Job description mining to understand undergraduate co-operative placements

by

Shivangi Chopra

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Management Sciences

Waterloo, Ontario, Canada, 2017

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Co-operative education has become popular worldwide. In this thesis, we use a text mining methodology to analyze over 17,000 co-op job postings in order to understand the co-op market in a large post-secondary institution. First, we develop a parser that extracts informative terms from freetext job descriptions. These terms include soft skills, technical skills as well as perks and indicators of company culture. Second, we group the job descriptions by discipline and academic year and analyze the differences between various segments of the co-op market. We obtain insight that can benefit students, employers and the institution.

Acknowledgements

I would like to thank my supervisor, Professor Lukasz Golab, for his guidance throughout my graduate studies. I would also like to thank Professor Ken McKay and Professor Mark Smucker for spending time in reviewing my thesis.

I would also like to extend my gratitude to Judene Pretti and Rocco Fondacaro for the enriching discussions and insights into the co-op process.

Finally, I would like to thank my family for spending time in proofreading my thesis. I would not have been able to do this without your support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Co-operative (co-op) education is being adopted at a fast pace [71, 73]. Co-operative programs allow students to apply the concepts learnt in class to real world situations and make it easier to find employment upon graduation [44, 55, 66]. While institutions use co-operative education to provide an integrated learning environment [14, 28], it also helps attract new students. Employers use co-op as a talent pipeline. Due to its popularity, many researchers are studying various aspects of co-operative education [21, 22, 31, 36, 58, 64, 71, 73, 74].

In brief, the co-op process proceeds as follows. At the beginning of every semester, employers post job advertisements. Students apply to selected jobs and employers interview selected candidates. Finally, hiring decisions are made, and, at the end of the semester, students and employers may evaluate each other. In most programs, students alternate between study and work terms, with each work term possibly taking place at a different employer.

In a large co-op program, the participating entities may not have full knowledge of the job market. For example, students in different academic programs, especially junior students with limited work experience, may not know what employers are looking for and what types of jobs are available. Additionally, Coll et al. [22] surveyed students and employers to find that students' and employers' perspectives about the workplace competencies required by graduates entering the workforce differ; this points towards a gap between the students' and employers' understanding of the job market.

From the institution's viewpoint, it has been reported that co-op coordinators view service quality in the recruitment process more favorably than employers [19]. Apart from being dissatisfied by the low after-placement support provided to students, employers also reported low satisfaction levels in the co-op coordinators ability to suggest students based on personality fit, writing ability and oral communication [19]. As a result, the institution may not be aware of job market needs and thus, may lack the information to decide what types of new employers to attract. Furthermore, the institution may struggle to understand the talent needs of employers and adjust curricula if necessary. In fact, it has been reported that professors have strong views about required workplace competencies that differ from the employers' views towards the same [75], resulting in a gap between their understanding of the co-op market's needs.

Finally, from an employer's viewpoint, employers may not realize the extent of competition for students with various skillsets and may not be expressing their requirements clearly. This makes it difficult to attract top students. In fact, many researchers have outlined various suggestions to modify job descriptions in order to attract more applicants [6, 30, 32, 43, 59, 72].

In this thesis, we analyze the co-op market using job postings to help address the above problems. We mine over 17,000 co-op job descriptions from a large post-secondary institution. These job descriptions are posted directly by employers and 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 useful information from job descriptions and use it to understand the characteristics of the co-op market.

We address the above challenge by designing a parser that extracts job-related attributes from unstructured job descriptions. These attributes include technical and soft skills, work profiles, company culture, media presence, perks etc. By extracting informative attributes and comparing them across various fragments of the co-op market, we obtain interesting insights for three groups of stakeholders.

First, from the perspective of students, our results could inform them about the trends of the job marketplace, hence helping them make informed decisions about their careers and become more employable. *Second*, the institution could use our results to advertise the types of available co-op jobs and attract new students. Furthermore, the institution can use the knowledge of the co-op market to make informed curriculum decisions. *Third*, employers could use our results to understand the trends and competition for talent within their discipline.

To recap, the two contributions of this thesis are 1) a novel text-mining methodology for understanding a co-op job market and 2) a case study using a large data set from a North American university which showcases the utility of the proposed methodology. To the best of our knowledge, this is the first work to apply text mining to co-op data and the first work to analyze the characteristics of the market in the co-op context.

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

Chapter 2

Related Work

This thesis is related to two bodies of work: text mining and co-operative education. In the context of text mining, we use standard parsing and information retrieval techniques. We do not make any new algorithmic contributions in text mining; instead, our contribution is to apply these standard parsing and information retrieval techniques to a unique context to gain new insight. In terms of co-op education, our methodology is the first to extract the characteristics of the co-op market from job descriptions and enables new insight that, to the best of our knowledge, has not appeared before.

Research in co-op education revolves around students and the impact it has on the students' career growth. While many researchers study how co-op plays a vital role in a student's career in the long term [33], research has also been done about how co-op affects grades [10] and the development of soft skills [57]. While Blair et al. found co-op to have a positive impact on students' grades [10], other researchers revealed that co-op hones students' leadership skills [57]. Additionally, research has been done to characterize competition between students of different academic programs [38], revealing the similarities between the skills possessed by the students of the various academic programs. Competition among employees and employers has also been studied, indicating the most and least sought after jobs in the co-op market [68]. Furthermore, many studies explore how the variety of work experiences during co-op helps students to get a head start on their career [44, 55, 66]. Lastly, research has also been conducted on the methods used for assessing the overall co-op experience and learning it provides [31, 64, 76] and on students' satisfaction with the co-operative experience [39]. Jiang et al. [39] identified that students reported higher satisfaction levels when in leadership roles. While all the studies aim at improving the co-op process, most of them are prospective studies that examine the eventual impact of co-op, and do not delve into how students can prepare themselves to obtain better co-op jobs. Our research attempts to fill this scholarly gap and focuses on the trends of the co-op market, such that students are informed as to how they can equip themselves to gain better co-op jobs.

A few researchers survey the various stakeholders of the hiring process to identify factors or skills that are vital for a job [21, 22, 36, 51, 58, 74, 75]. Coll and Zegwaard conducted a four-part study in which they surveyed 172 employers [21], 71 students [22], 143 graduates [74] and 72 university faculty [75] of a science and technology industry to understand which workplace competencies they deem important for new graduates entering the workforce. Each group ranked a list of 24 workplace competencies (containing hard and soft skills) and the results revealed differences amongst the views of the various stakeholders. While students thought that they should have both hard and soft skills to increase employability, employers placed a higher emphasis on hard skills. While students' perceptions seemed to gravitate towards the faculty, graduates seemed to drift towards the employers' perspectives. Similar studies were carried out by Hodges et al. [36] and Rainsbury et al. [58] to understand the employers', students' and graduates' perceptions of the workplace competencies required by business graduates. Hodges et al. [36] noticed a performance gap between the expectations of the employers and the performance of the graduates in particular workplace competencies; e.g., written communication and technical skills [36]. Deriving their results from surveys, all the studies listed above are limited to the perceptions of the different groups of stakeholders. On the other hand, our research provides data driven insight.

In prior work, job descriptions have been used for three purposes. *Firstly*, employers use it to communicate their needs. Barber indicated that job descriptions should help prospective employees make decisions about applying to the job [5]. Hesketh found that employers preferred to describe the job roles using the desired skillsets instead of meta-data including academic program, degree, etc., as it made it easy for them to address the right audience [35].

Secondly, employers use job postings to attract applicants. A considerable amount of research has been conducted to study the contribution of job descriptions to attract more applications. While the works of Rynes et al. [61] provides a general approach for attracting more applicants, Barber et al. [6] and Reeve et al. [59] study the contribution of job descriptions, in particular, in attracting applicants. Barber et al. [6] collected verbal reports of potential applicants while they evaluated job descriptions to decide whether they should apply to the job or not. Barber et al. found that while location and compensation mattered the most in the applicant's decision-making process, the inferred probability to hire and the amount of information provided also played a role in their decision. While Breaugh argued that the level of accuracy and completeness of a job description attracts more applications [11], many studies report that the reputation of employers is the most important reason behind receiving applications [13, 23, 24, 32]. Moreover, some studies surveyed students to see how they responded to postings that were detailed and specific versus those that were general and vague [30, 60]; Roberson et al. [60] found that students preferred detailed job postings with specific recruitment information as it enhanced their perception of the organization's attributes and their person-organization fit. Yuce et al. examined the effect of the number of attributes contained in a job description and found that the higher the number of attributes mentioned (relevant or irrelevant), the more attractive it is to the reader [72]. Smith et al. re-ordered the valuable versus other information of a job posting to find that if the valuable information is presented first, it would increase the chances of the applicant's decision to apply [63]. While the research mentioned above is based on synthetic job postings, Barber emphasized the need to work with real job descriptions to reach to the real trends of the marketplace [5]. In 2007, Leung analyzed 127 real job descriptions and determined whether the presence of certain components of the job description attracted more applications [43]. Leung found that apart from the reputation of the employer and location of the job, the information quality of the job description affected its attractiveness the most.

Thirdly, online recruitment systems use job descriptions to find similarities between jobs and job seekers and provide suggestions to both parties to improve the hiring process. With the advent of online recruitment, a tool which can match applicants to employers is beneficial for both applicants as well as employers to help them narrow their search [27]. Diaby et al. [27, 50] use structured fields from social media accounts of job seekers to match them to the structured fields of the job descriptions. Suggestions are made based on the amount of similarity of the user profile and the job description, while taking into consideration the social connections of the user. Stephane et al. [65] did not use structured job postings and profiles. They extracted the required information including information about past work, education, hobbies, interests, etc., from user profiles and matched them with the extracted attributes of the job descriptions (technical requirement, company culture etc.) to suggest matches.

The above studies investigated how job descriptions could attract or match applicants, instead of using job descriptions as an independent resource to understand the needs of the employers. In our research, we propose a methodology to extract information from the job descriptions and use the job description itself to understand the co-op market.

To the best of our knowledge, this study is the first to use text mining in the context of co-op education. We apply parsing and text mining technologies to extract useful information from the job descriptions to understand the characteristics of the co-op market.

Chapter 3

Data and Methodology

Beginning with the necessary information regarding the co-operative (co-op) employment process of the university under study, this Chapter provides an overview of our data set in Section 3.1. We then present our methodology in Section 3.2 and end with a discussion of the limitations of the proposed methodology in Section 3.3.

3.1 Data

The university under study has three terms in each academic year. Undergraduate co-op at the university is dominated by programs in engineering, information technologies and finance. Other undergraduate programs also offer co-op, but their enrollment is much lower.

Undergraduate students enrolled in the co-op option alternate between academic and work terms every 4 or 8 months. At the beginning of each academic term, employers post job advertisements for the next term to an internal online portal. A job posting consists of a job title and a job description. Students can view all the job advertisements posted on the portal and apply to any job by submitting a resume. Cover letters, transcripts, previous projects and references may need to be provided if required by the employer. Students define their academic level based on the academic term they are currently enrolled in. Employers then conduct interviews and make offers. At the end of the work term, employers and students evaluate each other.

Our dataset consists of data from the 3 terms of 2014. It comprises of 17,057 undergraduate job postings that were advertised and filled during that year. As seen in Figure 3.1, for each job posting, we have the job title and job description. We also have the year of study and the program of the student who ultimately obtained each job (Figure 3.2).

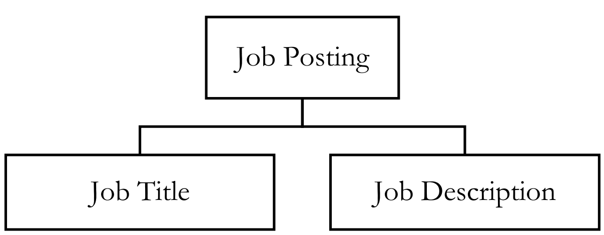

Figure 3.1 Information about a job posting

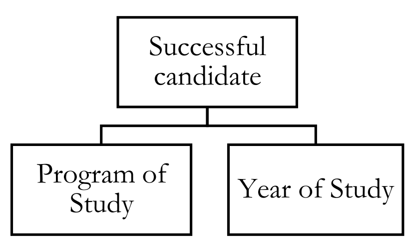

Figure 3.2 Information about a successful candidate

As shown in Figure 3.1, we have two text fields related to each Job Posting. A brief description of each of these fields is provided below:

1. Permitted to be 50 characters long, the Job Title generally consists of the position and/or the nature of the work. However, we observed that some job titles may include location or even a list of key job requirements. Common Job Titles include "Web Developer", "Engineering Intern" and "Planning Assistant".
2. The Job Description is an unstructured free text field that contains various details about the job. With no restriction on the length, employers list any information they want to communicate in any order they want. Some even format the job description using special characters. The information employers generally provide to the students through the Job Description include their company profile, the job profile (requirements, responsibilities), compensation (salary, perks etc.), contact information and other administrative details. We show a sample job description in Figure 3.3. To maintain data privacy, the job description in Figure 3.3 is not taken from the corpus but it matches the style seen in the real job descriptions.

Comprising of only the aforementioned fields, most job postings do not contain a target academic program or an industry code. Even if they do, the target programs and industry codes are often too general or incorrect. Instead, we use the characteristics of the successful candidate as a proxy for the targeted discipline of the job, as explained below.

As shown in Figure 3.2, we have two pieces of information about each candidate who successfully obtained a job.

1. Program of Study identifies the academic program the student was enrolled in when applying to the job.
2. The Year of Study represents the academic year of the student. In this thesis, Year 1 and 2 are considered to be the Lower Years while Year 3 and above are considered to be Upper Years of study.

The institution provided us with a mapping from academic programs to job disciplines. In this thesis, we study the three largest disciplines in the institution's co-op market: Information Technologies (abbreviated IT), Finance (abbreviated Fin) and Mechanical (abbreviated Mech). These three disciplines cover over 50 percent of jobs. IT includes academic programs such as Computer Science, Computer Engineering and Information & Technology Management. Finance includes Accounting and Actuarial Sciences. Mechanical includes Mechanical Engineering and Electronics.

3.2 Methodology

The goal of the thesis is to understand the co-op job market (with the help of job descriptions) and highlight the differences between disciplines and academic years. In order to achieve this, Section 3.2.1 describes the job descriptions and highlights the need for a parser. Section 3.2.2 introduces the parser's implementation. Further, Section 3.2.3 discusses the major groups of the co-op market and Section 3.2.4 introduces the ranking algorithm that will be used to find the trends of the market of a particular group. Finally, Section 3.2.5 introduces the tools that will be employed to compare groups.

3.2.1 Understanding job descriptions

Understanding the co-op job market is not limited to understanding the work profiles or technical skill requirements of its jobs; it also includes soft skill requirements, company profiles and culture, the administrative processes involved in obtaining a job as well as other requirements. For example, to understand the co-op job market of software jobs, we not only want to know what programming languages are in demand but also whether employers value certain soft skills, offer particular perks or shortlist students on the basis of particular attributes. We refer to these descriptive terms as job attributes and classify them into the following types.

1. *Specific Job Requirements* include technical skills (e.g. “java”), work profiles (e.g. “implementing a system”) and company profiles (e.g. “providing net banking solutions”).
2. *Soft Skills* include terms such as “teamwork”, “communication”, “passion” etc.
3. *Perks* include tangible (e.g. “free food”, “free transportation”) and intangible benefits (e.g. “fun”, “mentorship”) that may be part of a company’s culture.
4. *Media Presence* includes references to social media, magazines, television channels, etc.
5. *Admin* includes administrative requirements associated with applying for a job (e.g. “transcripts”, “past projects”)
6. *Insider* includes knowledge of specific courses of a university that are required by the jobs in the co-op market. They may also contain membership to specific clubs of the university.
7. *Internet Slang* includes casual instant messaging language.

Leung [43] lists 25 common components of a job description, summarized in Table 3.1, and suggests that a job description not only contains the above listed attributes (in its various components) but it also contains extra information related to the logistics/meta-data of the job. In Figure 3.3, we show a marked-up sample job description containing the above listed attributes and other information.

