

Text mining on business and computer science undergraduate internship job postings

Lakshmi Priya G K and Ruxin Dai
Department of Computer Science
University of Wisconsin - River Falls
River Falls, WI 54022
priyareddy1995@gmail.com

Abstract

Co-operative education has become popular worldwide. In this paper, we use a text mining methodology to analyze co-op internship job posting in order to understand the co-op market for students of College of Business and Economics (CBE) in University of Wisconsin - River Falls (UWRF). First, we develop a parser that extracts informative terms from job description data collected from the Internship Coordinator of CBE. These terms include soft skills, technical skills as well as benefits and indicators of company culture. Second, we group the job descriptions by attributes, sub-field and academic year and analyze the differences between various segments of the co-op internship from the GPA of intakes to market. We obtain insight that can benefit students, employers and the institution. The sub-fields categorization helps organizations to use inventory. It also guides individuals with career planning as well as academic institutions in advancing their business analytics curricula.

Introduction

Co-operative (co-op) education is being adopted at a fast pace. It is the bridge between the employer, the student and the academic institution and benefits everyone involved. Co-op programs prepare business students for senior management roles in business which includes accounting, finance, marketing, and human resources. As for computer science students, it is a study of computer programming and other specializations as they pertain to technology which makes it easier for them to find employment upon graduation. Co-op education combines academic study with alternating work terms which gives students an opportunity to test skills learned in the classroom, and to expand their knowledge through related work experience. In brief, the co-op process proceeds as follows. From the beginning of the year employers post job ads, jobs take vacations too that is during the summer, autumn represents the final hiring spree of the year internship job postings are posted approximately eight months prior employers interview selected candidates based on their resume. After few rounds of interview depending on the job. Finally, hiring decisions are made, and, at the end of the semester, students and employers have background check.

Co-op tries its best to fill gap between skilled employees needed and potential candidates for the job. Job market expects more skills from candidates. They will need more skills in terms of technical, soft skills and all aspects to fit in a suitable designation of their preference. Students opt between study and work terms. They try to get an internship so that they can enhance their skill set and prepare themselves for the organizational atmosphere. Students keenly wait to take up an Internship to make themselves stand ahead among the peers. National Association of colleges and employers act as a bridge between students and employers act as a bridge between students and employers nationwide. ACE connects more than 8,100 the college career services professionals at nearly 2,000 colleges and universities nationwide, more than 3,100 university relations and recruiting professionals, and the business affiliates that serve this community. Requirements from organizational point of view differs from each designation to designation and department to department. Organization look for technically strong candidates for IT, Development job and soft skill-oriented person for management and marketing job. Finance and Accounting must be carried with technically strong candidates. Likewise, requirements vary accordingly.

Students have different perspective at workplace competencies and proper understanding of job market helps avoid the gap between students and employers. Co-op provides an excellent quality of service. As per the survey results of employers and students, students state that they have not received quality training

from placement support and employers also indicate about their low satisfaction of hiring graduates directly from school. Eventually, the low satisfaction among freshers leads to search of only skilled labors and hiring from school gradually decreases. The educational institution is confused about training that must be given to students in order to acquire the attributes that are on demand in the job market. The mutual availability to both the ends makes the efforts fruitful. In this study, we analyze the co-op market using job postings to help address the above problems. We try to mine maximum possible number of job descriptions. These job descriptions are not designed as per the order or not well structured. It contains academic requirements, URL, Email, contact numbers, Qualifications and so on. The technical part of research is to extract the characteristics of the job and use it to understand about the co-op market. Designing a parser extracts job-related attributes from unstructured job descriptions. Soft skills, hard skills, company culture, amenities all information is a part of job description. By extracting and comparing the various types of job descriptions with the market trend. We can get insights from three group of stakeholders. From institution view point, it helps them use the information about the former students and attract the upcoming students. To train the current students regarding the trend in the job market. In student's viewpoint helps them to find and acquire the job that completely matches with their skill set. For an organization it helps to get potential candidate to make the designated job very productive. It also helps in accurate placement of people in their respective stream. As per now this is the outer framework for the research. Further work will be carried on accordingly. Chapter 3 - Experiments and Methodology Chapter 4 – Conclusion and Future Work and Chapter 5 – Reference.

Experiments and Methodology

Before getting hands on into the project or building the code there was a lot of research done on how Internship job postings are listed and how they were in a sequence. Most of the study made on the job postings were by reading the points from the postings. I thoroughly went through each point from the beginning which was time consuming, I came up with an idea that I could categorize these job postings based on different companies and then experiment on them.

