Job Description Mining to Understand Work-Integrated Learning

Shivangi Chopra and Lukasz Golab
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
{s9chopra,lgolab}@uwaterloo.ca

ABSTRACT

Work-integrated learning, also known as co-operative education, allows students to alternate between on-campus classes and off-campus work terms. This provides an enhanced learning experience for students and a talent pipeline for employers. We observe that co-operative job postings are a rich source of information about the required skills, working environment and company culture. We present a text mining methodology to extract and cluster informative terms from unstructured job descriptions, and we demonstrate the utility of our methodology on a co-op job posting corpus from a large North American university.

Keywords

work-integrated learning, co-operative education, text mining, Latent Semantic Analysis (LSA)

1. INTRODUCTION

The World Association for Cooperative and Work-integrated Education reports that 275 institutions from 37 countries offer co-operative education (co-op) programs, also referred to as work-integrated learning programs[1]. Students enrolled in co-op programs usually alternate between on-campus classes and off-campus work terms at participating employers. Co-operative education has become popular for a number of reasons: it provides an enhanced learning experience for students, a talent pipeline for employers, and a recruiting tool for institutions.

Concurrent with the popularity of work-integrated learning is the desire to understand the co-op job market: students want to know what types of jobs are available and what skills could make them more employable; employers want to know what competition they are facing and how to attract top talent; and institutions want to align curricula with job market needs.

[1] http://www.waceinc.org/global_institutions.html

In this paper, we propose to answer the above questions by mining co-operative job postings. We make two contributions: 1) a text mining methodology to extract informative terms from job descriptions in order to understand a co-op job market, and 2) a case study using real data to demonstrate our methodology.

In practice, job descriptions are written directly by employers, and therefore they are not standardized or well-structured. In particular, job descriptions may include information that is unrelated to the nature of the job such as website URLs, contact emails, and of course common English words. Our technical challenge, therefore, is to extract and cluster useful information, such as required skills, working environment and company culture.

We address this challenge by designing a text mining methodology to understand a co-op job market through job postings. We start by building a parser that extracts relevant attributes from unstructured job descriptions. We then identify frequently occurring attributes in job titles and descriptions, and we employ Latent Semantic Analysis (LSA) and k-means clustering over the extracted attributes to characterize the types of available jobs.

To demonstrate the utility of our methodology, we analyze nearly 30,000 co-op job postings from a large North American university. We identify sought-after skills and mindsets, we identify the types of jobs available to junior and senior undergraduate students, and we discuss trends over time. We argue that our findings provide actionable insights for students, employers and the institution.

The remainder of this paper is organized as follows. Section 2 discusses related work; Section 3 describes our data and methodology; Section 4 describes the experimental results; and Section 5 concludes the paper with the implications of our findings and directions for future work.

2. RELATED WORK

This paper is related to three bodies of work: text mining, co-operative education and workforce studies. We use standard parsing and information retrieval techniques, and do not make any new algorithmic contributions in text mining. Instead, our contribution is to apply these techniques to a new application domain in order to obtain new insight.

Prior work on co-operative education has focused on its impact on students' skills (especially soft skills such as leadership and entrepreneurship), grades and post-graduate employment; see, e.g., [2, 14, 21, 26, 29]. There has also been research on what makes co-op students successful and what workplace competencies are expected (see, e.g., [6, 7, 16, 20, 30, 31]), understanding competition for co-op jobs (see, e.g., [17, 27]), and assessing the overall co-op process and experience (see, e.g., [12, 18]). These works are orthogonal to ours, which studies a different problem of understanding a co-op job market in terms of the types of available jobs and the required skills and attitudes.

Prior research on job advertisements studied how to write them in order to attract qualified applicants (see, e.g., [4, 11, 22]), and how to match job descriptions with qualified resumes (see, e.g., [9, 19]). Moreover, job descriptions have been studied from a gender perspective, e.g., by counting the occurrences of masculine and feminine words [25]. While these works investigated how job descriptions could attract or match applicants, we study a different problem of understanding a co-op market through job descriptions.

Workforce literature has applied machine learning to improve recruitment, reduce turnover and understand work profiles [1, 5]. Machine learning algorithms have been applied to understand the factors affecting work performance and retention [5]. Furthermore, Aken et al. cluster Information Technology job postings on job websites to understand the work profiles prevalent in the market [1]. Our research extends this analysis to understand the work profiles of various industries (not only Information Technology) in a co-operative education setup and how they have changed over time. Not limiting the scope to broad work profiles, our research also highlights the specific skills and attitudes required by various industries.

3. DATA AND METHODOLOGY

We obtained two datasets from a large undergraduate North American institution: 12,066 job postings corresponding to all co-op jobs that were advertised and filled in 2004, and 17,057 job postings corresponding to all co-op jobs that were advertised and filled in 2014. The job postings are written in English. Most of these positions were located in North America, with a small number of overseas jobs. We use the 2014 data to characterize the current co-op job market and we compare with the 2004 data to analyze trends over time. Each record in our datasets contains the following information:

- A job title, up to 50 characters long, which generally consists of the position and/or the nature of the work. Common titles include Web Developer, Engineering Intern and Planning Assistant.
- A job description, with unlimited length and no standardized structure or formatting.
- The year of study of the successful candidate who secured the job. We refer to jobs obtained by first and second year students as *junior jobs* or *lower-year jobs*, and those obtained by third and fourth year students as *senior jobs* or *upper-year jobs*.

Note: EMPLOYMENT BASED IN THE USA* This work opportunity will be based in the USA; therefore all applicants must determine whether they are eligible to work in the USA.

Aqua Book Club (ABC), is a global eReading service <href=www.abc.ca. Ranked 1st in Bloomberg Magazine's annual ranking of startups, we have a strong employee culture that promotes teamwork and open communication.

ABC is looking for Javascript/HTML5/CSS/RoR experts who are obsessed with technology and who love what they do. As part of our small team of software engineers, you will be responsible for architecting and implementing the UI designs, and working with other members on the team to integrate the the application into our platform.Deep understanding of the front end web, from delivery to working with Ajax is required. Experience in Ruby on Rails or other MVC web frameworks is a plus.

Applications are due by 05/30/2014 12 a.m. Applications wont be accepted after that. Attaching a transcript is highly recommended. (Include #503482 in the name) - Currently enrolled in BASc or CS at the Intermediate level with the Co-op option – Students who have taken cs326 will be prefered

At ABC, you will get a chance to work closely with the CEO Tim while having the flexibility you need to make a real contribution to our system. If you have a past history of excellence, are un-put by challenges, are a team-player and have demonstrated ability to learn rapidly on the job, we want to talk to you. Other perks: - Get to work on really challenging and diverse problems in a casual environment. - We have a ping-pong and a foosball table (We will surely beat you in ping pong)! - A well stocked fridge - free lunch on release days!!! ie we're basicaly a really F*U*N place to work. The office is located downtown and is easily reached by TTC.