Table 3.1: Components of a job description [43]

S. No.	Components of a Job Description [43]	Type of Attribute
1	Job responsibilities [11, 16, 41, 70]	*Specific job requirements and/or Soft skills*
2	Coworkers [62]	*Specific job requirements and/or Soft skills*
3	Ethic identity [42]	*Specific job requirements and/or Soft skills*
4	Organization description and values [45]	*Specific job requirements and/or Soft skills*
5	Job qualifications [16, 41, 70]	*Specific job requirements and/or Soft skills and/or Insider*
6	Dress code [41, 62, 70]	*Specific job requirements and/or Soft skills and/or Perks*
7	Working hours [52, 70]	*Specific job requirements and/or Soft skills and/or Perks*
8	Working environment [11, 70]	*Specific job requirements and/or Perks*
9	Career development and support [41, 70]	*Perks*
10	Career path [5, 11, 41, 70]	*Perks*
11	Company surrounding environment [11, 70]	*Perks*
12	Housing [69]	*Perks*
13	Interesting work [41, 62, 70]	*Perks*
14	Local transportation [69]	*Perks*
15	Opportunities for promotion [5, 11, 41, 52, 70]	*Perks*
16	Travel requirement [41, 70]	*Perks*
17	Workforce Diversity [9, 11]	*Perks*
18	Organization reputation [5, 9, 11, 12, 32, 45, 61]	*Media Presence*
19	Prestige and recognition [70]	*Media Presence*
20	Application process information [41, 52, 70]	*Admin*
21	Compensation [11, 41, 61, 62, 70]	*Admin and/or Perks*
22	Supervisor [62]	*Meta-data*
23	City information [11]	*Meta-data*
24	Geographic location [5, 6, 11, 41, 52, 70]	*Meta-data*
25	Website information [18, 41, 52, 70]	*Meta-data*

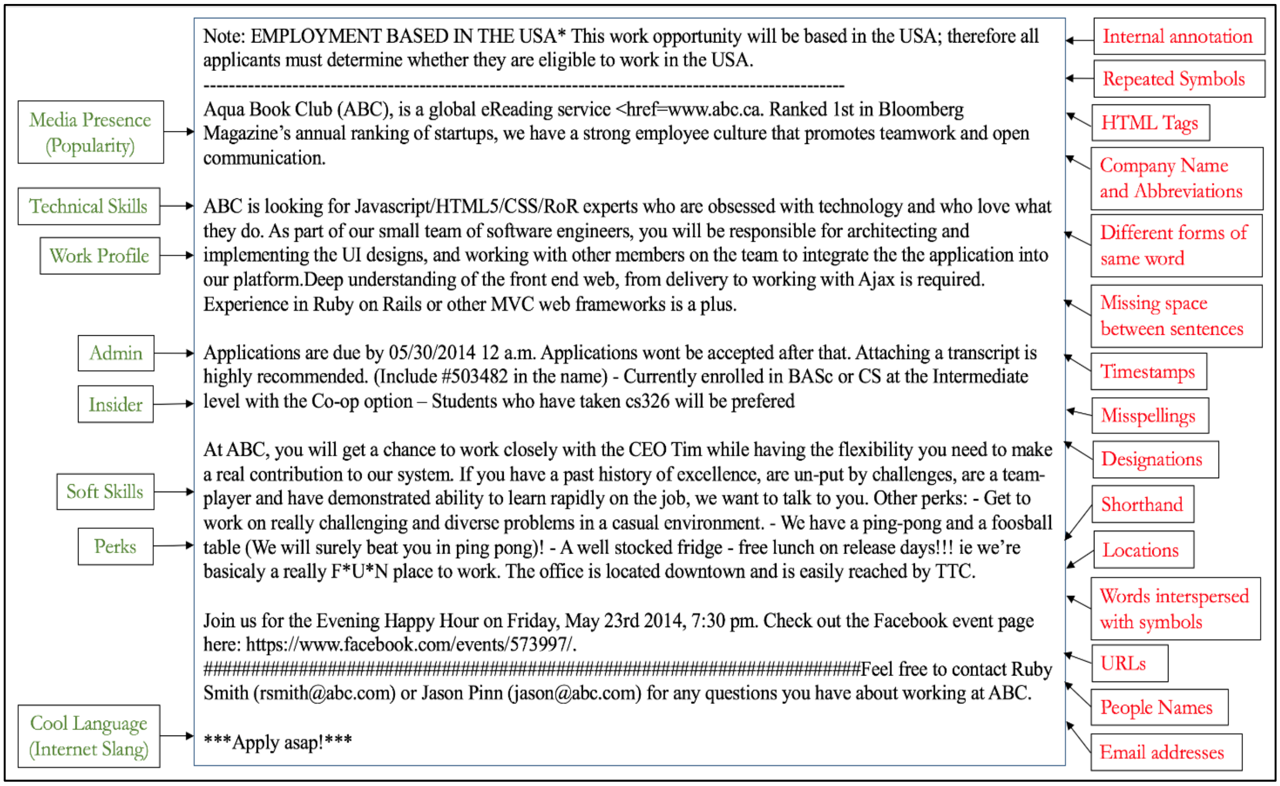

Figure 3.3 A sample job description

Apart from meta-data, Figure 3.3 suggests the need to eliminate other uninformative parts of a job description, including formatting, common English and inconsistencies/mistakes in writing. These elements arise as the job descriptions are free text inputs without any pre-defined structure and every employer writes the job descriptions as they see fit. Below, we list the elements we want to remove in order to identify informative job attributes.

Meta-data: These include words that are specific to the logistics of the company and the university. In line with the components of a job description that Leung [43] outlined, the Sample Job Description above includes meta-data such as person and company names ("Ruby Smith", "Jason Pinn", "Aqua Book Club"), abbreviations (ABC), locations ("downtown"), dates and times ("05/30/2014"), contact information ("rsmith@abc.com"), website URLs ("www.abc.ca") and internal notes appended by the institution.

Formatting: These include the printable and non-printable special characters that format the job description to give it a desired structure and/or flow [15]; e.g., ASCII control characters such as carriage return, line feed etc. [4]. Consecutive special characters that are used to divide the job description into sections and/or draw attention to specific parts of the job description are also considered part of Formatting. This can be seen in the Sample Job Description above (Figure 3.3). Other things considered part of Formatting include punctuation, special characters that are used as bullets (seen in the Sample to describe Perks), special characters embedded in words to increase emphasis (Sample contains an example: "F*U*N") as well as HTML tags [37] (the sample job description contains the HTML markup tag <href>).

English: Stopwords, common English words, Inflections, Derivations (with prefixes), Contractions (or compounded words), shorthand and abbreviations are all part of this constituent. As Stopwords [26], e.g., "are", "the" etc., and common English words, e.g., "able", "about", etc. [49], form the bulk of any natural language text, they are filtered out to improve query performance in a search engine [26]. As the job descriptions are also natural language text and we are searching for attributes embedded in them, we do the same. Inflections are different forms of the same word (with different word endings) to express tense, voice, number etc. [7]. Derivations are words with a prefix or a suffix attached to them [7]. Inflections ("implementing", "architecting", "obsessed") and Derivations ("un-put") can be seen in the Sample Job Description in Figure 3.3. It is common practice in Information Retrieval to standardize these forms into their root form using Stemming [26, 56]. Contractions or compounded words [7] are a shortened form of a group of words, e.g. "it's", "you're" etc. These can also be found in our Sample Job Description (e.g. "we're"). Finally, common shorthand notations ("i.e." in Figure 3.3) and abbreviations are also included in the English constituent.

Inconsistencies in Writing: These include common mistakes, shortcuts and different punctuation styles. Common mistakes that can be seen in the sample job description include misspellings ("prefered", "basicaly"), missing space between sentences ("…into our platform.Deep understanding..."), omitting dots in abbreviations ("ABC") and omitting special characters in contractions ("wont"). There could be many variations of writing the same pair or words, e.g., a ping pong table is mentioned twice in the sample job description of Figure 3.3 but is written differently each time ("ping pong" vs. "ping-pong"). Similarly, JavaScript could be written as "java-script", "java script" or "javascript". Different pairs of words could also be used to covey the same meaning; e.g., in the Sample job description, "teamwork" and "team-player" and "RoR" and "Ruby on Rails" communicate the same need. Finally, different spellings of the same word ("analyze" vs. "analyse") and the different meanings of the same word in different contexts ("Ruby on Rails" vs. "Ruby Smith") are also examples of inconsistencies.

Figure 3.4 summarizes the five constituents of a Job Description and emphasizes the need to remove the four aforementioned constituents to identify job attributes (and in turn the attributes of the co-op market). We developed a parser in Python to execute what is pictorially represented by Figure 3.4. Our parser eliminates Meta-data, English, Formatting and Inconsistencies in Writing to arrive at job Attributes.

3.2.2 Implementation of the parser to extract attributes from job descriptions

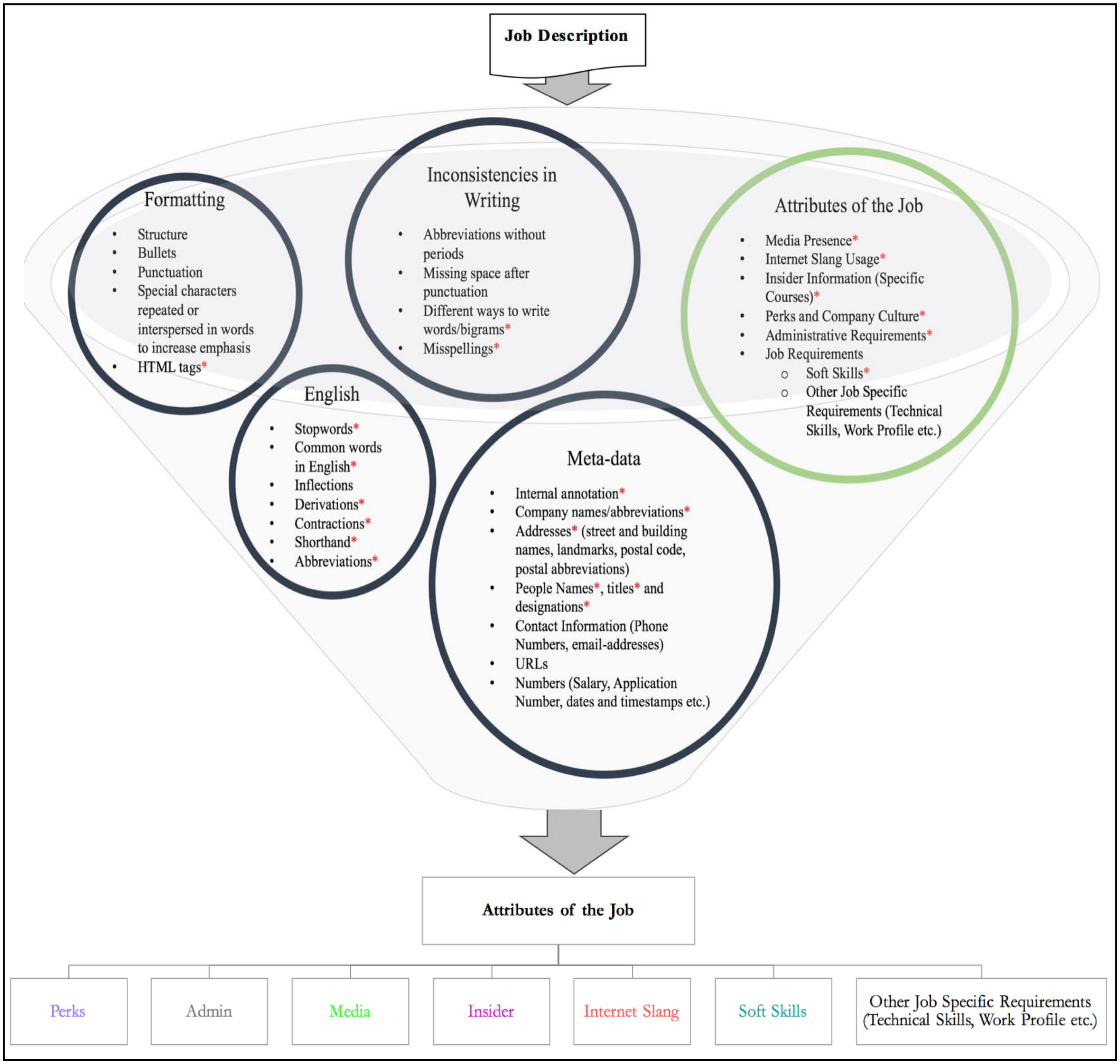

Figure 3.4 Constituents of a job description

To identify and thus eliminate uninformative elements of a job description, the parser requires external vocabularies. These are marked with an asterisk (*) in Figure 3.4. Not all vocabularies are available and therefore we created some of them manually. These are listed in Table 3.2. The "Internal annotation" vocabulary was provided by the institution.

Table 3.2 External Vocabularies

Constituents of a job description	Number of external vocabularies it requires	Existing Vocabulary	Manually Curated Vocabulary
Meta-data	6	Internal annotation	Company names/abbreviations
			Addresses (street and building names, landmarks, postal code, postal abbreviations)
			People Names
			Titles
			Designations
English	6	Stopwords [3]	Shorthand
		Common English words [49]	Abbreviations
		Derivations (Requires list of Prefixes) [53]	
		Contractions [48]	
Formatting	1	HTML tags [37]	
Inconsistencies in Writing	2	Common Misspellings [46]	Different ways to write words/word bigrams

Taking aid from the institution's internal databases, we manually created the proper nouns vocabulary of company names, addresses and people names as, to the best of our knowledge, there was no existing vocabulary which contained such information from around the world.

The external vocabulary of Abbreviations includes abbreviations of titles, designations, government institutions, businesses, academic disciplines and academic degrees [25]. We created this vocabulary by combining lists from various sources [1, 2, 40, 79] and revised it iteratively.

Finally, a list of different variations of the same words and bigrams was constructed. A mapping was built which used regular expressions [67] to convert all the forms into a single form. This list was built specifically for the co-op job market using domain knowledge and common occurrences in the job descriptions.

The manual vocabularies curated above may not necessarily be exhaustive. They are built to help remove unrequired material to arrive at the job attributes.

As the parser works by elimination, we need to be careful to not accidently discarding any useful words (attributes). For this, we create a seed list of words that are not to be eliminated. For example, "Ajax" is the name of a city in Canada as well as a programming language. Thus, "Ajax" appears in the proper noun list of addresses and would be eliminated. We include "Ajax" in the seed list to make sure it is not removed. Another example of a Specific Skill attribute sharing its name with a proper noun is the start-up company "Maple" and "Maple Software". Finally, an example of a proper noun sharing its name with a Soft Skill attribute is "teamwork" ("Teamwork Freight Solutions" is a company name).

The seed list is also required as many common English words that the parser would remove are Specific Skills; e.g., "analyze", "present", "write".

As summarized in Figure 3.5, our seed list contains common Specific Skill attributes and some common English words, e.g., "fun", "love". Note that the seed list only includes a subset of all possible Specific Skills.

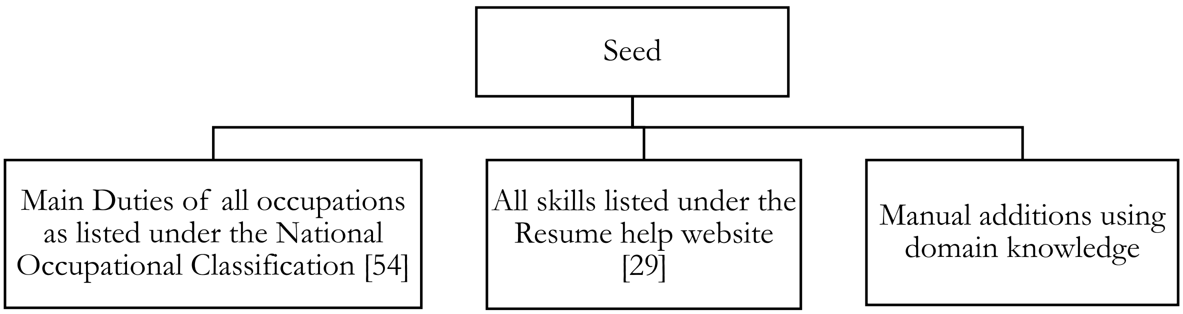

Figure 3.5 Components of the seed list

Once we have established the inputs required by the parser, Table 3.3 through Table 3.6 explain how the Parser handles different elements of a Job Description (defined in Section 3.2.1) and Figure 3.6 shows the sequence of operations carried out by the parser.

From here on, any reference to Tokens corresponds to the word forms returned by the parser after tokenizing the job description [26]. Further, any reference to an External Vocabulary in Table 3.3 through Table 3.6 corresponds to the External Vocabulary for handling that particular element.

Table 3.3 Operations of the parser that remove Meta-data

Meta-data	**Operation in Parser**	**Parser Procedure**
Internal annotation*	(Job Description – {External Vocabulary}) Using Regular Expression matching	Miscellaneous Filter
Company names/abbreviations*	{Tokens of the Job Description – {External Vocabulary – Seed}}	Discard Filter
Addresses* (street and building names, landmarks, postal code, postal abbreviations)	{Tokens of the Job Description – {External Vocabulary – Seed}}	Discard Filter
People Names*, titles* and designations*	{Tokens of the Job Description – {External Vocabulary – Seed}}	Discard Filter
Contact Information (Phone Numbers, email-addresses)	Remove sequences of numbers, sequences of numbers with special characters and email addresses from the Job Description using Regular Expression matching	Miscellaneous Filter

URLs	Remove URLs from the Job Description using Regular Expression matching	Miscellaneous Filter
Numbers (Salary, Application Number, dates and timestamps etc.)	Remove sequences of numbers and sequences of numbers with special characters from the Job Description using Regular Expression matching	Miscellaneous Filter

Table 3.4 Operations of the parser that remove English

English	Operation in Parser	Parser Procedure
Stopwords*	{Tokens of the Job Description – {External Vocabulary}}	Discard Filter
Common English Words*	{Tokens of the Job Description – {External Vocabulary – Seed}} {Lemma(Tokens of the Job Description) – {External Vocabulary – Seed}} Lemma is the root of a word [7]. As we want to remove all forms of common English words from the Job Description, the Parser removes any word whose lemma is in the Common English External Vocabulary.	Discard Filter
Inflections	Stem(Every token of the Job description) Using the Snowball Stemmer [56]	Stemmer
Derivations*	For every token of the job description, check if it starts with an item present in the List of Prefixes (external vocabulary). If yes, remove Token if (Token – Prefix item) is in the external vocabulary of Common English words*	Discard Filter
Contractions*	{Tokens of the Job Description – {External Vocabulary – Seed}}	Discard Filter
Shorthand*	{Tokens of the Job Description – {External Vocabulary – Seed}}	Discard Filter
Abbreviations*	{Tokens of the Job Description – {External Vocabulary – Seed}}	Discard Filter

Table 3.5 Operations of the parser that remove Formatting

Formatting	Operation in Parser	Parser Procedure
Structure	Remove ASCII control characters from the Job Description	Miscellaneous Filter
Bullets	Remove standalone special characters or numbers from the Job Description using Regular Expression matching	Miscellaneous Filter
Punctuation	Remove special characters separating sentences or words from the Job Description using Regular Expression matching	Miscellaneous Filter

Special characters repeated *or* interspersed in words to increase emphasis	Remove sequences of special characters surrounded by whitespace from the Job Description using Regular Expression matching	Miscellaneous Filter
	For every Token, if every other character of a Token is a special character: • remove Token if in external vocabulary Abbreviations • remove Token if the resulting token after concatenating consecutive alpha-numeric characters is in any external Vocabulary to be eliminated	Typo Filter
HTML tags*	{Tokens of the Job Description – {External Vocabulary}}	Discard Filter

Table 3.6 Operations of the parser that remove Inconsistencies in Writing

Inconsistencies in Writing	**Operation in Parser**	**Parser Procedure**
Abbreviations without periods	For every token, add periods (.) after all combinations of consecutive characters and remove Token if any combination matches an item in the external vocabulary Abbreviations	Typo Filter
Missing space after punctuation	For every token that contains a special character, split by the special character to form *x* resulting tokens and then {*x* resulting tokens – {All External Vocabularies to be eliminated – Seed}}	Typo Filter
Different ways to write words/bigrams*	Using Regular Expression matching, replace the multi-word tokens with one form for all items of the external vocabulary	Process multi-word tokens
Misspellings*	{Tokens of the Job Description – {External Vocabulary – Seed}}	Discard Filter

By sequentially applying the procedures illustrated in Figure 3.6, the parser removes the unrequired elements and keeps only the attributes. The output of the Parser is a set of unique tokens that correspond to one of the seven types attributes contained in the Job Description. An example of the Input (Job Description) and Output (attributes found in it) of the Parser is shown in Table 3.8. For completeness, the other inputs associated with the sample Job Description are shown in Table 3.7.

