According to a survey 84% of the HR verify the job seekers social media profile before taking a hiring decision. By generating advanced reports that analyzes the job seekers social media profile, we can make this a premium service and can generate revenue.
We can also generate revenue by following the google model. Basically our website works like a search engine but focused on jobs and training on technology. We can ask the recruiters or the training companies to pay to us so that their company gets listed in top 100 search result.
The advertising model is more common, and can be used by us to capitalize on our growing popularity and earn some money on the side.
Our website has revenue models for different type of user’s targeted.

✓ **Revenue Plan for Job-Seekers:**
The revenue plan for job seekers involves dividing the applicants into 4 different categories and providing different features for each one of these individual categories. The Free Category is not charged any monthly fee while the other categories are charged a monthly premium.
Revenue Plan for HR Recruiters:
The revenue plan for recruiters involves dividing the recruiters into 4 different categories and providing different features for each one of these individual categories. The Free Category is not charged any monthly fee while the other categories are charged a monthly premium.
Revenue Plan for Search Engine Optimizations (SEO):
The revenue plan for SEO’s involves division into 4 different categories and providing different features for each one of these individual categories. The Free Category is not charged any monthly fee while the other categories are charged a monthly premium.

Revenue Plan for Advertisements:
The revenue plan for advertisements involves division into 4 different categories and providing different features for each one of these individual categories. The Free Category is not charged any monthly fee while the other categories are charged a monthly premium.
**Projection Plan/ Estimated Monthly Turnover:**

Our estimated monthly turnover is assuming we have on an average 100 users, 40 recruiters, 4 SEO’s and 4 Ad companies who are willing to use our website. We have further divided our assumptions and categorized each of these users into different categories namely Bronze, Silver, Gold and Platinum respectively.

<table>
<thead>
<tr>
<th></th>
<th>Bronze</th>
<th>Silver</th>
<th>Gold</th>
<th>Platinum</th>
<th>Monthly turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Job seeker</strong> (100 users)</td>
<td>$0 * 25</td>
<td>$25 * 25</td>
<td>$50 * 25</td>
<td>$75 * 25</td>
<td><strong>$3750</strong></td>
</tr>
<tr>
<td><strong>Recruiter</strong> (40 users)</td>
<td>$0 * 10</td>
<td>$25 * 10</td>
<td>$50 * 10</td>
<td>$70 * 10</td>
<td><strong>$1450</strong></td>
</tr>
<tr>
<td><strong>Search engine optimization</strong> (4 companies)</td>
<td>$15 * 1</td>
<td>$49 * 1</td>
<td>$99 * 1</td>
<td>$199 * 1</td>
<td><strong>$362</strong></td>
</tr>
<tr>
<td><strong>Ads</strong> (4 companies)</td>
<td>$15 * 1</td>
<td>$49 * 1</td>
<td>$75 * 1</td>
<td>$100 * 1</td>
<td><strong>$239</strong></td>
</tr>
</tbody>
</table>

We estimate a total monthly turnover of **$5801** per month.

**Roles/contributions of members**

- **Shriram Chandra Anbalagan:**
  Shriram was involved in the following tasks:
  - UI Development
  - API Development
  - Mongo DB Database
  - System Integration
  - Business Case
  - Analytics

- **Kowsik Krishnan**
  Kowsik was involved in the following tasks
  - UI development
  - API Development
  - MySQL Database
  - System Integration
  - EC2 setup
- Visualization

- **Parama Dutta**
  - Parama was involved in the following tasks
    - API Development
    - Infrastructure Setup
    - MySQL Database
    - EC2 Setup
    - Presentation
    - Design and Aesthetics

- **Janakiram Pulipati**
  - Janakiram was involved in the following tasks
    - API Development
    - MySQL Database
    - Analytics
    - Visualization
    - Business Case

**Screenshots:**

**Home Page:**
The page below shows the launching page of our website. It primarily shows:
- ✓ Success Stories: Candidates who have submitted their profiles in our website and have found relevant jobs
- ✓ Top trending jobs depending on the current market scenario
- ✓ User profile based recommendations
- ✓ The various types of jobs available in our website, among different categories like engineering, medicine, IT etc
Jobs: Social Media Jobs
The social media jobs page lists jobs from a keyword search on different social media websites like Facebook, Twitter and LinkedIn.

In the example below we have searched for **CyberSecurity** related jobs.
Jobs: Government Jobs
Our website also caters to the need of US citizens who would be interested in Government Jobs in a specific category.
Example below lists government jobs in **cyber Security**
Novelty Features: Company Review (Sentiment Analysis)
The company review page of our website allows users to see the current trend in the companies and gives the positive and negative sentiment of those companies, along with the most positive or negative comment. We are also giving the option of choosing the month and seeing the company analysis.

Novelty Features: Career Review (Salary Analysis)
Our website also offers the unique feature of providing salary analysis of a user specified company. In the example below, different companies having Big Data are taken and the salaries of the same can be compared. Oracle is taken as an example and various departments and job titles within Oracle are displayed with their salaries.
Candidate Tips: Technical and Career Tips
The above screenshots represent the career tips results for **resume writing and technical tips for Android Training** which will list all the Youtube videos related to Android training.

**References**

http://ai.arizona.edu/research/cyber/
http://www.slideshare.net/Grifreaper/research-paper-hacker-plague
http://www.wired.com/threatlevel/2012/05/skype-location-tracking/
http://www.privacyrights.org/data-breach
http://secanalysis.com/a-brief-analysis-of-shodan/
http://raidersec.blogspot.com/2012/02/searching-for-devices-using-shodan.html
https://www.governmentjobs.com/