Join us for the Evening Happy Hour on Friday, May 23rd 2014, 7:30 pm. Check out the Facebook event page here: https://www.facebook.com/events/573997/.

##########Feel free to contact Ruby Smith (rsmith@abc.com) or Jason Pinn (jason@abc.com) for any questions you have about working at ABC.

Apply asap!

Figure 1: An anonymized job description

- The academic program of the successful candidate.

Since the job postings in our dataset do not include industry or discipline labels, we use the academic program of the student who obtained the job as a proxy. The institution provided us with a mapping from students' academic programs to job disciplines; e.g., positions filled by Computer Science or Software Engineering students are classified as Information Technology jobs. In our case study, we focus on the largest discipline in the institution's co-op market: Information Technologies (IT). We also point out interesting findings from other major disciplines: Finance, Health Studies, Arts, Biology, Environmental Studies, Chemical Engineering, Civil Engineering, Electrical Engineering and Mechanical Engineering.

Figure 1 shows an anonymized example of a job description from our dataset. It includes the following information:

- Technical skills: Javascript, Ruby on Rails
- Soft skills: team player, ability to learn
- Job duties: architecting and implementing UI designs
- Desired mindset and attitude: obsessed with technology
- Perks: ping-pong and foosball table, free lunch
- Company culture: casual environment

However, there is also some content that does not describe the job itself: names of people and locations, URLs, email addresses, HTML tags, timestamps, special formatting, and, of course, common English words. The first part of our methodology, therefore, is a parser that extracts job-related attributes from unstructured job descriptions. The parser, implemented in Python, consists of the following steps.

1. Using regular expression matching, we remove URLs, HTML tags, phone numbers and other numbers, email addresses, timestamps, administrative annotations added by the institution (such as the text following "Note:" in Figure 1), formatting characters such as bullet points, and sequences of special characters serving as separators (such as the sequences of dashes and hashtags in Figure 1).

2. We tokenize the remaining text and remove special characters embedded in words (such as F*U*N in Figure 1). To remove unimportant terms, we build a vocabulary, called *Remove-List*, consisting of common English words[2], misspellings[3] and abbreviations[4], as well as manually-curated lists of company names, locations, addresses and persons' names appearing in the institution's co-op system.

3. We have to be careful to not remove informative terms. For example, "Ajax" is a city in Canada and is therefore in *Remove-List*. However, Ajax is also a Web development toolkit. To address this problem, we create another vocabulary called *Keep-List*, of words that should *not* be removed. This vocabulary consists of skills found on a resume help Web site[5] and job duties from the Canadian National Occupation Classification[6]. Note that *Keep-List* only contains a subset of words we are interested in; e.g., it is missing many specific technical skills, perks and company culture descriptors.

4. We stem the remaining tokens using the NLTK snowball stemmer [7] and we remove stop words. Finally, we leverage our domain knowledge by converting important terms that can be written in different ways into a standard form; e.g., "java-script" and "javascript" both map to "javascript".

At the end of the parsing process, each job description is reduced to its stemmed words, minus those in *Remove-List* but not in *Keep-List*. In the remainder of the paper, we will refer to these stemmed words as "words", "terms", "tokens" and "attributes" interchangeably.

The second part of our methodology is designed to analyze the extracted job attributes. We do this in two ways:

1. To identify popular skills, attitudes, working environment and perks, we report attributes that occur at least once in a large percentage of job descriptions. Notably, and in contrast to other text mining applications, we do not count the number of occurrences of an attribute within a posting–we observed that important job requirements such as knowledge of the "Java" programming language are usually mentioned only once. We also identify attributes mentioned by more junior than senior jobs (and vice versa), and we compare attributes mentioned by more jobs in 2014 than 2004 (and vice versa) to characterized trends over time.

2. We use clustering to identify the different types of available co-op jobs within a discipline. Following previous work on text clustering [10, 23, 24], we start by applying Latent Semantic Analysis (LSA) to the job descriptions, with each job description represented as a *job vector*. The ith coordinate of a job vector is equal to the inverse document frequency (IDF) of the ith word in the set of possible words, provided that this word is mentioned in the given job description at least once (and zero otherwise). Following previous work, we use LSA to reduce the dimensionality of job vectors from the number of distinct words down to one hundred [28]. Each reduced dimension corresponds to a latent concept in the data. We then run k-means clustering on the transformed job vectors, and we report a few top terms (again, ranked by IDF) from each cluster centroid as representatives.

[2] http://www.lextutor.ca/freq/lists_download/longman_3000_list.pdf
[3] https://en.wikipedia.org/wiki/Wikipedia:Lists_of_common_misspellings/For_machines
[4] https://media.gcflearnfree.org/ctassets/modules/48/common_abbr.png
[5] https://www.thebalance.com/list-of-the-best-skills-for-resumes-2062422
[6] http://noc.esdc.gc.ca/English/noc/welcome.aspx?ver=16
[7] www.nltk.org/_modules/nltk/stem/snowball.html

Table 1: Top 10 frequent tokens in IT job titles

Token	Freq. in 2014	Token	Freq. in 2004
softwar	45%	develop	37%
develop	44%	softwar	27%
analyst	8%	analyst	17%
applic	7%	programm	11%
web	5%	assist	9%
support	4%	web	8%
assist	4%	support	7%
programm	4%	applic	6%
system	3%	system	6%
quality	3%	specialist	4%

4. RESULTS

In this section, we demonstrate the utility of our methodology. We show in-depth results for the largest discipline in our dataset: Information Technologies (IT), including frequent term analysis (Section 4.1), analysis of significant differences in term frequencies between 2014 and 2004 and between senior and junior jobs (Section 4.2), and clustering analysis (Section 4.3). We summarize our results for other disciplines in Section 4.4.

4.1 Frequent Term Analysis

Table 1 shows the top 10 attributes occurring in the most IT job *titles* in 2014 and 2004; for example, the first row indicates that the token "softwar" appears at least once in 45% of job titles in 2014 and 37% in 2004. Not surprisingly, nearly half the titles mention software development.