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!

Lower-case the entire Input and Process multi-word tokens

Apply the Miscellaneous Filter

- Structure
- Internal annotation*
- Contact Information (Phone Numbers, email-addresses)
- URLs
- Numbers (Salary, Application Number, dates and timestamps etc.)
- Bullets
- Punctuation
- Sequence of special characters to increase emphasis

Tokenize by whitespace

Apply the Typo Filter

- Abbreviations without periods
- Missing space after punctuation
- Special characters interspersed in words to increase emphasis

Apply the Discard Filter

- Misspellings*
- HTML tags*
- Stopwords*
- Common words in English*
- Derivations*
- Contractions*
- Shorthand*
- Abbreviations*
- Company names/abbreviations*
- Addresses* (street and building names, landmarks, postal code, postal abbreviations)
- People Names*, titles* and designations*

Stem using Snowball Stemmer

{"rank", "bloomberg", "promot", "servic", "magazin", "startup", "team", "communic", "love", "implement", "experi", "web", "ror", "javascript", "applic", "integr", "softwar", "ui", "html5", "mvc", "engin", "obsess", "contribut", "framework", "deliveri", "architect", "css", "design", "ajax", "platform", "transcript", "intermedi", "cs326", "recommend", "offic", "demonstr", "foosbal", "ttc", "stock", "system", "excel", "quicklearn", "histori", "problem", "releas", "pingpong", "lunch", "fun", "challeng", "divers", "flexibl", "event", "question", "asap"}

Figure 3.6 Process flow diagram of the parser

Table 3.7 Sample of the inputs associated with a job description

Input Required	
Program of Study of Successful candidate	Computer Science
Year of Study of Successful candidate	3
Discipline of the Job Posting	IT
Level of the Successful Candidate	Upper

Table 3.8 Sample input job title and description and its attributes extracted by the parser

Original Job Title and Description	**Parsed Job Title and Description**
Web Developer	{"web", "develop"}
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!***	{"rank", "bloomberg", "promot", "servic", "magazin", "startup", "team", "communic", "love", "implement", "experi", "web", "ror", "javascript", "applic", "integr", "softwar", "ui", "html5", "mvc", "engin", "obsess", "contribut", "framework", "deliveri", "architect", "css", "design", "ajax", "platform", "transcript", "intermedi", "cs326", "recommend", "offic", "demonstr", "foosbal", "ttc", "stock", "system", "excel", "quicklearn", "histori", "problem", "releas", "pingpong", "lunch", "fun", "challeng", "divers", "flexibl", "event", "question", "asap"}
Total number of tokens in the job description	354
Total number of distinct tokens in the job description	235
Number of attributes of the job description	54

Following the same nomenclature as in Section 3.2.1, we use manually-created vocabularies to label the attributes returned by the Parser as *Perks, Admin, Media, Insider, Internet Slang or Soft Skills*. These vocabularies, though not exhaustive, provide a way to segment the attributes. Borrowing from the lists found on various online sources, these lists are iteratively revised using domain knowledge and with help from co-op experts at the university. The vocabularies are shown as word clouds (Figure 3.7 to Figure 3.10) where the size of the words represent their frequency in our corpus of job descriptions.

Gathering information from various online sources [77, 78], the *Perks* vocabulary is built iteratively to include 52 perks which are valued by co-op students (Figure 3.7). *Perks* also contains attributes that describe company culture. The *Admin* vocabulary, containing 39 tokens, is solely created using domain knowledge (Figure 3.8). It contains attributes related to the application process, pre-requisites for the job and other administrative aspects that do not describe the nature of the work at the job. *Media* (Figure 3.9) borrows from an online list of social networking websites [47]. Names of commonly occurring television channels and magazines are added to this list to contain a total of 211 tokens. Out of the 2500 slang words available in an online list [80], our job description corpus contains 31; they are categorized under *Internet Slang* (Figure 3.10). The *Insider* vocabulary contains the courses and clubs of the Institution (list provided by the institution). Its word cloud has, thus, been omitted for data privacy. The *Soft Skills* vocabulary borrows from a resume help website [29] and is iteratively revised in consultation with co-op experts to reflect the soft skills that co-op employees value. It contains 94 tokens (Figure 3.11). The job attributes that occur in none of these vocabularies are assumed to belong to *Specific Job Requirements*.

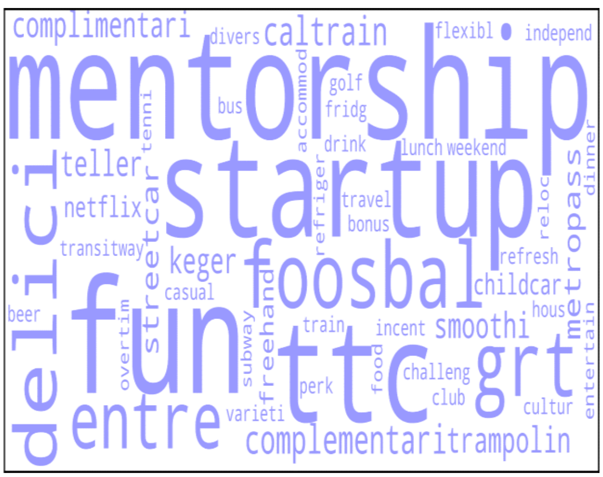

Figure 3.7 External Vocabulary of Perks

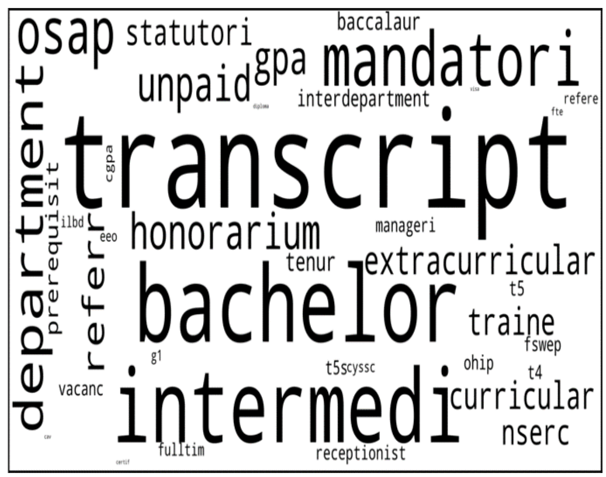

Figure 3.8 External Vocabulary of Admin

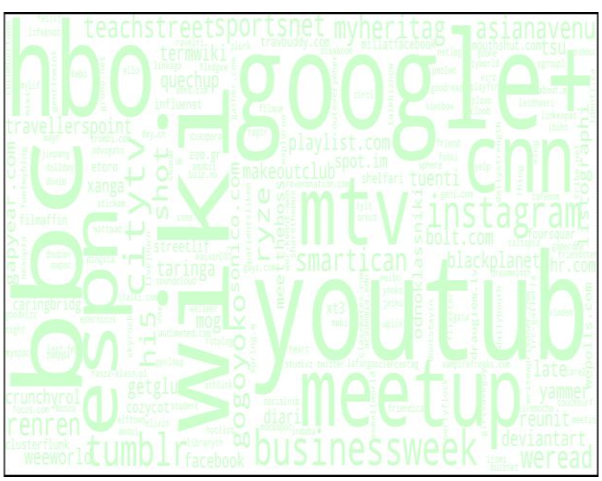

Figure 3.9 External Vocabulary of Media

Figure 3.10 External Vocabulary of Internet Slang

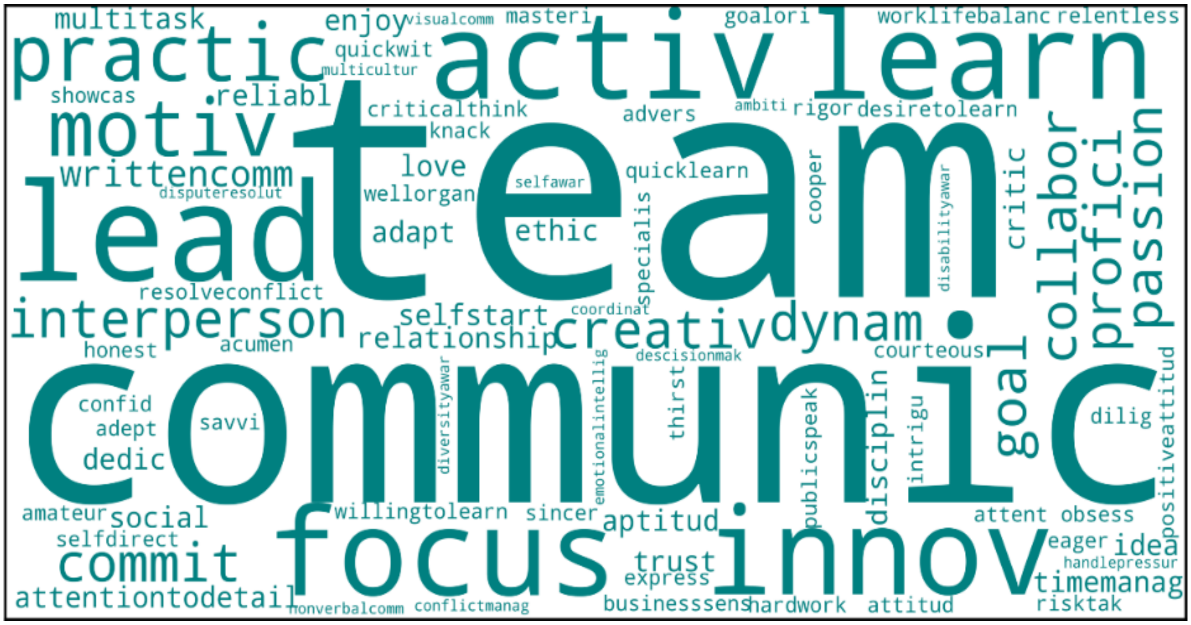

Figure 3.11 External Vocabulary of Soft Skills

All 17,057 job descriptions are parsed and labelled as outlined above. A vocabulary containing all the attributes of the co-op market is generated by parsing all the job descriptions and listing all the unique attributes that exist in at least 10 of the job descriptions. Let this Vocabulary of attributes of the job descriptions be represented by V and its size be defined by |V|. For each unique attribute, the Document Frequency (DF) is calculated as the number of job descriptions that contain that attribute, thus, quantifying how common the Attribute is in the corpus. The Inverse Document Frequency (IDF) is another common metric used in Information Retrieval to quantify the popularity of a word (in our case, Attribute) in the corpus [26]. The higher the IDF, the rarer the word is in the corpus. Where N is the total number of documents in the corpus (in our case N = 17,057 as we have 17,057 job descriptions), IDF is defined as:

$$\mathrm{IDF}_i = \log\left(\frac{N}{DF\ of\ Attribute\ i}\right) \tag{3.1}$$

These metrics quantify how common or rare each attribute is in our corpus and in turn in the co-op market. Apart from helping us understand the vocabulary of attributes, these metrics are a precursor to the methodology in Section 3.2.4. The process of generating the vocabulary of attributes is summarized in Figure 3.12.

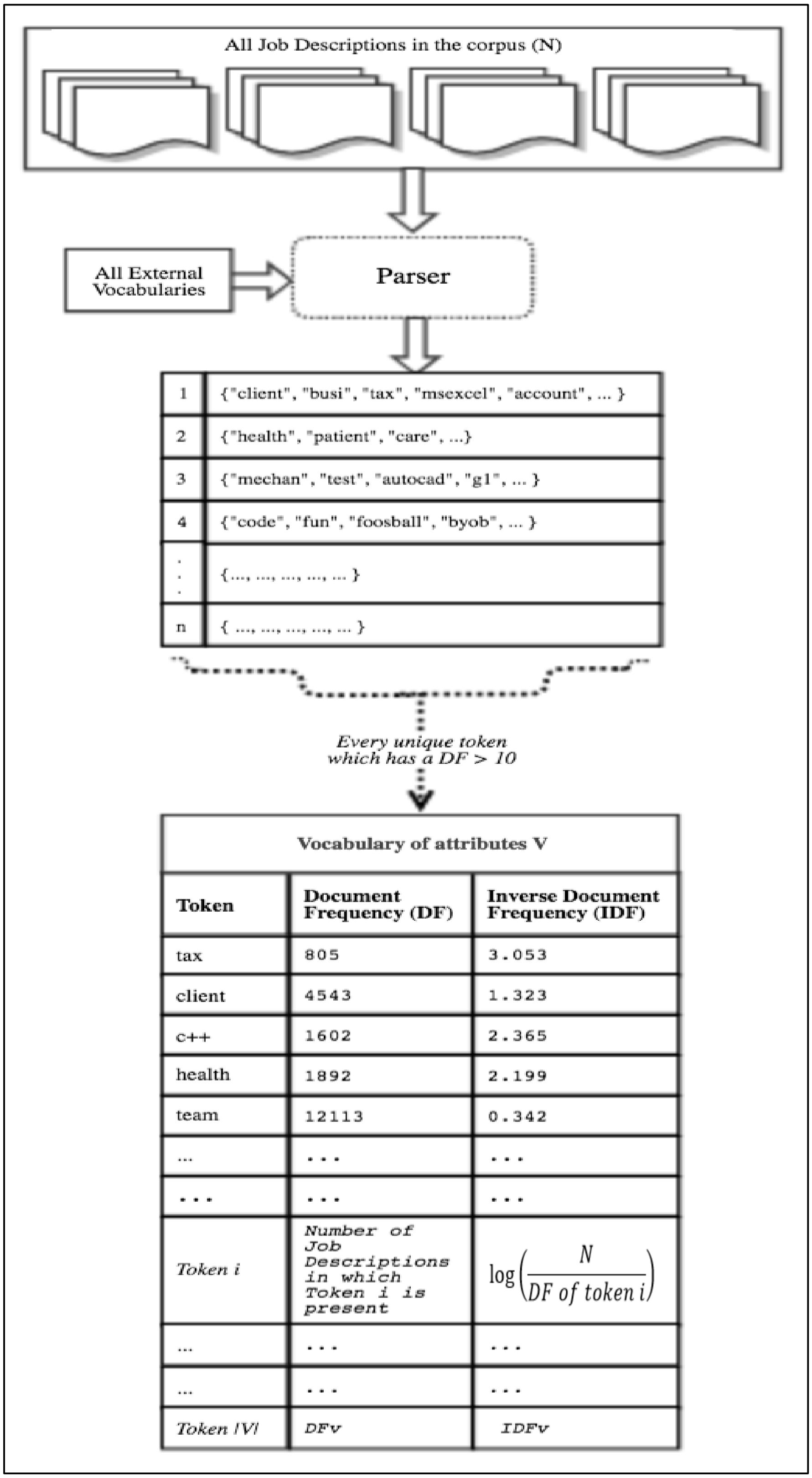

Figure 3.12 Process of generating the vocabulary of attributes

Another precursor step for the next section includes representing the attributes of each job description in vector form [26]. This is done by the Vectorizer and the process is shown in the Figure 3.13 below.

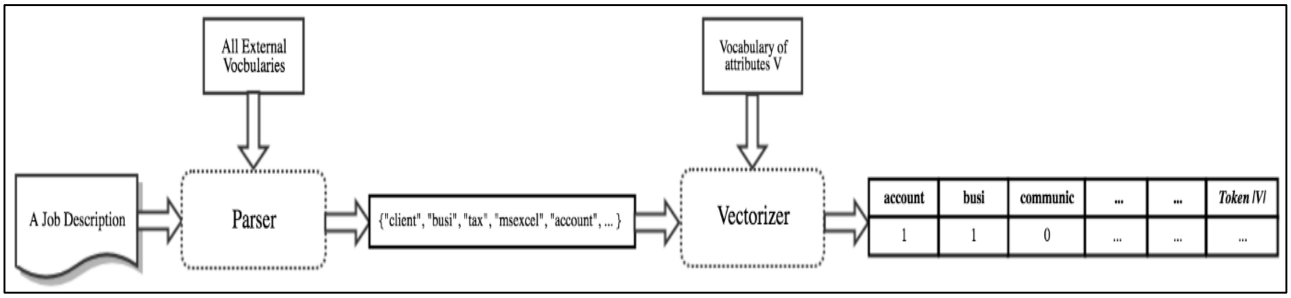

Figure 3.13 Process of converting a job description to a vector of attributes

As seen in Figure 3.13, the Parser converts a free text Job Description into a set of attributes and a Vectorizer then converts it into a vector which shows whether each attribute of the Vocabulary V is present or absent in the particular Job Description.

Even though the above processes are defined in terms of job descriptions, the same can be applied to Job Titles. As Section 3.1 mentions that some Job Titles contain other information besides Attributes, the above processes are run on the Job Titles too to give a Vocabulary of attributes found in Job Titles W. Unless stated explicitly, "Attributes of the Job" refers to the attributes found in their job descriptions and not their job titles.

3.2.3 Grouping the job descriptions

Having explained the tools used to extract attributes from job descriptions, we now outline our methodology to rank the attributes for a particular group of Job Descriptions defined by discipline or academic level of successful candidates.

We segment job descriptions into various Groups as follows.

Academic Discipline: By analyzing each academic discipline and comparing them, we want to answer questions such as "Are software skills becoming important in non-IT jobs?". As mentioned earlier, we label each job description with the Academic Discipline of the student who obtained the job. For example, the job descriptions of the jobs that were obtained by Finance students belong to the Finance Group. The Attributes of Finance refer to the attributes found in the job descriptions of the Finance Group.

Level of Study: Prior work has identified differences between co-op jobs for junior and senior students [17], and we want to use our dataset to confirm these; e.g., "Do lower-year students get more entry-level jobs than upper-year students?". Again, as most job postings do not specify the desired academic level of the student, we use the year of study of the successful candidate as a proxy.

From here on, Attributes of the Job Descriptions belonging to a given Group are referred to as Attributes of the Group.

Next, Section 3.2.4 develops a methodology to extract the attributes of a Group from its job descriptions. Section 3.2.5 explains how to compare the attributes associated with two Groups.