When I separated every job posting based on the company. I found it a little easier to know and understand them better, later I started researching the job postings and found that they could be categorized even more deeper. This is when I started making types, where I separated every type of job associated with each company.

This made the overall process easier and efficient. When I made everything into split categories I could see that there were various roles of internship job postings and few sets of postings were related to same company and few were related to different companies but associated with same or similar or closely related to the job titles.

This gave me a clear picture about the roles and companies which have kind of same or related roles. When this picture was drawn on my mind, it made me easier to get into the next step of experiment which was to understand the functionality of each job title and the skills associated with them. Most jobs which had the same title or job functionality had the same skills needed in-fact. Where I made a deep research and found that job with same title also had very few skills set changes.

The next experiment was to make a layer from the categories I had separated and see how many had the job description similar or few same or few extremely different from each other. This was pretty much a challenging research. After all these kinds of experiments and research I concluded that I could find or describe few methods on achieving them. The steps I had to take was very clear by now. To get these things going, I had to build a code base then I had to also show the students the impact behind the analysis. I also came up with an approach to show graphical representation and implementations of the frequency of skills used and frequency of hot jobs in the present market.

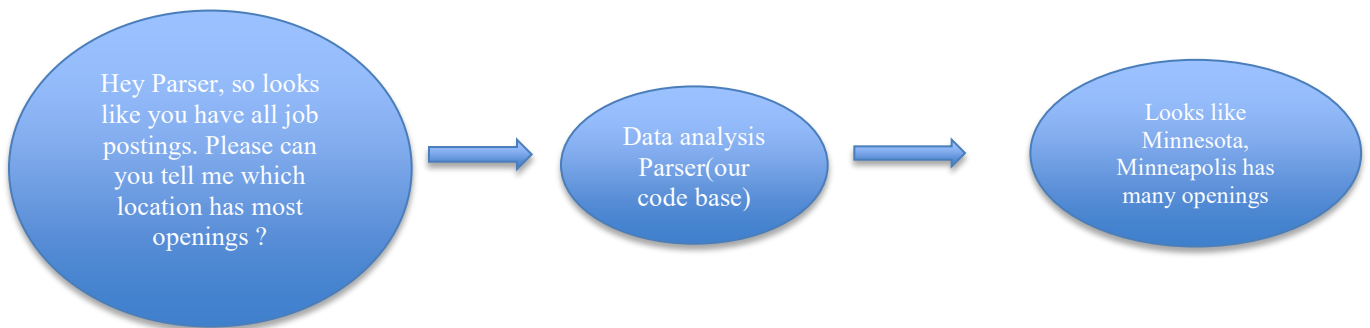
If this research reach students, it would start getting innovative and would start working on skills which they prefer are good for their job search.

I followed various Research methods for this parser which I built, the primary research I did is to understand how python reads or accepts word documents and the most efficient way to do so. Secondly, I made research on understanding the job posting, there are more than 4000 job postings and the efficient way to understand this is by grouping same company job postings at one place. Then comes the major part of research which contains different keywords regarding job postings like location, skills, job description etc. which are posted directly and indirectly and at same position in a word document and sometimes in a different position of the document. Lastly made research on how to build function which works on understanding this postings or word documents.

Analyzing the code which I developed and the function I wrote accepts the word document and the algorithm runs on each line of the document and searches for few keywords based on our research. This then repeats a cycle for all the job posting documents. The analysis for each part is then written as output to a path where files are automatically created. This is where most of our analysis gets

written to answer mostly all questions. From this point of analysis, we calculate the frequency and other aspects to answer the pointed questions in the document. Our results would bring difference because it's not possible for any organization or individual to go through all the job posting and find out which location has too many openings? (or) which skill is highly in demand?

Below shown diagram should give a basic user experience of the data analysis parser.