Table 2 shows the top 25 attributes occurring in the most IT job *descriptions* in 2014 and 2004. Overall, most IT co-op jobs appear to be software developer jobs. In 2014, hardware was mentioned in only 14% of the postings and embedded systems in 7%; in 2004, these percentages were slightly

Table 2: Top 25 frequent attributes in IT job descriptions

Token	Freq. in 2014	Token	Freq. in 2004
develop	91%	develop	80%
team	84%	applic	65%
softwar	76%	softwar	62%
applic	66%	system	61%
design	65%	team	61%
product	62%	program	54%
program	60%	design	53%
system	58%	communic	50%
project	53%	comput	49%
comput	52%	product	47%
test	50%	support	43%
build	48%	test	43%
communic	48%	servic	42%
web	47%	project	41%
code	46%	lead	39%
help	46%	excel	39%
learn	45%	solut	38%
servic	44%	web	38%
java	43%	tool	37%
manag	43%	assist	36%
creat	43%	busi	36%
solut	42%	manag	35%
technic	42%	java	35%
tool	41%	custom	34%
excel	40%	oper	33%

higher, at 22 and 9, respectively (and the actual number of hardware and embedded systems jobs was slightly higher in 2004). Furthermore, about half the job descriptions mention testing. Notably, mentions of some soft skills such as communication are more frequent than mentions of specific technical skills such as Java in both years.

By inspecting other frequent attributes, we obtain the following insights about frequently mentioned programming languages, platforms and applications in 2014:

- Programming languages: Java (43%), C++ (33%), JavaScript (31%), C (24%), Python (22%), C# (20%), HTML (19%), CSS (17%), PHP (12%), .NET (12%), jQuery (10%), Perl (10%), XML (9%), Ruby (9%)
- Development: web (47%), mobile (32%), game (12%)
- Databases: database (29%), SQL (26%), mySQL (8%), Oracle (7%)
- Mobile applications: android (19%), iPhone (7%)
- Operating Systems: linux (21%), unix (13%), iOS (14%)
- User-centered development: user (35%), agile (18%), deploy (16%)
- Other applications: server (29%), distributed (17%), security (17%), cloud (9%), graphic development (8%), big data (4%)
- Concepts: OOP (Object-Orient Programming) (24%), algorithms (18%), scalable (14%)

In terms of the working environment and company culture, the strongest result is that the word "team" is very frequent, suggesting a collaborative environment. Other frequent terms include challenging (32%), dynamic (20%), fun (16%), flexible (15%) and diverse (12%). Amenities such as free food, foosball and ping-pong tables are also frequent. The word start-up is mentioned in 11% of the job postings.

We also note the occurrence of mindset-related terms such as learn (45%), innovation (32%), passion (25%), focus (23%), creativity (22%), motivation (20%), love (15%) and enjoy (10%).

Similarly, for 2004, we identify the following frequently mentioned programming languages, platforms and applications:

- Programming languages: Java (35%), C++ (31%), C (21%), HTML (22%), XML (15%), ASP.NET (12%), Perl (11%), .NET (10%), JavaScript (10%), JSP (8%), C# (7%)
- Development: web (38%), mobile (10%), game (5%)
- Databases: database (30%), SQL (27%), Oracle (13%), mySQL (2%)
- Operating Systems: unix (22%), linux (15%)
- User-centered development: user (21%), deploy (7%), agile (0.5%)
- Other applications: server (25%), security (15%), graphic development (10%)
- Concepts: OOP (13%), algorithms (7%), scalable (4%)

Compared to 2014, the word "team" was again frequent in 2004, but words related to mindset, company culture and perks were less frequent.

Our results indicate that IT positions focus on software rather than hardware, especially web and Java development. The work environment appears team-oriented. In 2014, descriptions of mindset and company culture are appearing frequently.

4.2 Significant Differences

Next, we investigate the differences between 2014 and 2004 IT job descriptions which we began to see in the previous section. Table 3 summarizes the results by listing 20 attributes with most significant differences in frequencies between 2004 and 2014 (on the left), and 2014 and 2004 (on the right). We define a difference in frequencies, abbreviated Δ, as the percentage of job postings mentioning an attribute in one year minus the percentage of job postings mentioning this attribute in the other year. Both lists are sorted by Δ, and all results shown are statistically significant with P-values less than 0.05 using a proportion test [13]. We omit the analysis of job title differences between 2004 and 2014 which gave similar results. We also show a Venn diagram in Figure 2, which illustrates the overlap among the top 100 frequent attributes in 2004 and 2014 IT jobs.

Table 3: Differences in frequency between job description attributes of 2004 and 2014 IT

Token	2004	2014	Δ	Token	2014	2004	Δ
assist	36%	22%	14%	build	48%	15%	33%
asp	12%	2%	10%	help	46%	19%	26%
internet	18%	9%	9%	team	84%	61%	24%
unix	22%	13%	8%	code	46%	24%	22%
hardwar	22%	14%	8%	mobil	32%	10%	22%
sort	8%	0%	8%	javascript	31%	10%	21%
clarifi	8%	1%	8%	passion	25%	5%	20%
interperson	18%	10%	7%	featur	30%	10%	20%
oper	33%	26%	7%	creat	43%	23%	20%
msaccess	8%	1%	7%	python	22%	3%	19%
manufactur	10%	4%	6%	learn	45%	26%	19%
cost	11%	5%	6%	collabor	23%	5%	18%
xml	15%	9%	6%	agil	18%	0%	18%
support	43%	37%	6%	product	62%	47%	16%
expens	8%	2%	6%	contribut	27%	12%	15%
intranet	7%	1%	6%	problem	34%	19%	15%
oracl	13%	7%	5%	improv	25%	10%	15%
prepar	11%	6%	5%	solv	20%	6%	15%
supervis	12%	7%	5%	app	15%	1%	14%
xp	8%	3%	5%	peopl	33%	18%	14%

Table 4: Differences in frequency between job description attributes of junior and senior jobs in 2014 IT

Token	Jr.	Sr.	Δ	Token	Sr.	Jr.	Δ
document	29%	16%	13%	c++	46%	21%	24%
support	42%	31%	11%	algorithm	28%	9%	20%
assist	27%	16%	11%	scale	28%	9%	19%
communic	53%	43%	10%	scienc	49%	31%	17%
manag	48%	38%	10%	featur	39%	22%	17%
test	54%	45%	9%	python	31%	14%	16%
report	26%	17%	9%	scalabl	23%	7%	16%
busi	42%	34%	9%	build	57%	41%	15%
written	21%	13%	8%	code	54%	40%	15%
activ	23%	15%	8%	complex	27%	13%	13%
educ	17%	10%	7%	comput	59%	46%	13%
standard	15%	8%	7%	c	31%	18%	13%
interperson	13%	6%	7%	product	69%	57%	13%
instal	9%	3%	7%	structur	21%	9%	12%
troubleshoot	15%	9%	6%	field	23%	11%	12%
msoffic	8%	2%	6%	java	50%	38%	12%
summari	24%	18%	6%	data	42%	30%	12%
execut	15%	9%	6%	distribut	23%	12%	11%
detail	11%	5%	6%	search	16%	6%	10%
account	12%	6%	6%	problem	40%	29%	10%