3.2.4 Ranking the attributes of a group of job descriptions

A Group contains a subset of jobs and the Job Descriptions associated with them. While the Parser can extract the attributes from job descriptions, not all attributes may be important to the Group. For example, if a Finance student secures a co-op job related to Biology, then the attributes of that Biology job would not represent the Finance group. To understand the job market within each Group, we not only need to extract the attributes from their job descriptions, but we also need to identify those which are important to the Group.

The notion of importance of an attribute is two-fold.

a) Identifying attributes that are widely demanded by many jobs in the Group helps understand the general trends of its market (referred to as *Frequent attributes*),

b) It is as important to know the specific attributes that differentiate the Group from other Groups (referred to as *Representative Attributes*).

For example, "auditing" is a skill that represents Finance as students from other disciplines are not likely to possess that skill. On the other hand, "java" may be a frequent attribute in the IT group.

We use

a) Term Frequency (TF) [26] to calculate Frequency and the
b) Term Frequency * Inverse Document Frequency (TFIDF) [26] to calculate Representativeness

While the TFIDF scores provide a simple ranking function used in Information Retrieval, they are also used by many text-based recommended systems [8], making them one of the most popular term-weighting schemes. The TF represents the number of times a word appears in a document, thus showing its importance in the document [26]. The TFIDF score offsets the TF with the IDF, which as defined in Section 3.2.2 represents the importance of a word in the entire corpus. Thus, the TFIDF score represents the weight of the word in the document [26] i.e. how essential is the word in defining the document. As we need to calculate the Frequency and Representativeness of an attribute in the entire Group, we need to consider all the job descriptions of the Group as a single document before calculating the TF and TFIDF scores of each attribute.

Based on our application, we make a slight variation to the above definition. Even though the repetition of an attribute within a job description might emphasize its importance for that job, it does not communicate anything about the importance of the attribute for the entire Group. For example, if a job requires a skill, e.g., "Word Perfect" and mentions it five times within the job description, it does not imply that "Word Perfect" is important to the entire Group. It simply means that the particular skill is essential for that particular job. Considering that we want to measure the trends of the co-op market of the entire Group, the repetition of an attribute within a job description should not be accounted for while calculating an attribute's importance in the Group. Thus, for measuring the importance of an attribute in a Group, each job description is reduced to its distinct attributes and then made part of the document containing all the job descriptions of the Group. The TF and TFIDF score is calculated for every attribute in the vocabulary according to the definition above. The IDF for all of V is calculated during the generation of the vocabulary of attributes of the job descriptions (Section 3.2.2).

The metrics Frequency and Representativeness are derived from the TF and TFIDF score. For example, if "teamwork" is required by 90 out of the 100 jobs in IT, it is said to have a Frequency of 90%. Sorting the attributes of a Group from their highest to lowest Frequency gives us a ranked list of the most Frequent Attributes of that Group. Now, let us say "teamwork" is required by 90 out of the 100 jobs in IT and 80% of the corpus in general. Then "teamwork" does not distinguish IT jobs from the rest. Sorting the attributes of a Group by their highest to lowest TFIDF scores gives us a ranked list of the most Representative Attributes of that Group.

Figure 3.14 provides an overview of the method used for identifying the Frequent and Representative attributes of a Group. Taking N_x Job Descriptions belonging to a Group as input, Figure 3.14 shows how to extract the most frequent and/or representative attributes.

As shown in Figure 3.14, the output can be interpreted in an ordered or unordered fashion. The Ordered output (also referred to as Ranked Lists) can be obtained by sorting the attributes of a Group by the metric required by the application. Once sorted, all or the Top K elements of the sorted list can be considered as the Ordered output. Removing the order from the Ordered output and considering all its attributes as a set constitute the Unordered output (also referred to as Sets).

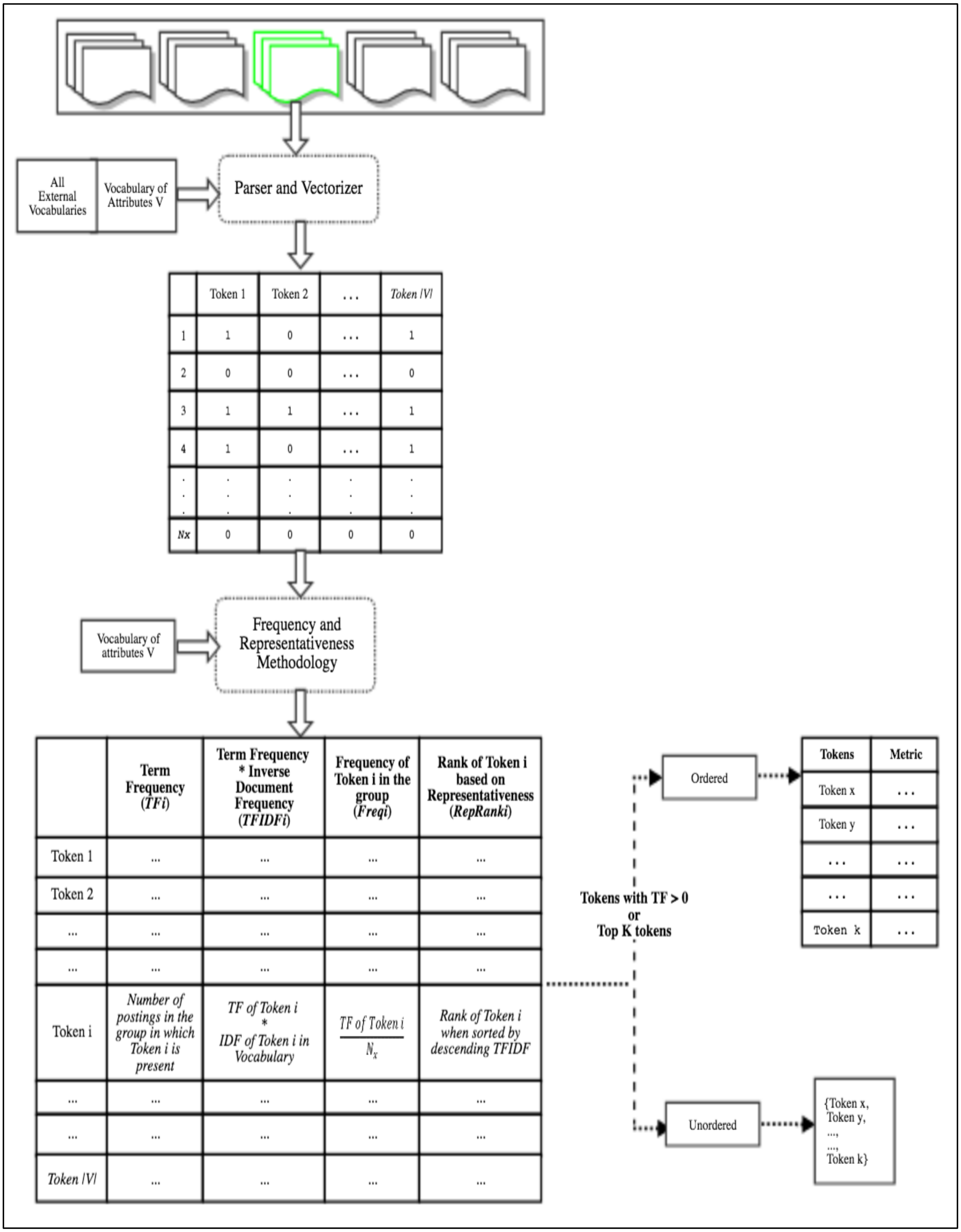

Figure 3.14 Overview of the method for identifying Frequent and Representative attributes of a Group

3.2.5 Comparisons of two groups of job descriptions

To analyze the differences between two Groups, we compare their Top 100 most Frequent and their Top 100 most Representative attributes. Even though we do not expect much overlap between the Top 100 Representative Attributes of two Groups (as they define the Group and thus, would not have much importance in other Groups), we compare them for completeness.

We make these comparisons using the following tools.

- **Venn Diagrams:** We take the Top 100 (Frequent or Representative) attributes of the two Groups and represent them as Venn Diagrams [67].

- **Jaccard Similarity (JS):** Using the Top 100 (Frequent or Representative) attributes of the two Groups, we calculate their Jaccard Similarity [26]. Ranging from 0 to 1, with 1 being most similar, JS provides a quantitative measurement of similarity.

- **Distribution of types of attributes:** Recall that the different types of attributes are Specific Job Requirements, Soft Skills, Perks, Admin, Internet Slang, Insider and Media (defined in Section 3.2.1). We will compare the distributions of these among the Top 100 (Frequent or Representative) attributes of two groups.

- **Difference in Frequency of attributes:** We also compare two Groups by identifying attributes whose frequency in one group is much lower or higher than in the other.

Figure 3.15 summarizes our techniques for comparing the attributes of two Groups.

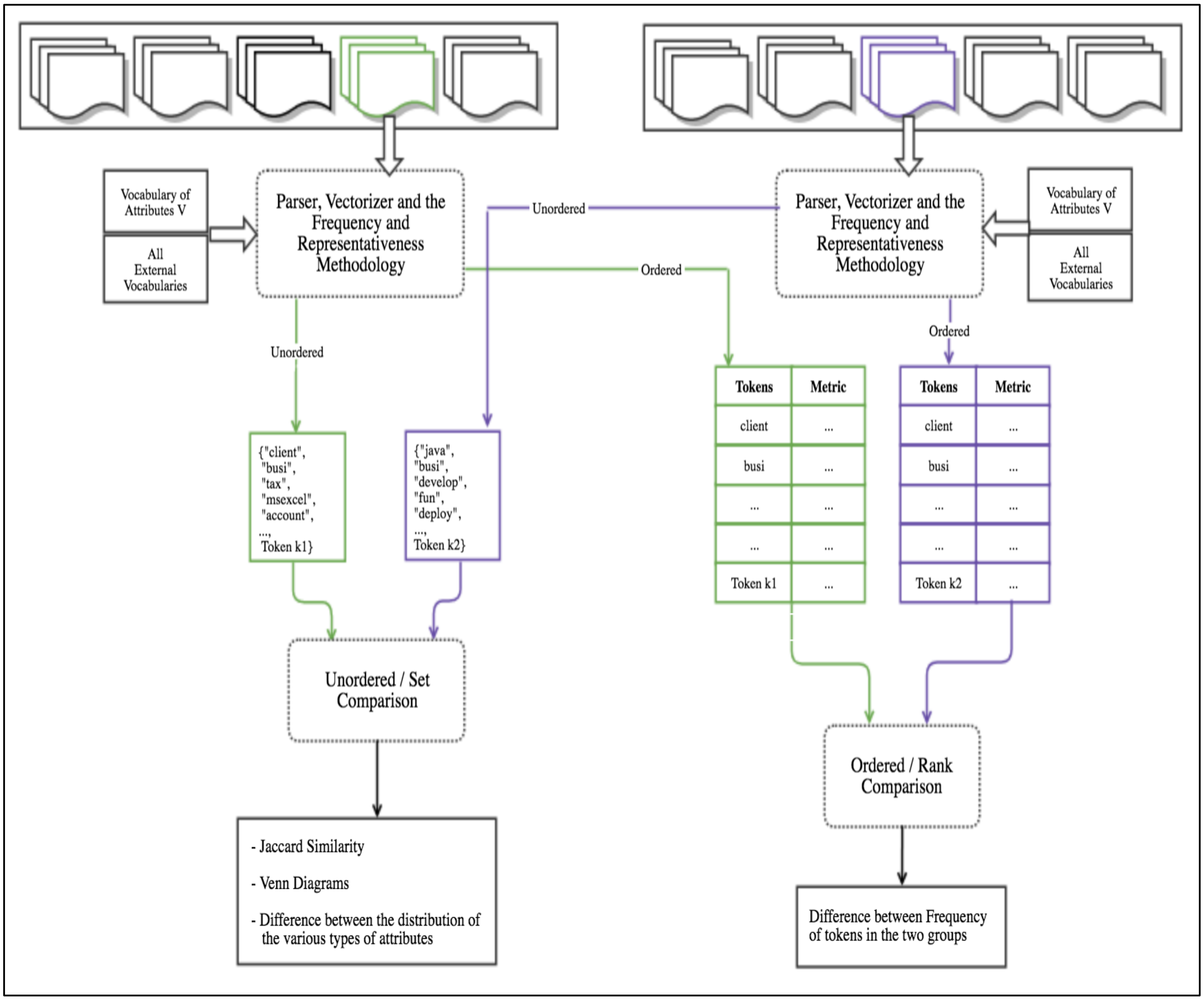

Figure 3.15 Comparing two groups of job descriptions

3.3 Limitations and Assumptions

Our results have two key limitations. First, we do not know whether job description communicate the actual nature of work. Second, a job description is considered to be part of a Group based on the characteristics of the successful candidate. Even though we are unaware of an employer's rationale in selecting the particular student, the student's rationale in taking that job and the student's performance on the job, we know that the successful candidate was selected from a pool of competing students. Thus, we assume that the student was qualified for the position.

Chapter 4

Results and Discussion

This chapter begins with an analysis of the job titles and job descriptions of the entire corpus to understand the general characteristics of the co-op market (Section 4.1). In Section 4.2, we study the trends of the three main disciplines and compare them. We conclude with an investigation of how lower year jobs of a discipline differ from their upper year counterparts and identify the general trend in lower and upper year jobs (Section 4.3).

4.1 Attributes associated with the entire job corpus

For an overview of the co-op job market as a whole, we examine the attributes present in the job titles of all the 17,057 job postings. Figure 4.1 illustrates a word cloud of the attributes that appear in at least 10 job titles; the higher the frequency, the larger the font. Table 4.1 corresponds to the word cloud of Figure 4.1 and provides the frequency of the top 25 most frequent attributes in the job titles of the corpus.

Figure 4.1 Word cloud of attributes occurring in job titles sized by frequency

Figure 4.1 suggests that "engin" (representing words like "engineer", "engineering", "engine" etc.), "assist", "develop", "research", "software" and "analyst" are the most common, while some "test" positions are also present.

With "assist" (representing "assistant", "assisting", "assistance" and "assist") being mentioned in many job titles, we hypothesize that many co-op positions are junior positions. As seen in the word cloud of Figure 4.1, the co-op market also has some "specialist" positions, but they are more rare than the "assist" co-op positions. Table 4.1 indicates that 2% of the job titles mention "specialist" while 19% mention "assist". Zooming into lower and upper year positions will verify this, which we will do in Section 4.3. On a similar note, more job titles mention "support" than "manage".

Attributes related to the Fin, IT and Mech disciplines (including "engin", "manufacture", "lab", "web", "software", "analyst", "account", "actuari" etc.) also appear in the word cloud (Figure 4.1). As seen in Table 4.1, some of these attributes are even part of the Top 25 most frequent attributes of the corpus. This is because of their noticeably bigger size in the co-op of the institution we are studying. Hence, we focus only on these disciplines in this thesis.

As shown in Figure 4.1, attributes labelled as soft skills (marked in green in the word cloud) can also appear in job titles.

Next, we examine the attributes that appear in the job descriptions (Figure 4.2 and Table 4.2).

Figure 4.2 Word cloud of attributes occurring in job descriptions sized by frequency

As seen in Table 4.2, "experi" and "develop" are the most frequent attributes found in the job descriptions. Notably, "develop" is mentioned more often than "test" (71% of the job descriptions contain "develop" in comparison to the 30% that contain "test").

“experi” represents the different forms of the word “experience” as well as “experiment”. This is an artefact of using stems of words to represent attributes instead of whole words. Thus, the size of “experi” represents a combined frequency of “experience” and “experiment” in the corpus. Other attributes that might have been affected due to stemming or the lack of context include “excel”. While “excel” represents the different forms of the word “excellent”, it might also include the software name “Excel” if written without some form of the word “Microsoft” preceding it. Various forms of the software “Excel” including “MS Excel”, “Microsoft Excel” etc. have been converted to the attribute “msexcel” in the Process multi-word tokens filter of the parser (Section 3.2.2).

Furthermore, Figure 4.2 indicates that soft skills such as “team” and “communication” are frequent while terms related to mindset, such as “motivation”, “learn”, “passion”, “selfstarter” and “dynamic”, are less frequent. It is interesting to note that more than 70% of the jobs in the corpus require teamwork skills. While past research identifies employers’ emphasis towards soft skills using survey data [20, 34, 36], Figure 4.2 and Table 4.2 provide data-driven evidence of this.

Figure 4.2 and Table 4.2 also suggest that while “assist” and “support” are frequent, “manage” and “lead” are less frequent in the co-op job corpus.

Table 4.1 Top 25 Frequent attributes in the job titles

S. No.	Token	DF (↓)	Frequency in Corpus
1	**engin**	3427	20%
2	**assist**	3296	19%
3	**develop**	2381	14%
4	**softwar**	1985	12%
5	**analyst**	1872	11%
6	**research**	1383	8%
7	**design**	613	4%
8	**busi**	581	3%
9	**account**	578	3%
10	**support**	507	3%
11	**technician**	488	3%
12	**project**	452	3%
13	**system**	410	2%
14	**actuari**	374	2%
15	**qualiti**	374	2%
16	**program**	365	2%
17	**applic**	361	2%
18	**specialist**	359	2%
19	**product**	355	2%
20	**manag**	353	2%
21	**web**	338	2%
22	**servic**	334	2%
23	**market**	330	2%
24	**oper**	310	2%
25	**lab**	296	2%

Table 4.2 Top 25 Frequent attributes in the job descriptions

S. No.	Token	DF (↓)	Frequency in Corpus
1	**experi**	12946	76%
2	**develop**	12164	71%
3	**team**	12113	71%
4	**communic**	9261	54%
5	**project**	9122	53%
6	**program**	8981	53%
7	**assist**	8312	49%
8	**applic**	8283	49%
9	**design**	8186	48%
10	**excel**	8177	48%
11	**manag**	8050	47%
12	**product**	7751	45%
13	**support**	7662	45%
14	**engin**	7220	42%
15	**softwar**	7137	42%
16	**system**	6952	41%
17	**busi**	6823	40%
18	**report**	6683	39%
19	**servic**	6656	39%
20	**process**	6550	38%
21	**lead**	6402	38%
22	**learn**	6208	36%
23	**data**	6159	36%
24	**perform**	6155	36%
25	**comput**	5930	35%

Overall, the results in this Section indicate that many co-op jobs appear to be assistant or junior positions, and that teamwork and communication are important to many jobs.