Note: This user experience picture above may just serve as an example only

Conclusion and Future work

Research was carried on for a job/internship finding portal for students of University of Wisconsin-River Falls, College of Business and Economics. Text mining methodology was used to study a Co-op market. The reading and research were not an easy task. The different attributes required for the job had to be identified. There were many attributes regarding different types of designations. Students with different fields of study required extra added attributes in order to acquire a job. Some were also dependent on seniority levels. There were some of the qualities required by all students irrespective of stream like Interpersonal skills. There were more assistant and Junior level jobs for undergraduates enter organization. An interesting insight gained was that there were number of summer internships available for both Undergraduates and Graduates. Summer seems to be more suitable time for an Internship rather than spring or summer. Summer had wide range of opportunities for both IT and Non-IT professionals. All job descriptions almost had some similar attributes requested like team player, strong oral and written skills, passionate to work etc. IT jobs had requirements of

software and their development like SQL, JAVA, C#, Python and so on. There was an upward trend in the IT jobs, as technical workforce has greater demand than the remaining work force. Finance jobs more likely needed analytical skills with minimal software requirements. But, IT had great boom in the job market and Hence, many junior or entry level jobs of Finance and Mechanical industry tend to move a little towards IT sector. Work atmosphere also has a great impact on the attributes that is been listed. Mechanical had labs and manufacturing firms, Finance consisted of office type firms, Casual and fun working environment in case of IT. There are many factors that affect the proper functioning a) Size of a firm b) Scale of operation c) Altering or modifying the work culture in order not to damage the public image of a company. Fun and casual work culture increases the productivity of an employee. But, sometimes nature of a business will not allow to have such an environment. For an instance, Financial environment cannot be controlled in the medium of protecting relationships more than business.

The Co-op study mainly focuses on the three stakeholders connected. First one being an institution. Institution can use the findings to train the students as per the job market requirements. Students who have secured a place at a prestigious organization can be kept track of and their successful rate can be used as a statistics to illustrate for getting the promotion done for an organization. It can help the institution to give wholesome training for students to reach the higher hierarchy of an organization. Second one being the organization, The findings say that some organization may have low satisfaction for freshers hired. But some freshers have taken up an intensive training and has converted themselves into an asset of the firm. Seniority level posts can also be filled with a help of this as it also exhibits seniority levels along with their skill set. Last but never the lease its Students who take utmost use of this. They get an opportunity to crave their own path as per their area of interest. To enhance their skill sets and match the prospective employer for efficient working and usage of skills. The major idea of Co-op is to find a right employee to employer and vice versa.

We can also make the finding to be clearly understood by following ways. a) Plot graph of most likely attribute b) Ratings of soft skills c) Companies with different job descriptions d) Plotting graph for most needed skill in computer science e) Company size and number of employees f) Academic record of students plotted on a graph. The current thesis has concentrated on Job description structure and extracting it and make it structural for employers and as well as employees.

This can also be improvised by adding some technological features to the system and making it more flexible and user friendly. An online interview interface to be created, online soft skills enhancement sessions can be helpful, Automatic resume screening and recommendations.

References

Text mining towards Data Science accessed on 2 Feb 2019, at <https://towardsdatascience.com/tagged/text-mining>

Undergraduate Internship job postings at CBE school from University of Wisconsin – River Falls accessed on 2 Feb 2019, at <https://www.uwrf.edu/CBE/Internship/Offerings/Index.cfm>

Employers point of view for internship job postings at University of Wisconsin – River Falls accessed on 2 Feb 2019, at <https://www.uwrf.edu/CBE/Internship/EmployerInformation.cfm>

University of Wisconsin – River Falls Data Collection accessed on 11 Feb 2019, at <https://www.uwrf.edu/CEPS/EPP/data-collection.cfm>

Internship preparation for a life after college accessed on 11 Feb 2019, at <https://www.uwrf.edu/CareerServices/Internships.cfm>

First destination survey accessed on 5 March 2019, at <https://www.uwrf.edu/CareerServices/FirstDestinationSurvey.cfm>

Jeffrey J. Selingo (2015), on internship job postings accessed on 5 March 2019, at https://www.washingtonpost.com/news/grade-point/wp/2015/05/18/are-internships-the-only-way-for-recent-college-grads-to-grab-entry-level-jobs/?noredirect=on&utm_term=.5cedf01e6a56

Dawn Rasmussen (2015), on how to make internship translate to employers accessed on 2 Feb 2019, at <https://www.workitdaily.com/how-internships-translate-employers>

CBE academic advising accessed on <https://www.uwrf.edu/CBE/Students/Advising.cfm>

Most popular Employers Perks accessed on <https://www.glassdoor.com/blog/15-of-the-most-popular-employee-perks/>

Integrating workplace and classroom learning accessed on https://link.springer.com/chapter/10.1007/978-90-481-3939-2_10