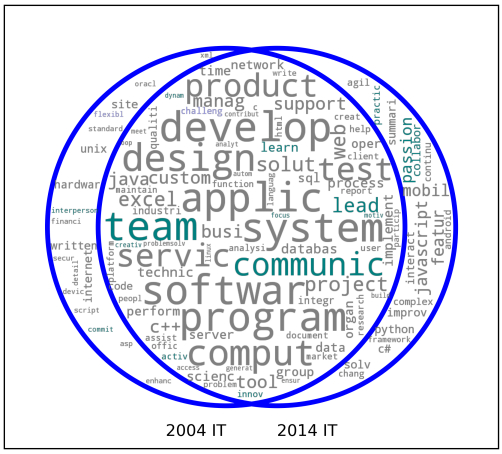

Figure 2: Overlap between the top 100 most frequent attributes of IT jobs in 2004 and 2014

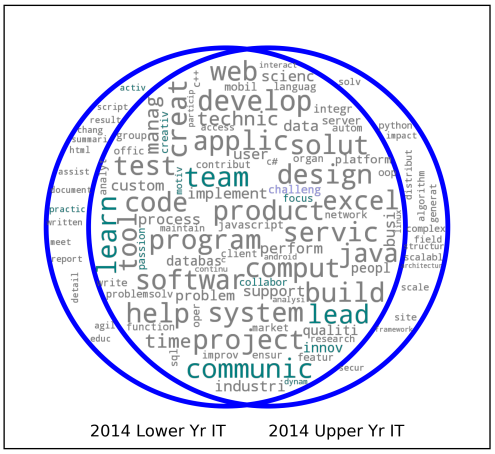

Figure 3: Overlap between the top 100 most frequent attributes of Junior and Senior IT jobs in 2014

Our results suggest that 2004 job postings include more entry-level positions (suggested by attributes such as "assist", "support", "prepare", "arrange" and "document"), and mention technologies and software popular at the time such as ASP, XML, Windows XP and Microsoft Access. Additionally, the fraction of hardware-oriented jobs was higher in 2004. On the other hand, job postings in 2014 include words representing current technologies such as mobile, Javascript, Python, agile and app (and, further in the list, scalable and distributed systems). Notably, many soft skills and mindset-related terms are more frequent in 2014: "passion", "create", "learn", "collaborate" and "contribute". Although not shown in Table 3, other terms that are more frequent in 2014 include company culture descriptors such as "innovative","challenging", "fun" and "diverse".

The next important difference is that between junior and senior jobs. Table 4 shows two lists: top terms appearing in more junior than senior jobs (on the left), and top terms appearing in more senior than junior jobs (on the right), both in 2014 and both sorted by the difference of percentages. Table 5 shows the same two lists, but for 2004. Figures 3 and 4 show Venn diagrams that illustrate the overlap among the top 100 frequent terms from junior and senior jobs in 2014 and 2004, respectively.

We observe that in 2014, junior jobs are more likely to be entry-level documentation, testing or troubleshooting jobs. Junior job postings are more likely to mention soft skills such as communication and interpersonal skills. In terms of specific technologies, junior jobs mention HTML, SQL and Web 5 percent more frequently than senior jobs. On the other hand, senior jobs in 2014 mention technical concepts and specific programming languages such as algorithms, scalability, data, C++, C and Python. Other interesting differences not shown in the table are OOP (9% more frequent

Table 5: Differences in frequency between job description attributes of junior and senior jobs in 2004 IT

Token	Jr.	Sr.	Δ	Token	Sr.	Jr.	Δ
maintain	23%	14%	9%	c++	45%	21%	24%
support	47%	38%	9%	c	32%	14%	18%
updat	13%	5%	8%	design	63%	46%	17%
html	26%	18%	8%	cost	20%	5%	14%
excel	43%	35%	8%	clarifi	16%	2%	14%
msoffic	13%	5%	8%	expens	16%	2%	14%
troubleshoot	14%	7%	8%	arrang	17%	4%	13%
document	30%	23%	7%	sort	16%	3%	13%
user	24%	17%	7%	solut	44%	33%	11%
qualiti	26%	20%	6%	challeng	29%	19%	10%
report	26%	20%	6%	develop	85%	76%	9%
web	40%	34%	6%	linux	20%	11%	9%
mainten	14%	8%	6%	complex	16%	7%	9%
instal	13%	7%	6%	algorithm	12%	3%	9%
interperson	20%	15%	5%	unix	27%	18%	9%
server	27%	22%	5%	code	29%	21%	8%
hardwar	24%	19%	5%	lead	44%	36%	8%
xp	10%	5%	5%	innov	27%	19%	8%
time	31%	26%	5%	oop	18%	10%	8%
offic	21%	17%	5%	scale	11%	3%	8%

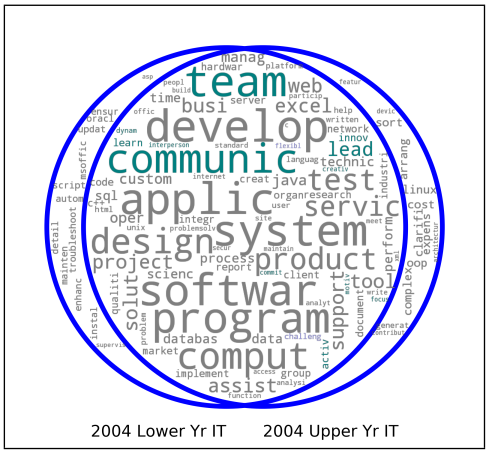

Figure 4: Overlap between the top 100 most frequent attributes of Junior and Senior IT jobs in 2004

than in junior jobs), linux (8%), cloud (8%), security (7%) and data science (5%).

We observe similar patterns in Table 5 and Figure 4. In 2004, junior jobs also included terms suggesting entry-level positions, whereas senior jobs included more mentions of programming languages and computing concepts.

To summarize, there are clear differences between 2014 and 2004 IT jobs, and between junior and senior jobs. In addition to differences due to new technologies, soft skills, mindset and company culture are more frequently mentioned in 2014. In both years, junior IT jobs are more likely to mention documentation, testing and troubleshooting, whereas senior jobs are more likely to mention technical concepts.