4.2 Attributes associated with each discipline

In this section, we first examine the attributes associated with the job titles and job descriptions of the three disciplines: Fin (Section 4.2.1), IT (Section 4.2.2) and Mech (Section 4.2.3). After understanding which attributes are frequent and representative in each discipline, Section 4.2.4 compares the disciplines based on their Top 100 frequent or representative skills.

4.2.1 Finance job analysis

We begin with the attributes present in the Job Titles of Finance. Sized by the frequency and representativeness, respectively, Figure 4.3 shows the most frequent attributes and Figure 4.4 shows the most representative attributes. Table 4.3 and Table 4.4 provide the corresponding metrics (frequency and representativeness rank) that have been used to size the attributes of Figure 4.3 and Figure 4.4, respectively. We obtain the following insight.

- With few "specialist" and "manager" positions, Finance has frequent "trainee", "support" and "assist" positions.
- Job titles tend to specify the level of student who should apply, e.g., "intermediate".
- Even though they are not frequent, "program" and "software" appear Finance, which could indicate a trend towards IT.
- Specific financial skills include "analyst", "cpa" (Certified Professional Accountant) and "actuari" (representing actuary). Jobs related to "account" (representing "accounting", "accounts", "accountants" etc.), "audit", "tax", "risk management", "business", "market", "bank", "treasuri", "pension", "equity" and "capital" also seem to be popular.

Figure 4.3 Word cloud of all the attributes of the job titles of Fin sized by frequency

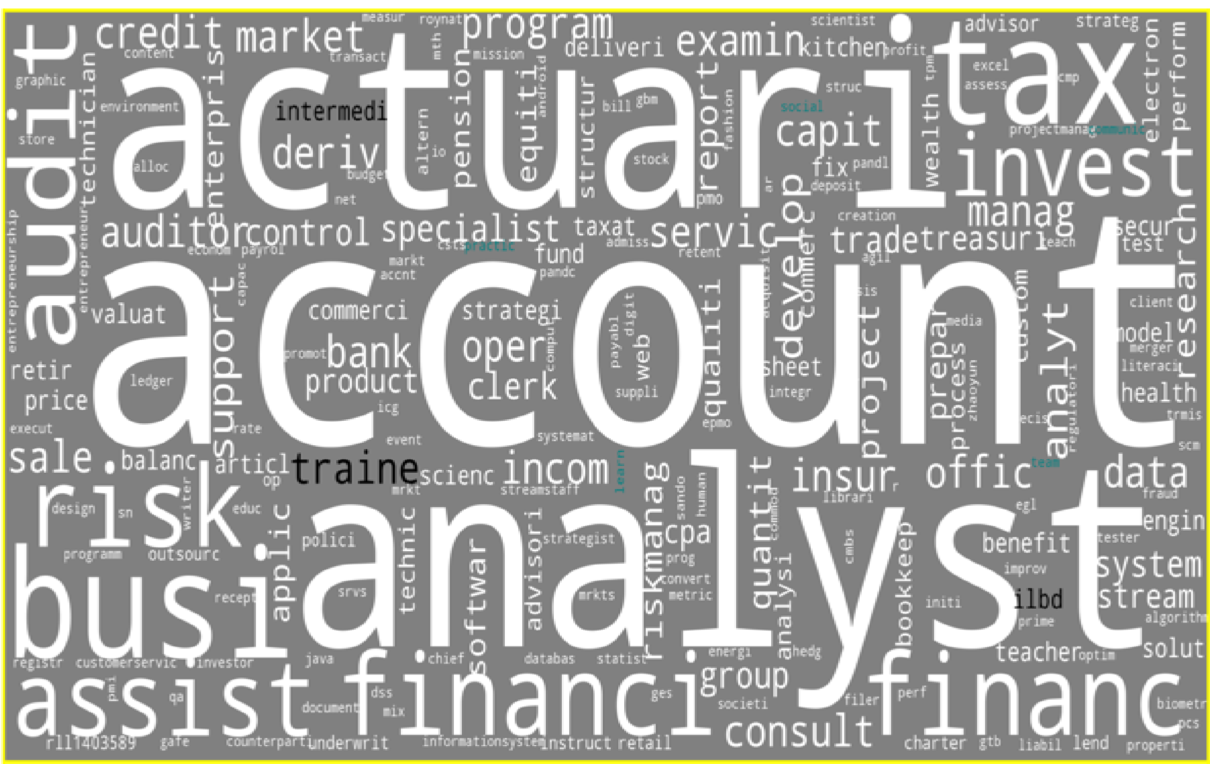

Figure 4.4 Word cloud of all the attributes of the job titles of Fin sized by representativeness

Figure 4.4 reveals that Fin is represented by specific attributes such as “actuari”, “account”, “risk”, “audit”, “tax” and “invest”. Attributes like “software”, “data”, “java” and “web” appear but are small and thus do not represent Fin (not in the Top 25 representative attributes of Fin’s job titles Table 4.4).

Table 4.3 Top 25 Frequent attributes of job titles of Fin

S. No.	Token	TF (↓)	Frequency in Group	Rank in Representative attributes
1	**analyst**	749	29%	2
2	**account**	512	20%	1
3	**actuari**	290	11%	3
4	**assist**	257	10%	10
5	**busi**	229	9%	4
6	**tax**	156	6%	5
7	**financi**	154	6%	6
8	**financ**	122	5%	7
9	**risk**	114	4%	8
10	**develop**	97	4%	24
11	**program**	93	4%	12
12	**invest**	84	3%	11
13	**audit**	83	3%	9
14	**manag**	83	3%	14
15	**market**	73	3%	17
16	**traine**	66	3%	15
17	**auditor**	64	2%	13
18	**servic**	61	2%	19
19	**data**	58	2%	18
20	**research**	58	2%	36
21	**incom**	52	2%	16
22	**oper**	51	2%	22
23	**support**	51	2%	28
24	**offic**	48	2%	21
25	**softwar**	47	2%	52

Table 4.4 Top 25 Representative attributes of job titles of Fin

S. No.	Token	TFIDF (↓)	Rank in Representative attributes	Frequency in Group
1	**account**	1732.9	1	20%
2	**analyst**	1654.9	2	29%
3	**actuari**	1107.8	3	11%
4	**busi**	773.9	4	9%
5	**tax**	718.9	5	6%
6	**financi**	686.2	6	6%
7	**financ**	580.0	7	5%
8	**risk**	530.2	8	4%
9	**audit**	440.0	9	3%
10	**assist**	422.4	10	10%
11	**invest**	385.2	11	3%
12	**program**	357.5	12	4%
13	**auditor**	351.7	13	2%
14	**manag**	321.9	14	3%
15	**traine**	313.3	15	3%
16	**incom**	295.6	16	2%
17	**market**	288.0	17	3%
18	**data**	251.5	18	2%
19	**servic**	239.9	19	2%
20	**deriv**	238.8	20	2%
21	**offic**	205.1	21	2%
22	**oper**	204.4	22	2%
23	**consult**	193.0	23	2%
24	**develop**	191.0	24	4%
25	**capit**	185.9	25	1%

To understand the Fin job market in detail, we next examine all the attributes of the Fin job descriptions. Figure 4.5 shows all the attributes sized by their frequency. Table 4.5 shows the Top 25 frequent attributes with their frequency and representativeness rank. Figure 4.6 and Table 4.6 show the same for the most representative attributes.

Figure 4.5 Word cloud of all the attributes of Fin sized by frequency

Examining the frequent attributes suggests the following.

- Soft skills including “team” and “communic” frequently appear in Fin job descriptions (“team” in 77% and “communic” in 63%)
- Confirming the findings from job titles, Fin has fewer “lead” and more assistant roles, shown by frequency of “assist” and “support”.
- The high frequency of “client” and “service” suggests a consumer orientation.

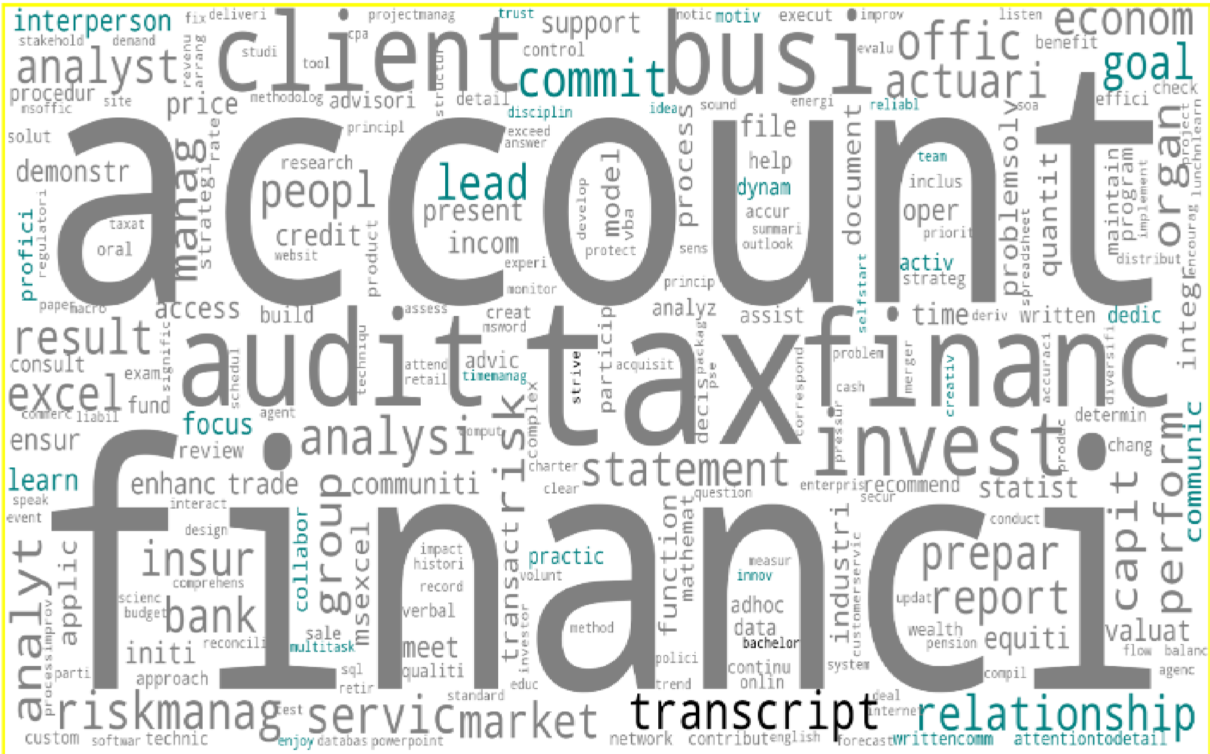

Figure 4.6 Word cloud of all the attributes of Fin sized by representativeness

Figure 4.6 and Table 4.6 suggest that the Finance related skills of “account”, “tax”, “audit”, “invest”, “risk management” etc. are most representative.

- “client” is among the Top 10 defining attributes of Fin, suggesting the importance of client-related skills. Soft skills such as “commitment”, “relationship”, “interpersonal skills” and “communication” are also representative of Fin.
- “transcript” is a defining attribute of Fin with almost 25% of the job descriptions requiring students to include transcripts of their grades with their applications.
- “office” is the 12th most representative attribute of Fin and is mentioned in almost 50% of its postings, emphasizing the formal office environments in Fin.

Table 4.5 Top 25 Frequent attributes of Fin

S. No.	Token	TF (↓)	Frequency in Group	Rank in Representative attributes
1	team	2004	77%	157
2	experi	1949	75%	234
3	busi	1914	74%	6
4	excel	1743	67%	21
5	develop	1730	67%	207
6	communic	1651	64%	68
7	manag	1613	62%	31
8	financi	1600	62%	1
9	servic	1562	60%	13
10	account	1487	57%	2
11	support	1455	56%	38
12	applic	1388	53%	69
13	program	1378	53%	96
14	report	1338	52%	26
15	assist	1327	51%	83
16	offic	1289	50%	12
17	client	1288	50%	7
18	lead	1233	47%	32
19	perform	1191	46%	30
20	project	1158	45%	148
21	analysi	1142	44%	15
22	process	1117	43%	57
23	time	1086	42%	40
24	learn	1059	41%	56
25	organ	1051	40%	34

Table 4.6 Top 25 Representative attributes of Fin

S. No.	Token	TFIDF (↓)	Rank in Representative attributes	Frequency in Group
1	financi	2579.5	1	62%
2	account	2488.8	2	57%
3	tax	2027.5	3	26%
4	financ	1999.8	4	32%
5	audit	1937.7	5	29%
6	busi	1753.7	6	74%
7	client	1704.0	7	50%
8	invest	1694.5	8	27%
9	transcript	1595.8	9	24%
10	risk	1577.3	10	25%
11	analyt	1524.4	11	39%
12	offic	1488.5	12	50%
13	servic	1469.9	13	60%
14	riskmanag	1436.3	14	17%
15	analysi	1409.4	15	44%
16	commit	1408.2	16	31%
17	relationship	1393.1	17	23%
18	analyst	1389.6	18	23%
19	statement	1388.1	19	17%
20	bank	1297.9	20	17%
21	excel	1281.5	21	67%
22	actuari	1274.2	22	14%
23	result	1273.1	23	28%
24	prepar	1265.1	24	37%
25	market	1259.4	25	33%

Overall, the results in this Section suggest that Fin jobs emphasize interpersonal skills and grades, are placed in formal office environments and are client-oriented.

4.2.2 IT job analysis

We start by analyzing frequent attributes in job titles of IT (Figure 4.7 and Table 4.7).

- Software terms, e.g., "software", "web", "programming", "mobile", "debug", "security", have a higher frequency than hardware terms "hardware" and "embedded" (representing "embedded systems").
- While IT also has "support" and "assist" positions like Fin, it seems to offer more "develop" positions. Furthermore, there are more "develop" positions than "tester" positions and more "design" positions than "qa" or "research" positions.
- Job titles also tend to specify the level of student who should apply, e.g., "intermediate".
- While IT job titles contain more traditional computer skills like "databas", "java", ".NET", "C++", "javascript", we also see emerging technologies like "cloud", "android", "ios", "python", "distributed" (distributed computing) and "data" (data science).
- Specific knowledge, e.g., "backend", "agile", "stack" etc. mentioned in the job titles emphasizes their importance.
- Notably, "team" (a soft skill) occurs as frequently as "java" (a core Specific Skill).

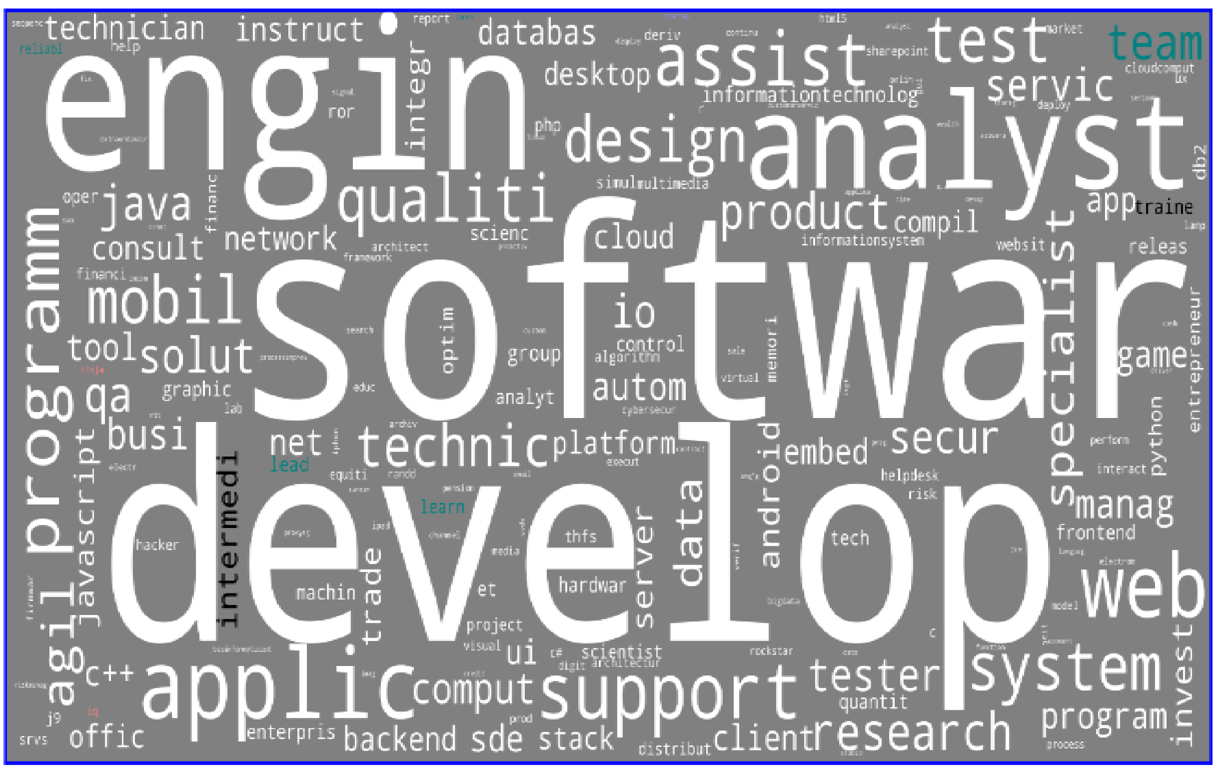

Figure 4.7 Word cloud of all the attributes of the job titles of IT sized by frequency

Following similar trends as in the frequent attributes, the representative attributes (Figure 4.8 and Table 4.8) that distinguish IT job titles from other disciplines focus on software skills.

- Apart from that, IT job titles mention “ninja” showing use of more casual language than Fin.
- They also mention “startup” and “entrepreneur” showing their inclination towards start-ups.