Table 6: Largest clusters of 2014 IT jobs

Label	Tokens in cluster centroid	%All	%Jr.	%Sr.
Web Development	javascript, html, web, css, sql, c#, server, java, net, jquery	22%	**64%**	36%
Programming	c++, c, languag, linux, python, oop, scienc, algorithm, perl, script	21%	46%	**54%**
Start-up Culture	startup, python, javascript, featur, code, web, love, stack, fun, passion	18%	39%	**61%**
Business Analyst	sql, analyst, test, solut, c#, script, execut, financi, document, busi	16%	**69%**	31%
Mobile Development	io, android, mobil, app, platform, java, agil, iphon, devic, c	10%	**61%**	39%
System Administrator	hardwar, troubleshoot, configur, instal, network, desktop, server, user, xp, resolut	6%	**87%**	13%

4.3 Clustering Analysis

After investigating frequently occurring terms, we now cluster the IT job descriptions to understand the types of available jobs. We experimented with different numbers of clusters between 2 and 30. We present results using ten clusters; using fewer clusters led to different types of jobs being assigned to the same cluster, whereas using more clusters led to similar types of jobs belonging to multiple clusters.

Table 6 shows the six largest clusters in 2014 sorted by size; the remaining four clusters had under 2% of the total number of jobs each. We report the representative tokens of each cluster centroid, a manually-assigned label summarizing the tokens, and three percentages: the percentage of all jobs assigned to this cluster, and the percentages of junior and senior jobs within this cluster. We highlight the higher of the last two percentages in bold font to indicate whether a cluster consists of more junior or senior jobs.

Based on the clustering results, we characterize the IT co-op market as follows. The five largest clusters cover 87% of IT jobs, spanning web development (22%), programming (21%), start-ups (18%), business analysis (16%) and mobile development (10%). The junior vs. senior split evident in the clustering is consistent with our earlier results from Section 4.2: troubleshooting jobs are mostly filled by junior students, whereas jobs mentioning company culture, many of which are startups, are filled by senior students.

Table 7 shows the 7 largest IT clusters in 2004; the remaining three clusters are small and one of them contains job postings from a specific large employer at the time. There is no longer a start-up cluster with mentions of the working environment, and there is an emphasis on hardware in the last cluster. These results align with our earlier results from Section 4.1.

To summarize, our clustering methodology segments the IT job market into web development jobs, general software development jobs, data analysis jobs, mobile

Table 7: Largest clusters of 2004 IT jobs

Label	Tokens in cluster centroid	%All	%Jr.	%Sr.
Software Development and Testing	java, test, sql, tool, server, qualiti, softwar, autom, custom, solut	20%	**61%**	39%
Web Development	html, sql, web, asp, javascript, server, java, xml, databas, net	19%	**66%**	34%
Databases	scienc, databas, model, comput, analysi, group, research, data, tool, msaccess	15%	**57%**	43%
System Development	c++, sort, clarifi, expens, arrang, cost, gui, code, java, softwar	15%	40%	**60%**
System Administrator	network, hardwar, troubleshoot, instal, user, configur, xp, desktop, msoffic, problem	9%	**87%**	13%
Programming	perl, script, languag, unix, c, java, rank, c++, enterpris, linux	7%	**57%**	43%
Embedded Systems and Graphics	video, digit, hardwar, c, multimedia, debug, embed, devic, c++, graphic	7%	36%	**64%**

development jobs and troubleshooting jobs. Mentions of mindset and work environments in 2014 are frequent enough to create a separate cluster for these jobs.

4.4 Analysis of Other Disciplines

In this section, we apply our text mining methodology to the other disciplines in our dataset. As before, we structure the results into frequent term analysis, difference analysis (2014 vs. 2004 and junior vs. senior jobs), and clustering analysis to characterize the types of available jobs within each discipline. We focus on job description analysis and only mention the results of job title analysis if they lead to additional insight.

4.4.1 Frequent Term Analysis

Overall, all the other disciplines have frequent mentions of soft skills ("team", "communication", "leadership") and basic computing skills (databases and Microsoft Office) in both 2004 and 2014. Below, we highlight additional frequent terms for each discipline.

Finance: soft skills indicating client relationships ("client", "interpersonal", "relationship"); finance-specific technical skills ("audit", "tax", risk assessment, asset valuation, market analysis); formal office working environment ("bank", "office")

Health Studies: soft skills ("active students", indicating physical fitness); health-specific terms ("patient", "care", "kinesiology", "therapy", "injury", "rehabilitation", "ergonomics", "physiotherapy", "recreation")

Arts: tokens related to editorial, technical and content writing ("edit", "write", "english", "proofread", "content"); additionally, media and social media were frequently mentioned in 2014.

Biology: discipline-specific technical terms ("molecular", "chemistry", "microbiology", "biochemistry", "disease", "cell", "tissue", "DNA", "genetics", "pharmaceutical"); lab-oriented work environment ("research", "lab", "technician")

Environmental Studies: discipline-specific terms (GIS (Geographic Information System), "water", "land", "soil", "map", "survey", "sample", "policy"); field work environment ("field", "site"). Frequent words in job titles: "assistant", "planner", "technician", "research", "analyst", "inspector", "project", "management".

Chemical Engineering: Discipline-specific technical terms ("chemistry", "process", "manufacturing", "equipment", "sample", "procedure", process improvement, "safety"); lab-oriented work environment (chemical plants, research labs). Additionally, frequent in 2014: project management; frequent in 2004: field-work.

Civil Engineering: construction-related tokens ("design", "AutoCAD", "site", "field", "concrete", "safety"); graphic design ("graphic", "PhotoShop").

Electrical Engineering: discipline-specific technical skills ("electrical", "hardware", "power", "schematic", "control", "embedded", "circuit"); computing skills ("code", Web, Java, SQL). Frequent terms in job titles: "design", "quality", "assurance", "testing", "research".

Mechanical Engineering: discipline-specific terms ("equipment", "assembly", "robot", "circuit", "material", "CAD", "SolidWorks", "AutoCAD", "control", "process", "improvement", "maintenance", "draw", "prototype", "test", "troubleshoot", "safety"); work environment ("plant", "shop", "floor", "manufacturing").

4.4.2 Significant Differences

Next, we highlight differences in frequent terms between 2004 and 2014. Overall, we observed that each discipline had more mentions of soft skills, and more mentions of project management and IT-related terms in 2014. Additional differences are summarized below for each discipline.

Finance: 2004 jobs mention actuarial science more; 2014 jobs mention risk management and assessment, "equity", "trade", "client" and "interaction" more. Additionally, 2014 jobs mention concepts related to data analysis (e.g., Microsoft Excel and VBA).

Health Studies: 2014 jobs include more research related terms: "research", "summary", "data", "review", "cancer". 2004 jobs have more mentions of "recreation", "kinesiology", "outdoor", "therapy" and "teach". In particular, "cancer" appears in 6% more job postings in 2014 than in 2004.

Arts: more 2014 jobs mention market analysis and media-related terms: "media", "project", "management", "PowerPoint", "client" and "relationship". 2004 jobs mention more writing-related terms such as "history", "newsletter", "proofread", "French" and HTML.

Biology: 2014 job postings include more research and project management positions, and mention computing

skills and clinic more often. 2004 job postings mention laboratory terms including "technique", "microbiology", "sample", "gel", "biochemistry", "microbe", HPLC (High Performance Liquid Chromatography blood test), "bacteria".

Environmental Studies: 2014 jobs mention project management, clients, research and computing skills more often. 2004 jobs mention "educate", "air", "waste", "treatment", "recycle" and ground water. It is interesting to note that "sustain" (sustainability) is mentioned 7% more often in 2014 than in 2004.

Chemical Engineering: 2014 jobs mention project management terms (e.g., "manage", "report", "project", "maintain"), "safety", "energy", "oil", "gas", "petroleum" and "sand" more often than 2004 jobs. On the other hand, 2004 jobs mention more computing skills and laboratory-specific terms ("lab", "technician", "sample", "treatment").

Civil Engineering: 2014 jobs mention more software ("software" and "AutoCAD" are mentioned 21% and 8% more often, respectively, in 2014 than in 2004). 2004 jobs mention "cost" and "expense" more often than 2014 jobs. It is interesting to note that "safety" is mentioned 13% more often in 2014 than in 2004.

Electrical Engineering: 2014 jobs mention "passion" and computing skills related to web development, core programming languages and mobile development. 2004 jobs mention more "manufacturing", "graphic", "multimedia", "processor", "hardware", "VHDL" (a hardware description language) and "Unix".

Mechanical Engineering: 2014 jobs mention research (suggested by "lab", "research", "simulate", "electron"), client-oriented development ("client", "customize") and computing terms (Python, Java, "mobile"). 2004 jobs are more likely to mention mechanical engineering terms: "blueprint", "draw", "cost", "weld", "hydraulics", "gear". It is interesting to note that "quality" is mentioned 9% more in 2014 than in 2004. While both AutoCAD and SolidWorks are CAD software, SolidWorks is mentioned 11% more in 2014 while AutoCAD is mentioned 5% more in 2004.

Next, we compare the differences between tokens in junior and senior jobs in each discipline. Overall, more senior jobs across all disciplines mention project management or deal with advanced concepts of the field (either through applications or research). Junior jobs appear to have more clerical work, computing-related responsibilities or mention less advanced concepts of the discipline (including testing, field work and lab work). We provide additional discipline-specific details below.

Finance: Senior jobs require more technical knowledge of the field ("audit", "invest", "risk" "management"). Junior jobs have a more clerical ("document", "arrange", "English") and computing (HTML, Java, databases) focus. Senior jobs are more likely to mention "commitment", "dynamism", "client" and "interaction". Additionally, senior jobs in 2004 mention more mathematical and statistical terms than junior jobs in 2004.

Health Studies: Senior jobs mention more research. Junior jobs mention more field work.

Arts: Senior jobs mention more project management (suggested by "manage", "PowerPoint", "client", "workload", "process", "improvement"). Junior jobs mention more clerical work, "English", "Web", "research" and "customer service". Additionally, senior jobs in 2004 appear to include more business analyst and editor roles than junior jobs in 2004.

Biology: Senior jobs mention more "research", "hospital" and technical terms including "genetics", "therapy", "cancer", "cardiovascular", "nanomedicine", "biomaterial" and "in vitro". Junior jobs are more likely to mention "office", "assistant", "support" and "campaign".

Environmental Studies: Senior job titles indicate more planner and analyst positions with more project management, policy-making and GIS terms mentioned in the descriptions. Junior job titles indicate more lab technician, inspector, and surveyor positions with more "lab", "survey", "test" and "outdoor" mentioned in the descriptions. 2004 senior jobs additionally mention environmental concepts including "ground", "water", "remedy", "contaminate", "river", "hydrology" and "hydrogeology".

Chemical Engineering: Senior jobs mention a more industrial working environment with more mentions of "energy", "product", "design", "cost", "process", "improvement" and "optimization". Junior jobs mention laboratory-specific terms ("research", "sample", "record") more often. In 2004, senior jobs mentioned more chemical manufacturing terms.

Civil Engineering: Senior jobs mention more "modelling", "design", "client", "interaction" and "software". Junior jobs mention more "inspection", "field", "survey", data recording and clerical work. 2014 senior jobs have more mentions of project management.

Electrical Engineering: Senior jobs mention more electrical concepts ("power", "circuit", "embedded", "distributed", PCB (Printed Circuit Board), "sensor", "chip", "schematic"). Junior jobs mention more quality assurance and basic computing terms ("web", "program") as well as more clerical work. In addition, junior jobs in 2004 contain system administrator positions and senior jobs in 2004 mention more programming languages (C++ and C).

Mechanical Engineering: Senior jobs are more likely to mention project management, designing and implementation. Junior jobs have more clerical (e.g., "update", "arrange", "email", "written"), computing (marked by "database", "Web", SQL, HTML, Java) and field-work, and requirement collection terms ("client", "custom", "meet"). Junior jobs in 2004 do not mention client interaction; instead they mention testing.

4.4.3 Clustering Analysis

Finally, we apply our clustering methodology to each discipline, both for 2014 and 2004. Our clustering results provide additional support for the findings in Section 4.4.1 and 4.4.2. Additionally, the main benefit of clustering is that it reveals

the different types of available jobs in each discipline. We discuss these findings below.

Finance: In 2014, the largest clusters were: several clusters mentioning finance-specific skills such as "trade", "equity", "tax", "reconciliation", "pension", asset valuation, risk management, "forecast", "causality" and "insurance" (63%); financial documentation (15%); and Web software development (10%). The jobs clustered under finance-specific skills were dominated by senior students, with the clerical (documentation) and IT (web development) clusters dominated by junior students. This result aligns with our analysis of significant differences from the previous section. Furthermore, in 2004, the largest clusters relate to financial analysis and documentation (51%), actuarial work including "valuation" and "pension" (18%), "tax" and "audit" (14%), and "causality" and "insurance" (5%). Thus, the 2004 clusters focus more on documentation and appear to describe a narrower range of jobs. All clusters except the last one mentioned have an equal split of junior and senior jobs.

Health Studies: The largest clusters in 2014 are related to organizing community events (21%), recreation camps for adults and children (14%) and therapy (13%), and are dominated by junior jobs. Smaller clusters dominated by senior jobs are related to research, cancer patient care and advanced aspects of health studies, including biomechanics, anatomy and statistics. The 10 clusters in 2004 are similar but exhibit equal proportions of junior and senior jobs in recreation, leisure and patient care.