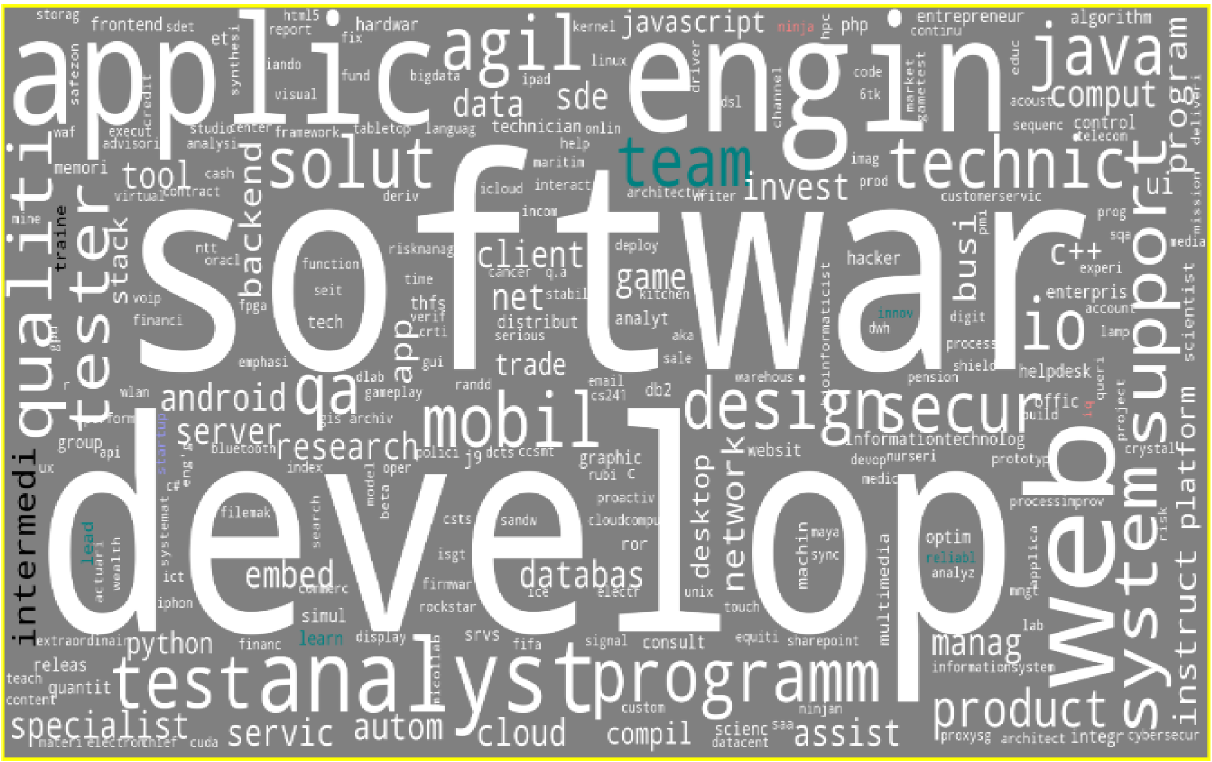

Figure 4.8 Word cloud of all the attributes of the job titles of IT sized by representativeness

Table 4.7 Top 25 Frequent attributes of job titles of IT

S. No.	Token	TF (↓)	Frequency in Group	Rank in Representative attributes
1	softwar	1531	45%	1
2	develop	1515	44%	2
3	engin	1062	31%	3
4	analyst	289	8%	6
5	applic	230	7%	4
6	web	184	5%	5
7	support	143	4%	8
8	assist	132	4%	25
9	programm	128	4%	7
10	system	109	3%	11
11	qualiti	105	3%	12
12	mobil	102	3%	9
13	test	101	3%	10
14	design	93	3%	15
15	technic	77	2%	14
16	agil	65	2%	13
17	research	62	2%	38
18	product	59	2%	23
19	qa	58	2%	16
20	specialist	58	2%	24
21	team	57	2%	18
22	solut	53	2%	20
23	java	51	1%	17
24	io	47	1%	19
25	tester	47	1%	22

Table 4.8 Top 25 Representative attributes of job titles of IT

S. No.	Token	TFIDF (↓)	Rank in Representative attributes	Frequency in Group
1	softwar	3292.9	1	45%
2	develop	2982.9	2	44%
3	engin	1704.3	3	31%
4	applic	886.7	4	7%
5	web	721.5	5	5%
6	analyst	638.5	6	8%
7	programm	573.6	7	4%
8	support	502.7	8	4%
9	mobil	490.6	9	3%
10	test	423.7	10	3%
11	system	406.4	11	3%
12	qualiti	401.1	12	3%
13	agil	342.3	13	2%
14	technic	315.1	14	2%
15	design	309.3	15	3%
16	qa	282.0	16	2%
17	java	281.8	17	1%
18	team	278.4	18	2%
19	io	268.8	19	1%
20	solut	265.9	20	2%
21	secur	242.8	21	1%
22	tester	234.6	22	1%
23	product	228.5	23	2%
24	specialist	223.9	24	2%
25	assist	217.0	25	4%

Figure 4.9 Word cloud of all the attributes of IT sized by frequency

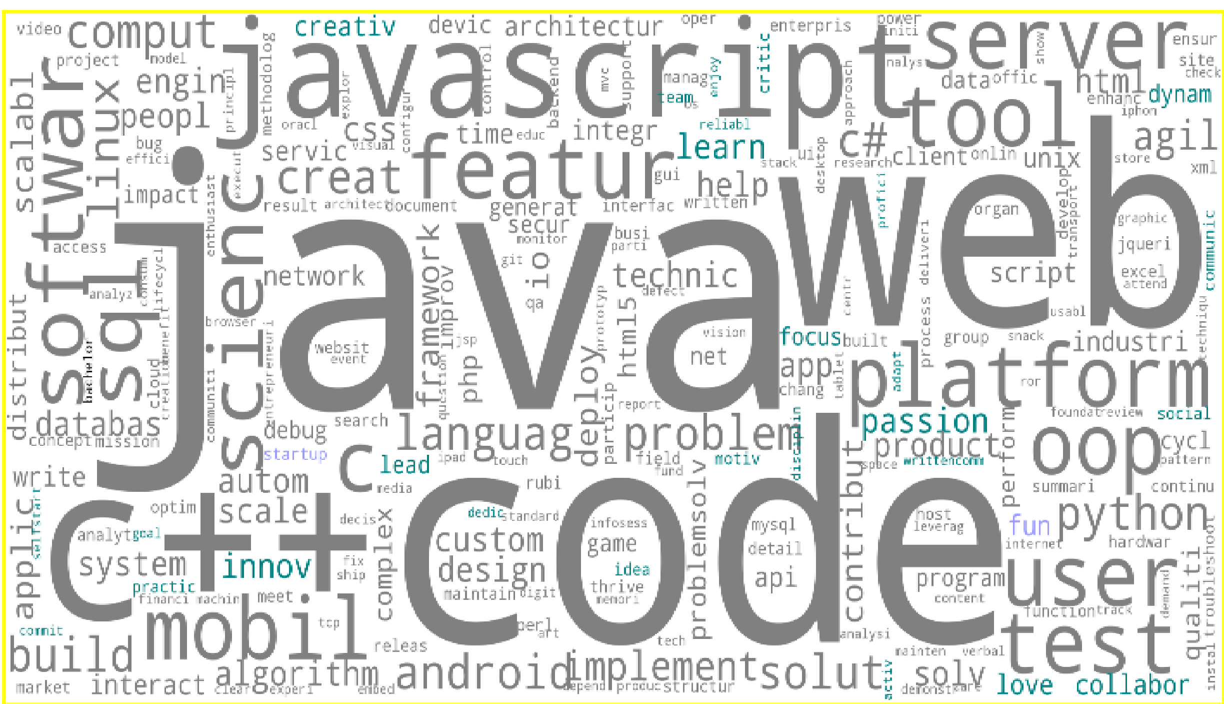

Figure 4.10 Word cloud of all the attributes of IT sized by representativeness

Table 4.9 Top 25 Frequent attributes of IT

S. No.	Token	TF (↓)	Frequency in Group	Rank in Representative attributes
1	**develop**	3114	91%	130
2	**experi**	2890	84%	193
3	**team**	2888	84%	143
4	**softwar**	2597	76%	12
5	**applic**	2256	66%	45
6	**design**	2227	65%	44
7	**product**	2133	62%	40
8	**program**	2043	60%	83
9	**system**	1978	58%	33
10	**engin**	1960	57%	38
11	**project**	1817	53%	109
12	**comput**	1775	52%	25
13	**test**	1702	50%	16
14	**build**	1658	48%	27
15	**communic**	1651	48%	141
16	**web**	1613	47%	3
17	**code**	1592	46%	2
18	**help**	1558	46%	32
19	**learn**	1541	45%	53
20	**servic**	1520	44%	65
21	**java**	1483	43%	1
22	**manag**	1472	43%	120
23	**creat**	1465	43%	26
24	**solut**	1453	42%	21
25	**technic**	1430	42%	37

Table 4.10 Top 25 Representativeness attributes of IT

S. No.	Token	TFIDF (↓)	Rank in Representative attributes	Frequency in Group
1	**java**	3114.1	1	43%
2	**code**	2828.5	2	46%
3	**web**	2683.2	3	47%
4	**c++**	2632.6	4	33%
5	**javascript**	2527.0	5	31%
6	**platform**	2370.2	6	33%
7	**featur**	2349.5	7	30%
8	**mobil**	2348.4	8	32%
9	**server**	2342.1	9	29%
10	**oop**	2294.1	10	24%
11	**user**	2278.1	11	35%
12	**softwar**	2262.7	12	76%
13	**c**	2091.4	13	24%
14	**scienc**	2082.5	14	39%
15	**sql**	2050.3	15	26%
16	**test**	2020.5	16	50%
17	**tool**	2008.5	17	41%
18	**languag**	2000.9	18	29%
19	**problem**	1991.6	19	34%
20	**python**	1985.5	20	22%
21	**solut**	1962.5	21	42%
22	**linux**	1956.3	22	21%
23	**c#**	1905.1	23	20%
24	**android**	1889.3	24	19%
25	**comput**	1875.4	25	52%

Analyzing the frequent attributes of IT (Figure 4.9 and Table 4.9), we draw the following insights.

- Not surprisingly, almost 91% of IT jobs require "development" skills and almost 50% of the jobs require testing skills.
    - With more than 43% of the jobs in IT requiring "java", Java is the most frequently mentioned programming language in IT. Other popular programming languages in IT include C++ (with 33% of the job postings mentioning it), JavaScript (31%), C (24%), Python (22%), C# (20%), HTML (19%), CSS (17%), PHP (12%), .NET (12%), jQuery (10%), Perl (10%), XML (9%) and Ruby (9%).
    - While web development is required by 47% of the jobs, mobile development is required by 32%. Android application development is required by 19% and IPhone application development is required by 7% of the jobs in IT.
    - Knowledge of databases is required by 29% of the jobs while 26% mention SQL, 8% mention MySQL and 7% mention Oracle.

    - Knowledge of Linux is required by 21% of the jobs and Unix by 13%.
    - Some jobs require advanced skills such as distributed systems (required by 17%), cloud computing (required by 9%) and Big Data (required by 4%).
- While "software" is mentioned in 76% of the postings, "hardware" is mentioned in only 14%.
- Apart from Specific Skills, soft skills such as "team", "lead", "communication" and "collaboration" as well as mindset related soft skills including "passion", "love", "enjoy", "selfstarter", "focus", "motivation" and "learn" (related to quick learning) are frequent in IT. The mention of "innovation", "creativity" and the above show that IT requires students who not only possess technical and interpersonal skills, but also a passion for the work they do.
- "teamwork" is required by almost 85% of the jobs in IT showing that IT jobs often feature a collaborative environment.
- Attributes labelled as Company Culture or Perks also appear in the frequent attributes of IT. A "fun" work environment and "mentorship" seem to be offered by many IT jobs.

In line with the observations made using the Frequent attributes of IT, Figure 4.10 and Table 4.10 show similar trends in the most representative skills of IT.

- "Java" seems to be the most defining skill of IT followed by "code", "web", "C++" and "javascript". More specific skills like "OOP" (Object Orient Programming), "Linux", "C#" and "android" also seem to represent IT.
- Representative attributes such as "platform", "feature" (related to features of a system), "user" and "deploy" suggest the development of consumer-oriented systems.
- Attributes such as "platform", "architecture", "framework" and "algorithm" rank among the most representative attributes, emphasizing the knowledge of computer systems in addition to specific programming languages.
- Attributes related to company culture and soft skills also represent IT.

Overall, the results in this Section indicate that IT positions focus on software instead of hardware and claim to offer a fun and collaborative work environment. In addition to other soft skills, IT includes mindset related soft skills such as passion.

4.2.3 Mech job analysis

To analyze the Mech Discipline, we examine all the attributes present in the Job Titles of Mech. Figure 4.11 shows the attributes sized by their frequency and Figure 4.12 shows the attributes sized by their representativeness. Table 4.11 and Table 4.12 provide the frequency and representativeness rank of the top 25 most frequent and representative attributes found in the job titles of Mech.

Zooming into the frequent attributes mentioned in the job titles of Mech (Figure 4.11 and Table 4.11), the following can be inferred.

- While Mech contains attributes like “engin”, “develop”, “mechan” (representing Mechanical) and “manufactur” (representing manufacturing jobs), it also mentions “software”, “java” and “web”. This could suggest Mech’s trend towards IT.
- The frequent attributes also contain mechanical-related attributes such as “hardware”, “electrical”, “control”, “processes”, “robot”, “circuit”, “material” and “CAD”.
- Placements in “labs” and “plants” seem to be frequent, unlike in other disciplines.
- Apart from “develop” and “design”, “quality”, “test” and “maintenance” jobs are frequent in Mech.
- Similar to other disciplines, Mech has more “support” and “assist” jobs than managerial positions (inferred by the size of “specialist” and “projectmanag”). Mech also has “technician” and “inspector” positions that were not seen so frequently in other disciplines.
- Many Mech jobs seem to be research oriented.
- With “team” appearing in the word cloud of the job titles, Mech seems to value “teamwork” skills, as was the case in IT.

An analysis of the most representative attributes mentioned in the job titles of Mech (Figure 4.12 and Table 4.12) reveals similar findings.

- A variety of mechanical skills represent the job titles of the Mech discipline: “fuel”, “electron”, “gas”, “seismic”, “fluid” and “robot”.
- “software” and “web” are part of the representative attributes of Mech Job titles. “software” is the 8^{th} most representative attribute of Mech.
- Some soft skills including “team” and “lead” are also representative of Mech job titles.

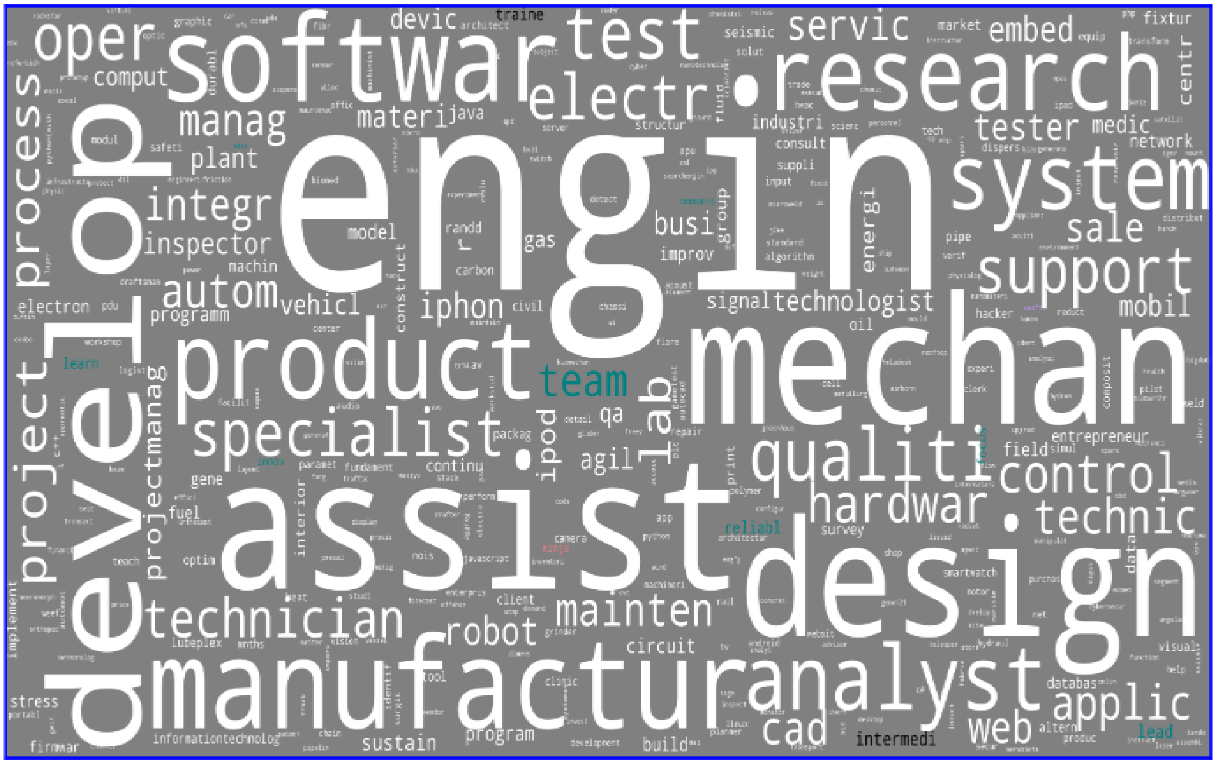

Figure 4.11 Word cloud of all the attributes of the job titles of Mech sized by frequency