Arts: The largest job clusters in 2014 include writing online content (24%), organizing events and providing customer service (22%), and writing, proofreading and summarizing research material (13%). These clusters have an almost even split of junior and senior jobs. Other clusters include project management (indicated by "stakeholder", "PowerPoint", "present"), market analysis ("campaign", "blog", "promote"), content writing ("Drupal", WCMS, standing for Web Content Management System), library liaisons and teaching (adult education, names of courses), which are dominated by senior students. Additionally, 52% of the jobs in 2004 fall in one cluster characterized by preparing English material for education and research on various topics including policy and politics. Other clusters include publishing newsletters and articles (with "graphics") (12%), office assistant positions (indicated by words such as "multitask", "file", "compile", "photocopy" and "fax") (8%), teaching and business analysis. Most of the clusters have an almost even split between junior and senior jobs. It appears that the Internet and social media have created new Arts jobs.

Biology: Our clustering results identify jobs in various fields of this discipline (microbiology, molecular biology, genetics, biochemistry), using various techniques (chromatography, electrophoresis).

Environmental Studies: The largest clusters in 2014 include project management (31%), education/research (25%), survey (18%), urban planning (13%) and advanced topics including GIS, cartography and geospatial analysis. (13%). On the other hand, half the jobs in 2004 mention educating people (largest cluster). While 8% of the jobs are related to advanced concepts, the other three clusters involve urban planning (20%), hydrogeology (14%) and waste water treatment (12%).

Chemical Engineering: Clustering 2014 Chemical jobs reveals additional insight: there is a cluster of jobs related to mechanical aspects of chemical plants, including the term "equipment". Additionally, a cluster with "nanotechnology", "lab", "material" and "physics" includes 10% of 2014 jobs. While 8% of the jobs are related to energy sources (including "oil", "gas", "petroleum", "sand" and "biofuel"), 5% of the jobs revolve around "emission", "environment", "pollution", "regulation" and greenhouse gases. Similar to 2014, 2004 clustering also contains clusters related to the mechanics of chemical plants, process improvement and research. It is interesting to note the differences in the field of application in both the years. While 2014 concentrates on nanotechnology, energy and emissions, 2004 deals with pharmaceuticals and waste water treatment.

Civil Engineering: Consistent with the previous section, junior students dominate the clusters including on-site field work (data collection and inspection), and senior students dominate the design clusters.

Electrical Engineering: The types of jobs in 2014 include System development (18%), web development (14%), electrical drawing (12%), PCB and circuit design (12%), system administration (9%), quality assurance (9%), simulation/research (8%), power (8%), embedded systems (8%) and research on advanced topics including transmitters, effect on climate, power grids, etc. (2%). In line with the findings of the previous section, there is a higher proportion of junior jobs in computing and system administration, and a higher proportion of senior jobs in core electrical clusters including circuit design and embedded systems. The main types of jobs in 2004 are related to power systems (26%), IT (19%), project management (18%), circuit design (15%), multimedia/graphics (6%), and transmission/telecommunication (4%).

Mechanical Engineering: Three-quarters of both 2004 and 2014 Mechanical Engineering jobs fall in the mechanical drawing cluster. While the other quarter of 2004 jobs mention plant-related terms including "assembly", "weld" and "motor", the other quarter of 2014 jobs is related to computing ("hardware", "automate", C++, Java, C, "web", "code"). Clustering 2014 jobs further reveals a 60-20-20 split among mechanical drawing, embedded systems and web development jobs.

To summarize, our clustering methodology identifies the types of available jobs in various disciplines. Through frequent term analysis, we found that soft skills and basic computing skills appear to be important in all disciplines in the 2014 job dataset.

5. DISCUSSION AND CONCLUSIONS

In this paper, we presented a text mining methodology to extract, compare and cluster important terms from freetext job descriptions. Our method identifies required skills as well as working environment and company culture descriptors. To demonstrate the utility of our methodology, we

analyzed a dataset containing nearly 30,000 undergraduate co-operative job postings from two years: 2004 and 2014. Our main findings are as follows.

- As expected in an undergraduate co-op marketplace, there are many assistant and junior positions, but less so in 2014 than in 2004.
- Basic computing skills are needed in almost all disciplines and at all levels. In other words, many non-IT disciplines appear to be trending towards IT.
- Soft skills are mentioned frequently by job postings from all disciplines, and more so in 2014 than in 2004. For example, over all disciplines, "team" was mentioned 20% more often in 2014 than in 2004. (in 71% vs. 51% of all job postings). These findings agree with those reported in [3, 8, 15]. Besides teamwork, communication and leadership were frequently mentioned in job postings across all disciplines, with IT postings additionally mentioning mindset-related terms (passion and love for the work), Finance jobs mentioning interpersonal relationships and Health Studies jobs recruiting active students.
- Regardless of discipline, lower-year positions were and are more clerical and/or involve more basic computing. Upper year positions tend to mention advanced concepts and solution methods.
- We identified several trends over time by comparing 2004 jobs with 2014 jobs. For example, IT jobs now emphasize mobile and cloud computing, Arts jobs involve social media and Chemical Engineering jobs mention sustainable energy.
- Job postings from different disciplines suggest different working environments: plants in Chemical and Mechanical Engineering, labs in Biology, and casual, fun and collaborative environments in IT.

We emphasize that our results should be interpreted carefully due to the following factors.

- Diversity in size and age of companies, e.g., the IT discipline has many modern companies that emphasize a fun work culture, while other disciplines such as Finance have more traditional companies which might emphasize client relationships.
- Incorrect job descriptions which may not reflect the true nature of the job; e.g., employers may write or modify the job descriptions to suit the company's public image.

Nevertheless, we believe that our findings are of interest to students, employers and the institution. We provide several examples of actionable insights below.

- We can provide students with a better understanding of co-op opportunities in various disciplines and therefore help them select the right academic program and career.
- In particular, we suggest that all students, regardless of discipline, acquire basic computer programming skills, which should help them secure co-op positions in their junior years.
- The institution can use our findings to manage the expectations of junior students. As we showed, it may take until senior years to obtain a co-op position that fully utilizes advanced discipline-specific skills.
- The institution may use frequently appearing job attributes and the clustering of jobs in various disciplines to produce more effective promotional material for its co-op programs and to help attract strong students.
- With the help of our findings, the institution can make an informed decision about how to change academic curricula to align with employers' needs. For example, as all disciplines seem to emphasize teamwork, the institution can incorporate more team exercises in the curriculum. Hackathons and other competitions could be organized to foster passion and other mindset-related skills for IT students, while mock client meetings could be arranged for Finance students so that they could hone their interpersonal skills. New tools and methods may be introduced in courses when the corresponding terms begin to appear in job descriptions.
- Employers may examine our findings to understand which skills are in high demand and to understand the extent of competition in the co-op market.
- Our lists of frequent attributes may be used to redesign the way employers submit job postings. For instance, separate fields (outside the job description) may be added for required skills and company culture descriptions, with drop-down lists populated with frequent terms obtained through our methodology. Additionally, our clustering methodology can be used to segment the job descriptions to make it easier for students to find jobs they are interested in.

Naturally, there is more data-driven work that can be done. The goal of a successful co-op system is to match the right student with the right employer. Thus, our long-term research objective is to minimize the gap between employers' needs and students' talents. In this paper, we focused on job descriptions, which provide an indication of what co-op employers are looking for and what working environments they offer. In future work, we will characterize what students have to offer by mining resumes. Furthermore, we plan to study the gap between what employers want and what is being taught in schools (e.g., by comparing job postings with course descriptions). Another interesting research direction is to determine if students are likely to obtain full-time jobs at one of their co-op employers after graduating. Finally, we are interested in comparing our job postings with those from other institutions worldwide. For example, the knowledge of foreign languages did not appear to be important in our dataset but it may be important in other countries.

6. REFERENCES

[1] A. Aken, C. Litecky, A. Ahmad, and J. Nelson (2010). Mining for computing jobs. IEEE Software, 27(1):78-85.

[2] A. Andrade, S. Chopra, B. Nurlybayev and L. Golab (2018). Quantifying the impact of entrepreneurship on co-operative job creation. International Journal of Work-Integrated Learning, 19(1):51-68.

[3] R. Bancino and C. Zevalkink (2007). Soft skills: The new curriculum for hard-core technical professionals. Techniques: Connecting Education and Careers (J1), 82(5):20-22.

[4] A. E. Barber and M. V. Roehling (1993). Job postings and the decision to interview: A verbal protocol analysis. Journal of Applied Psychology, 78(5):845-856.

[5] C. F. Chien, and L. F. Chen (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. Expert Systems with Applications, 34(1):280-290.

[6] R. K. Coll, K. E. Zegwaard and D. Hodges (2002). Science and technology stakeholders ranking of graduate competencies part 1: Employer perspective. Asia-Pacific Journal of Cooperative Education, 3(2):19-28.

[7] R. K. Coll, K. E. Zegwaard and D. Hodges (2002). Science and technology stakeholders ranking of graduate competencies part 2: Students perspective. Asia-Pacific Journal of Cooperative Education, 3(2):35-44.

[8] R. De Villiers (2010). The incorporation of soft skills into accounting curricula: preparing accounting graduates for their unpredictable futures. Meditari Accountancy Research, 18(2):1-22.

[9] M. Diaby, E. Viennet, and T. Launay (2013). Toward the next generation of recruitment tools: an online social network-based job recommender system. In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 821-828.

[10] C. Ding and X. He (2004). K-means clustering via principal component analysis. In Proceedings of the 21st International Conference on Machine Learning, 29.

[11] D. C. Feldman, W. O. Bearden and D. M. Hardesty (2006). Varying the content of job advertisements: The effects of message specificity. Journal of Advertising, 35(1):123-141.

[12] S. Ferns and K. Moore (2012). Assessing student outcomes in fieldwork placements: An overview of current practice. Asia-Pacific Journal of Cooperative Education, 13(4):207-224.

[13] J. L. Fleiss, B. Levin and M. C. Paik (2004). Determining Sample Sizes Needed to Detect a Difference between Two Proportions. Statistical Methods for Rates and Proportions, pages 64-85. John Wiley & Sons, Inc.

[14] J. Gault, J. Redington and T. Schlager (2000). Undergraduate business internships and career success: are they related? Journal of marketing education, 22(1):45-53.

[15] I. Grugulis and S. Vincent (2009). Whose skill is it anyway? Soft skills and polarization. Work, employment and society, 23(4):597-615.

[16] D. Hodges and N. Burchell (2003). Business graduate competencies: Employers views on importance and performance. Asia-Pacific Journal of Cooperative Education, 4(2):16-22.

[17] Y. Jiang and L. Golab (2016). On Competition for Undergraduate Co-op Placements: A Graph Mining Approach. In Proceedings of the 9th International Conference on Educational Data Mining (EDM), 394-399.

[18] Y. Jiang, S. W. Y. Lee and L. Golab (2015). Analyzing student and employer satisfaction with cooperative education through multiple data sources. Asia-Pacific Journal of Cooperative Education, 16(4):225-240.

[19] E. Malherbe, M. Diaby, M. Cataldi, E. Viennet, and M. A. Aufaure (2014). Field selection for job categorization and recommendation to social network users. In Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 588-595.

[20] E. Rainsbury, D. L. Hodges, N. Burchell and M. C. Lay (2002). Ranking workplace competencies: Student and graduate perceptions. Asia-Pacific Journal of Cooperative Education, 3(2):35-44.

[21] E. Ralph, K. Walker and R. Wimmer (2009). Practicum-education experiences: Post-interns' views. International Journal of Engineering Education, 25(1):122-130.

[22] C. L. Reeve and L. Schultz (2004). Job-seeker reactions to selection process information in job ads. International Journal of Selection and Assessment, 12(4):343-355.

[23] W. Song and S. C. Park (2007). A novel document clustering model based on latent semantic analysis. In Proceedings of the International Conference on Semantics, Knowledge and Grid, 539-542.

[24] S. E. Sorour, T. Mine, K. Goda, and S. Hirokawa. (2014). Efficiency of LSA and K-means in Predicting Students' Academic Performance Based on Their Comments Data. In Proceedings of the International Conference on Computer Supported Education (CSEDU), Volume 1, 63-74.

[25] S. Tang, X. Zhang, J. Cryan, M. J. Metzger, H. Zheng, and B. Y. Zhao (2017). Gender Bias in the Job Market: A Longitudinal Analysis. In Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), article 99.

[26] G. R. Thiel and N. T. Hartley (1997). Cooperative education: A natural synergy between business and academia. SAM Advanced Management Journal, 62(3):19.

[27] A. Toulis and L. Golab (2017). Graph Mining to Characterize Competition for Employment. In Proceedings of the Network Data Analytics workshop at the ACM SIGMOD Conf. on Management of Data, 3:1-3:7.

[28] TruncatedSVD [Computer software manual]. Accessed on 6 March 2018, at `scikit-learn.org/stable/modules/generated/sklearn.decomposition.TruncatedSVD.html`

[29] D. R. Young, D. N. Stengel, P. Chaffe-Stengel and R. M. Harper (2010). Assessing the academic and

workplace skills of undergraduate business interns. Journal of Cooperative Education and Internship, 44(1):13-22.

[30] K. E. Zegwaard and D. Hodges (2003). Science and technology stakeholders' ranking of graduate competencies part 3: Graduate perspective. Asia-Pacific Journal of Cooperative Education, 4(2):23-35.

[31] K. E. Zegwaard and D. Hodges (2003). Science and technology stakeholders' ranking of graduate competencies part 4: Faculty perspective. Asia-Pacific Journal of Cooperative Education, 4(2):36-48.