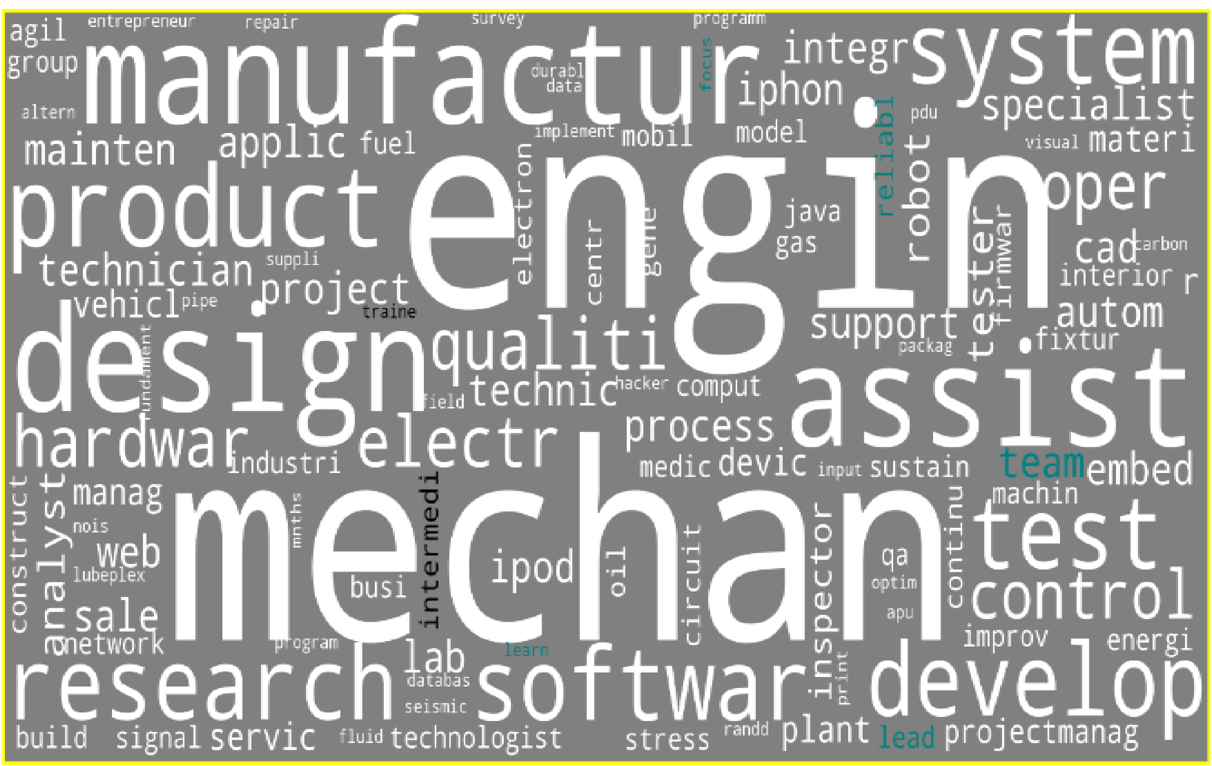

Figure 4.12 Word cloud of all the attributes of the job titles of Mech sized by representativeness

Table 4.11 Top 25 Frequent attributes of job titles of Mech

S. No.	Token	TF (↓)	Frequency in Group	Rank in Representative attributes
1	engin	923	50%	1
2	assist	274	15%	5
3	mechan	253	14%	2
4	design	211	12%	3
5	develop	201	11%	7
6	manufactur	147	8%	4
7	softwar	135	7%	8
8	product	108	6%	6
9	research	95	5%	10
10	system	73	4%	9
11	analyst	69	4%	19
12	test	55	3%	11
13	qualiti	55	3%	12
14	oper	44	2%	16
15	specialist	44	2%	17
16	support	43	2%	20
17	electr	41	2%	13
18	technician	39	2%	24
19	control	38	2%	15
20	project	38	2%	25
21	hardwar	37	2%	14
22	process	34	2%	18
23	lab	33	2%	26
24	autom	29	2%	21
25	applic	28	2%	32

Table 4.12 Top 25 Representative attributes of job titles of Mech

S. No.	Token	TFIDF (↓)	Rank in Representative attributes	Frequency in Group
1	engin	1481.2	1	50%
2	mechan	1041.5	2	14%
3	design	701.8	3	12%
4	manufactur	640.2	4	8%
5	assist	450.4	5	15%
6	product	418.2	6	6%
7	develop	395.8	7	11%
8	softwar	290.4	8	7%
9	system	272.1	9	4%
10	research	238.7	10	5%
11	test	230.7	11	3%
12	qualiti	210.1	12	3%
13	electr	191.2	13	2%
14	hardwar	186.6	14	2%
15	control	185.0	15	2%
16	oper	176.3	16	2%
17	specialist	169.9	17	2%
18	process	156.9	18	2%
19	analyst	152.5	19	4%
20	support	151.2	20	2%
21	autom	150.5	21	2%
22	cad	149.2	22	1%
23	mainten	148.4	23	1%
24	technician	138.6	24	2%
25	project	138.0	25	2%

Figure 4.13 Word cloud of all the attributes of Mech sized by frequency

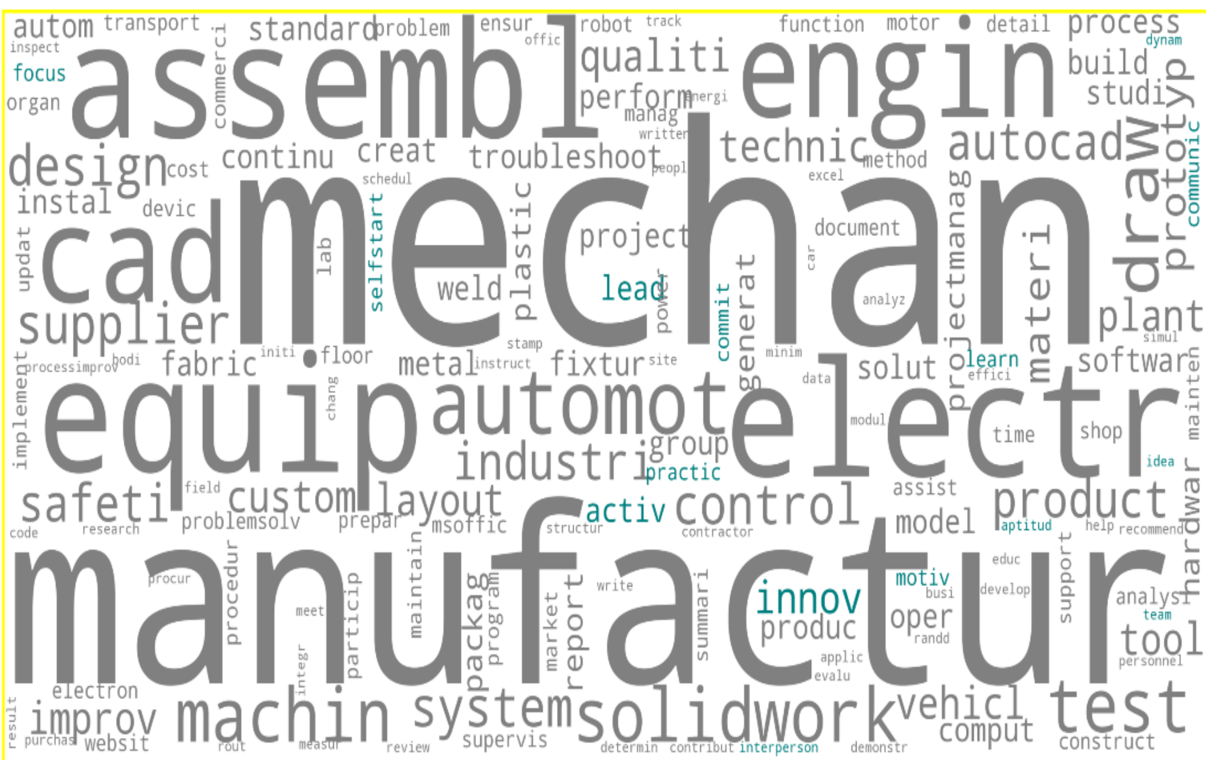

Figure 4.14 Word cloud of all the attributes of Mech sized by representativeness

Table 4.13 Top 25 Frequent attributes of Mech

S. No.	Token	TF (↓)	Frequency in Group	Rank in Representative attributes
1	engin	1415	77%	6
2	develop	1336	73%	140
3	experi	1296	71%	215
4	design	1267	69%	15
5	team	1237	68%	159
6	product	1130	62%	17
7	project	1080	59%	46
8	system	1043	57%	14
9	communic	896	49%	102
10	mechan	891	49%	1
11	program	883	48%	93
12	assist	867	47%	67
13	test	821	45%	10
14	softwar	798	44%	40
15	support	782	43%	64
16	manufactur	767	42%	2
17	process	762	42%	32
18	manag	755	41%	91
19	lead	737	40%	35
20	excel	730	40%	108
21	applic	719	39%	114
22	industri	694	38%	18
23	report	692	38%	59
24	perform	686	37%	38
25	equip	682	37%	3

Table 4.14 Top 25 Representative attributes of Mech

S. No.	Token	TFIDF (↓)	Rank in Representative attributes	Frequency in Group
1	mechan	2071.0	1	49%
2	manufactur	1663.0	2	42%
3	equip	1344.9	3	37%
4	assembl	1257.7	4	24%
5	electr	1218.3	5	27%
6	engin	1216.5	6	77%
7	cad	1202.5	7	21%
8	automot	1135.6	8	18%
9	solidwork	1041.2	9	14%
10	test	974.6	10	45%
11	draw	963.0	11	22%
12	machin	954.8	12	18%
13	supplier	944.4	13	16%
14	system	936.1	14	57%
15	design	930.1	15	69%
16	control	921.8	16	28%
17	product	891.3	17	62%
18	industri	885.0	18	38%
19	autocad	884.2	19	18%
20	safeti	864.4	20	21%
21	qualiti	851.8	21	34%
22	prototyp	851.1	22	16%
23	improv	842.6	23	30%
24	custom	833.8	24	33%
25	technic	786.1	25	36%

The most frequent attributes of the job descriptions of Mech are shown in Figure 4.13 and Table 4.13. We make the following conclusions.

- While general attributes including "engin", "develop", "experi" and "project" are frequent, specific attributes like "mechan", "manufactur", "process" and "equip" (related to equipment) can also be seen.
- "design" appears more frequently than "test".
- With emphasis on "team" and "communic", many soft skills reflecting mindset can be seen ("passion", "love", "focus", "attention to detail", "self-starter", "focus", "active" etc.). Notably, teamwork is mentioned in 67% of Mech jobs, but in 77% of Fin and 85% of IT jobs.

The representative attributes of Mech (Figure 4.14 and Table 4.14) show many attributes that distinguish Mech from other disciplines.

- With “mechan” and “manufacture” being the obvious attributes that distinguish Mech from other disciplines, specific Mech skills such as “equip”, “assembly”, “CAD”, “SolidWorks”, “AutoCAD” (design software), “draw” and “prototype” are also representative.
- Unlike other disciplines, “safety” is a representative attribute of Mech owing to their non-office environment.
- Testing and troubleshooting seem to be important skills to have in Mech.

Overall, the results in this Section indicate that Mech job descriptions mention mechanical and design concepts as well as IT related software skills. Teamwork and initiative are mentioned frequently, as is safety due to lab and plant environments.

4.2.4 Similarity between disciplines

This section examines how Fin, IT and Mech differ using their Top 100 frequent and representative attributes. Comparisons are made using all the methods listed in Section 3.2.5.

4.2.4.1 Attribute intersections

We start with a quantitative comparison of the Top 100 most frequent and Top 100 most representative attributes of the three disciplines. Figure 4.15 and Figure 4.16 show the Jaccard Similarity (JS). Recall that JS does not take into account the rank of the attributes. It simply calculates similarity based on the presence of an attribute in a Group.

Figure 4.15 suggests that about half the frequently mentioned attributes of any discipline are common to all of them. Next, Figure 4.17 shows a Venn Diagram with the top 100 frequent attributes of each discipline. The attributes are sized based on their frequency; the sizes of attributes in the intersections are based on the lowest frequency of the attribute among the groups that share it. All disciplines mention generic attributes like “experi”, “busi”, “perform”, “process”, “product” etc. Furthermore, all disciplines frequently mention soft skills including “team”, “communic”, “lead”, “learn”, “time” (representing time management skills), “focus”, “motivation”, “active” and “practice”.

Other insights from the intersection of disciplines include:

- All disciplines mention IT skills, e.g., “software”, “data” and “program”.
- While all the disciplines contain “create”, “design” and “develop”, they also contain “maintain” and “test” suggesting that co-op students get a chance to apply their knowledge in various ways.
- While “lead” and “manage” are common among all the disciplines, so are “assist” and “support” suggesting that all the disciplines offer both assistant and managerial roles.
- Documenting, reporting, research and problem-solving skills are common to all disciplines.

According to Figure 4.15, Mech vs. Fin has a higher Jaccard Similarity than Mech vs. IT and Fin vs. IT. Zooming in on Mech and Fin's intersection in the Venn Diagram (Figure 4.17), they share "modelling", "project management", "detailing", "reviewing" etc. They also share some soft skills including "interpersonal" skills and "commitment" which are absent from the frequent attributes of IT.

Other pairwise comparisons reveal the following insights:

- Fin and IT share a focus towards clients that is missing from Mech. They also provide a "dynamic" and "collaborative" work environment which is not frequent in Mech.
- Mech and IT mention innovation more than Finance.

Finally, looking at attributes that are frequent in only one discipline reveals additional insight.

- Finance shows a frequency of "transcript" indicating a greater emphasis on grades. Finance also appears to place greater emphasis on "goal orientation" and "relationships".
- Apart from typical IT skills, IT emphasizes "passion" and "creativity". This suggests the importance of mindset in IT.
- Other than typical Mech skills, Mech contains "MS Office" in the Top 100 most frequent attributes.

With JS for all comparisons being low, Figure 4.16 shows that the most representative attributes of the disciplines are different. As these skills define their disciplines, we did not expect them to be similar but included the analysis for completeness. The results match our expectations as the different disciplines share only 10-20 attributes from the top 100 attributes that represent them. We examine the Venn Diagram in Figure 4.18 to understand similarities among the three disciplines.

- It is interesting to see "program" in the intersection of the Top 100 representative skills of all the 3 disciplines. This may indicate a trend towards IT skills in other disciplines.
- Soft skills such as "problem solving", "time management", "learning" and "leadership" are important in all disciplines.

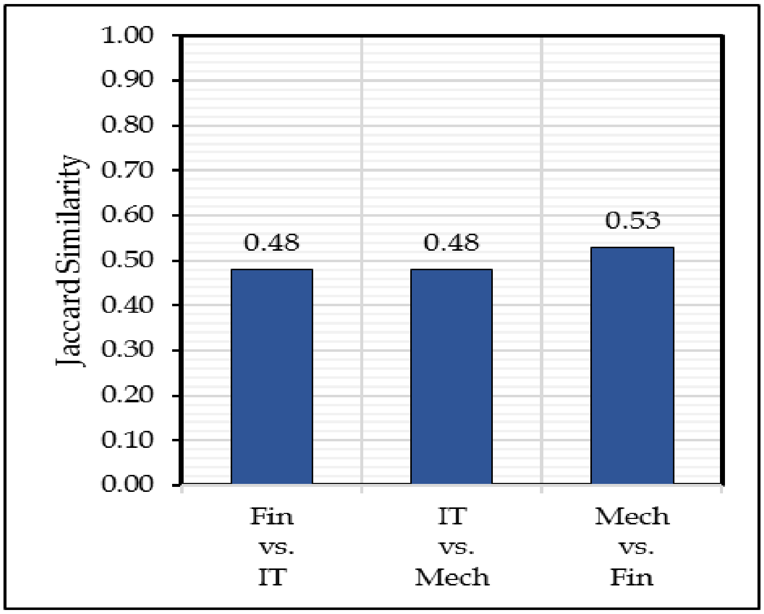

Figure 4.15 Comparison of the top 100 most frequent attributes of Fin, IT and Mech using Jaccard similarity

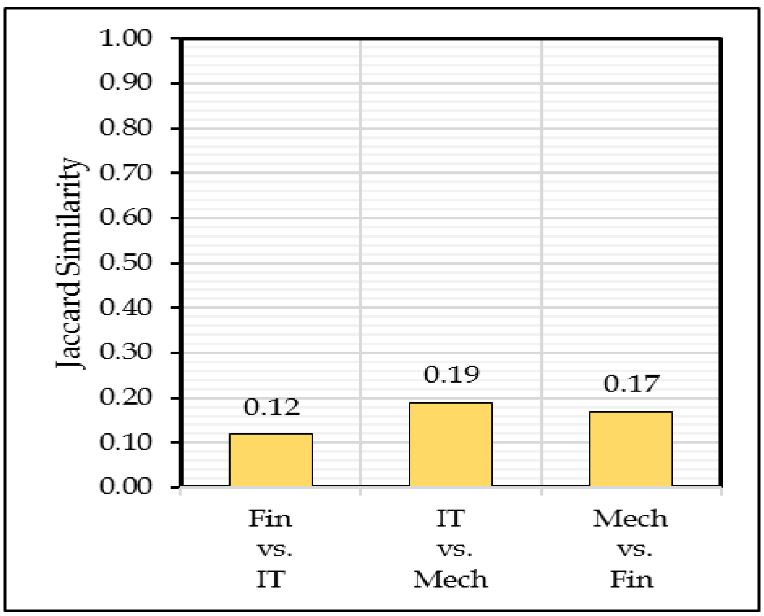

Figure 4.16 Comparison of top 100 most representative attributes of Fin, IT and Mech using Jaccard similarity

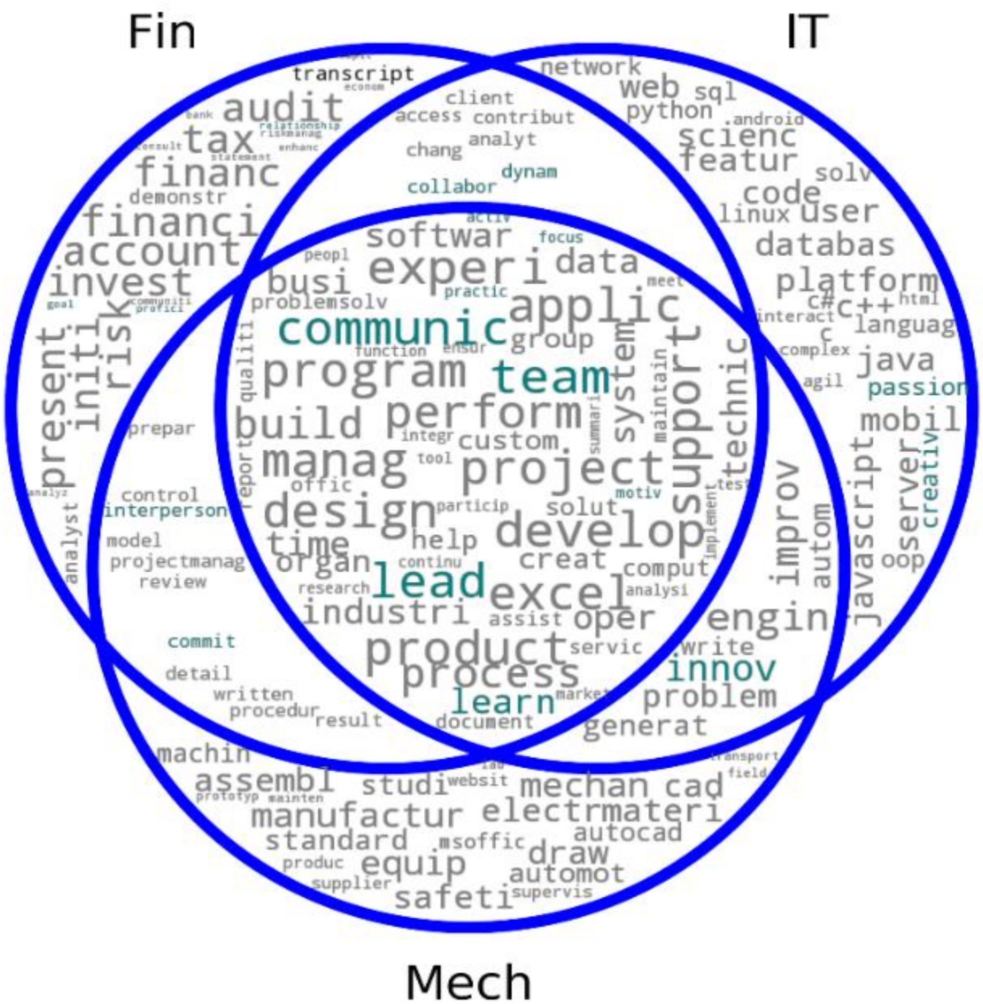

Figure 4.17 Overlap between the top 100 most frequent attributes of Fin, IT and Mech

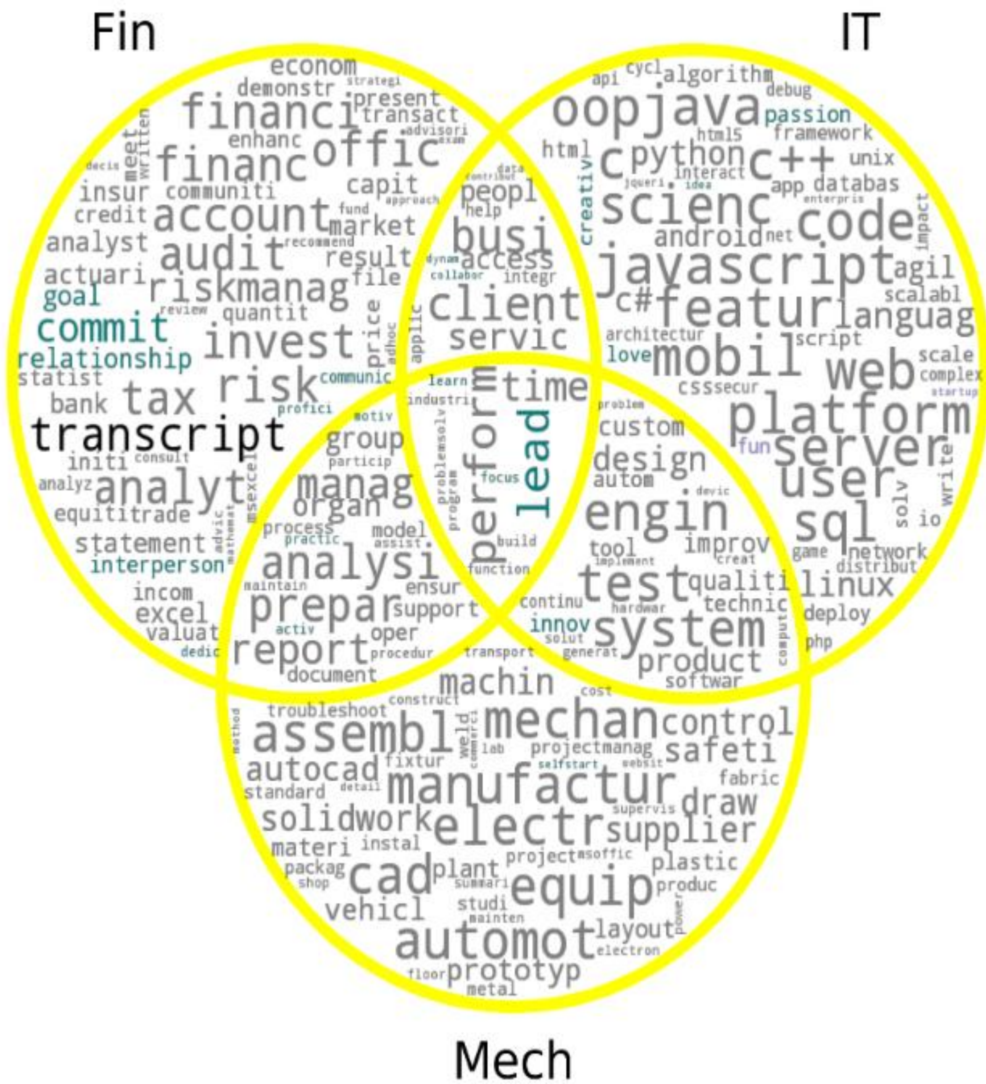

Figure 4.18 Overlap between the top 100 most representative attributes of Fin, IT and Mech

4.2.4.2 Distribution of types of attributes among the Top 100

To gain further insight into the differences between the three disciplines, we examine the differences in the distribution of the types of attributes within the Top 100 most frequent (Figure 4.19) and representative (Figure 4.20) attributes. Recall that the types of attributes include Soft Skills, Perks, Admin, Insider, Media, Internet Slang and Specific Job Requirements. We consider distributions of Perks, Admin and Soft Skills as no attributes of Insider, Media or Internet Slang appear in the Top 100 frequent or representative attributes. We do not examine the fraction of Specific Job Requirements as these vary among the three disciplines. The other types of attributes, on the other hand, have a common vocabulary.

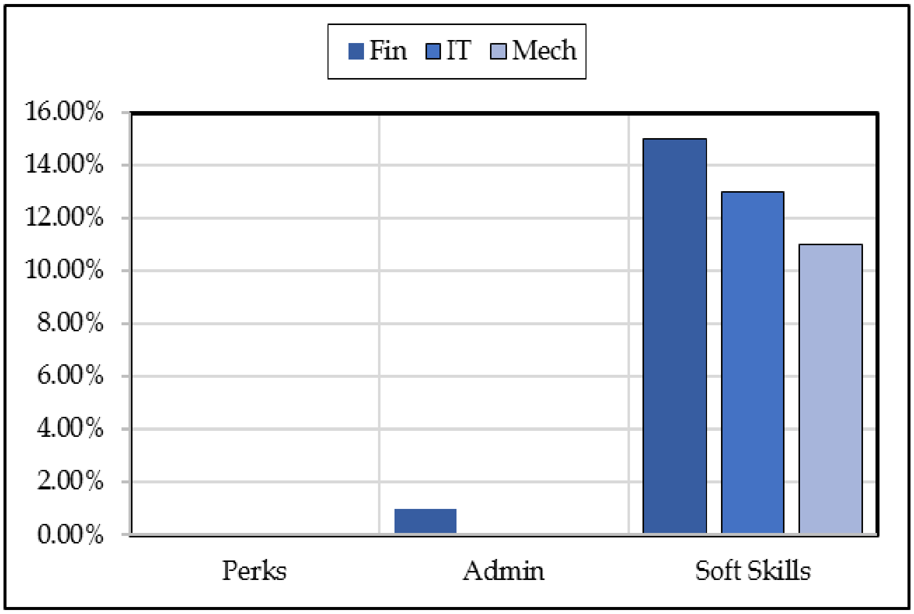

Figure 4.19 Distribution of various types of attributes in the top 100 most frequent attributes of Fin, IT and Mech

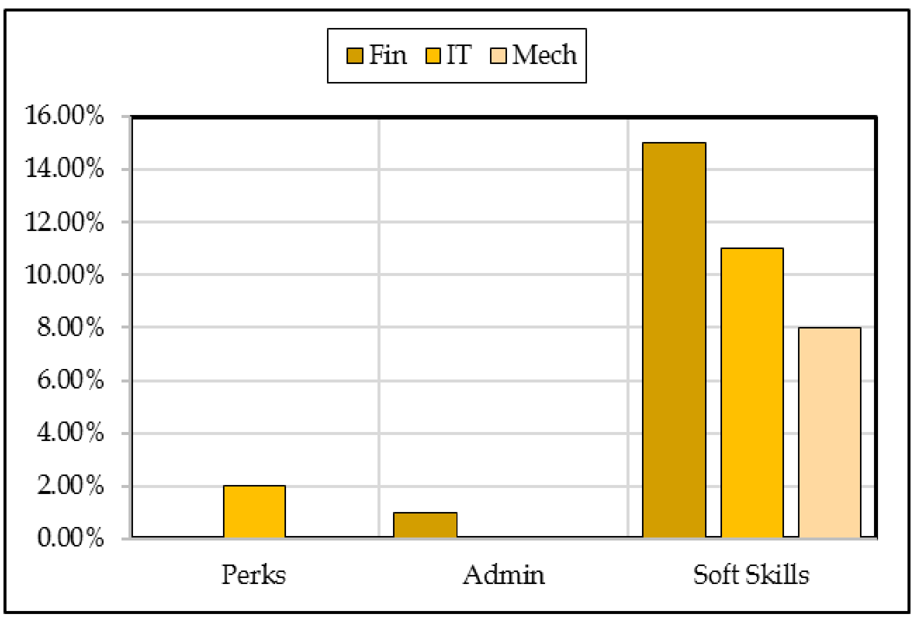

Figure 4.20 Distribution of various types of attributes in the top 100 most representative attributes of Fin, IT and Mech

As seen in Figure 4.19, Perk and Company Culture attributes are not frequent in any discipline. However, Company Culture (and/or Perks), e.g., a "fun" working environment and working at "startups" are more representative of IT (Figure 4.20). Furthermore, Figure 4.19 and Figure 4.20 reinforce the previous observation that Admin attributes (e.g. "transcript") are specific to Fin. Finally, both figures show that Soft Skills are most frequent in Fin and least frequent in Mech.

4.2.4.3 Attributes with higher frequency in one discipline than another

Next, we examine how the demand for an Attribute changes from one discipline to another. We start with analyzing how Fin differs from IT (Table 4.15) and Mech (Table 4.16):

- Apart from core Fin skills, e.g., "account", "tax", "invest", "audit" etc. which have a higher frequency in Fin than in any other discipline, Fin has higher demand for some soft skills.
  - Fin mentions "relationship" 18% more than IT or Mech,
  - Fin includes "interpersonal" and "communication" skills 15% more than IT or Mech.
- "client" and "service" appear in Fin 22% and 16% more often than IT (Table 4.15) and 36% and 38% more often in Mech (Table 4.16) suggesting that Fin jobs are more client oriented.
- Confirming the previous findings (Section 4.2.1) about Fin's administrative requirements, Fin jobs require grade transcripts almost 20% more than any other discipline.
- Speculating by the presence of "assist" in job descriptions, while 22% of IT jobs have assistant roles, more than 50% of the Fin positions are junior/assistant. Also, "assist" and "support" are mentioned 29% and 18% more in Fin than in IT (Table 4.15). Fin also mentions them 4% and 13% more than Mech. This indicates that Fin has more assistant positions than any other discipline.
- 29% and 14% more "report" is mentioned in Fin than IT or Mech suggesting that Fin requires more work related to reports and/or reporting.
- "office" is seen 24% and 28% more in Fin than in IT or Mech suggesting a more formal work environment.

Next, we move on to IT jobs.

Table 4.15 Attributes with higher frequency in Fin than in IT sorted by their difference

S. No.	Token	Frequency in Fin	Frequency in IT	Difference in Frequency (↓)
1	financi	61.6%	13.6%	48.0%
2	account	57.3%	9.7%	47.6%
3	busi	73.7%	38.5%	35.2%
4	prepar	37.0%	6.1%	30.9%
5	report	51.5%	22.1%	29.4%
6	assist	51.1%	21.9%	29.2%
7	financ	31.6%	3.8%	27.9%
8	excel	67.1%	40.1%	27.1%
9	audit	28.8%	1.8%	27.0%
10	tax	25.6%	0.6%	25.0%
11	analysi	44.0%	19.6%	24.4%
12	offic	49.7%	25.6%	24.0%
13	client	49.6%	27.3%	22.3%
14	invest	27.0%	7.6%	19.4%
15	risk	24.6%	5.2%	19.4%
16	manag	62.1%	43.0%	19.1%
17	transcript	24.4%	5.8%	18.6%
18	support	56.0%	37.5%	18.6%
19	commit	30.6%	12.4%	18.1%
20	relationship	22.8%	4.8%	18.0%
21	present	26.8%	9.1%	17.8%
22	analyt	39.4%	22.5%	16.9%
23	interperson	26.3%	10.2%	16.1%
24	statement	16.6%	0.8%	15.8%
25	servic	60%	44%	16%

Table 4.16 Attributes with higher frequency in Fin than in Mech sorted by their difference

S. No.	Token	Frequency in Fin	Frequency in Mech	Difference in Frequency (↓)
1	financi	61.6%	4.6%	57.1%
2	account	57.3%	7.3%	50.0%
3	busi	73.7%	27.7%	46.1%
4	servic	60.2%	22.1%	38.1%
5	client	49.6%	13.6%	36.0%
6	offic	49.7%	21.1%	28.5%
7	excel	67.1%	39.8%	27.3%
8	financ	31.6%	5.2%	26.4%
9	tax	25.6%	0.1%	25.5%
10	audit	28.8%	3.5%	25.3%
11	analyt	39.4%	14.2%	25.2%
12	invest	27.0%	2.7%	24.3%
13	transcript	24.4%	2.9%	21.4%
14	risk	24.6%	3.3%	21.3%
15	manag	62.1%	41.2%	20.9%
16	analyst	22.9%	3.3%	19.6%
17	relationship	22.8%	4.7%	18.1%
18	analysi	44.0%	26.9%	17.1%
19	riskmanag	17.3%	0.2%	17.1%
20	bank	16.9%	1.0%	15.9%
21	communiti	21.8%	6.2%	15.6%
22	econom	17.3%	1.7%	15.6%
23	statement	16.6%	1.1%	15.5%
24	communic	63.6%	48.9%	14.7%
25	insur	16%	1%	15%

As seen in Table 4.17 and Table 4.18, the majority of attributes that are more frequent in IT than Fin or Mech correspond to programming languages and software systems. Other insights include:

- While Table 4.17 indicates that 31% of the Fin jobs require the knowledge of "software", Table 4.18 specifies that 44% of Mech jobs require the knowledge of "software". This suggests that the knowledge of "software" is more important in Mech than in Fin.
- Consistent with the finding from Section 4.2.2, IT has more development jobs than Fin (24%) or Mech (18%). Suggesting a more user-oriented development, IT mentions "user" 25% and 28% more than Fin and Mech.
- IT mentions "test" 31% more than Fin and 5% more than Mech.
- With "user" being mentioned 25% more in IT than in Fin and "client" being used 22% more in Fin than in IT, we speculate that "user" and "client" are possibly used synonymously to refer to consumers.

- Confirming the speculations in Section 4.2.3, "team" occurs 7% more often in IT than in Fin and 17% more in IT than in Mech.
- Not seen in Table 4.18, soft skills related to mindset are >10% frequent in IT than in Mech.

Next, we discuss which attributes are more frequent in Mech.

Table 4.17 Attributes with higher frequency in IT than in Fin sorted by their difference

S. No.	Token	Frequency in IT	Frequency in Fin	Difference in Frequency (↓)
1	**softwar**	75.8%	30.6%	45.3%
2	**engin**	57.2%	12.4%	44.9%
3	**web**	47.1%	6.4%	40.8%
4	**java**	43.3%	2.8%	40.5%
5	**code**	46.5%	8.9%	37.6%
6	**test**	49.7%	18.4%	31.3%
7	**c++**	32.5%	2.5%	30.0%
8	**system**	57.8%	29.2%	28.6%
9	**javascript**	30.5%	2.0%	28.6%
10	**comput**	51.8%	23.9%	27.9%
11	**mobil**	31.5%	3.9%	27.6%
12	**featur**	29.6%	2.2%	27.5%
13	**platform**	33.0%	5.9%	27.1%
14	**design**	65.0%	38.4%	26.6%
15	**user**	34.9%	10.0%	24.9%
16	**server**	29.0%	4.2%	24.8%
17	**develop**	90.9%	66.6%	24.3%
18	**product**	62.3%	38.7%	23.6%
19	**scienc**	39.4%	16.1%	23.3%
20	**oop**	24.5%	1.6%	22.9%
21	**c**	24.1%	1.9%	22.2%
22	**tool**	41.1%	18.9%	22.2%
23	**languag**	28.9%	8.8%	20.1%
24	**python**	21.9%	1.8%	20.1%
25	**problem**	34%	15%	19%

Table 4.18 Attributes with higher frequency in IT than in Mech sorted by their difference

S. No.	Token	Frequency in IT	Frequency in Mech	Difference in Frequency (↓)
1	**java**	43.3%	7.7%	35.6%
2	**web**	47.1%	12.5%	34.6%
3	**code**	46.5%	13.6%	32.9%
4	**softwar**	75.8%	43.6%	32.3%
5	**scienc**	39.4%	9.9%	29.5%
6	**user**	34.9%	7.0%	27.9%
7	**applic**	65.9%	39.2%	26.6%
8	**javascript**	30.5%	4.9%	25.7%
9	**c++**	32.5%	8.8%	23.7%
10	**server**	29.0%	6.2%	22.8%
11	**platform**	33.0%	10.4%	22.6%
12	**servic**	44.4%	22.1%	22.3%
13	**featur**	29.6%	7.7%	22.0%
14	**mobil**	31.5%	9.9%	21.6%
15	**help**	45.5%	24.7%	20.8%
16	**oop**	24.5%	3.7%	20.8%
17	**languag**	28.9%	8.8%	20.0%
18	**sql**	25.8%	5.9%	19.9%
19	**databas**	29.4%	10.8%	18.6%
20	**develop**	90.9%	72.9%	18.0%
21	**comput**	51.8%	34.7%	17.2%
22	**c**	24.1%	7.3%	16.8%
23	**team**	84.3%	67.5%	16.8%
24	**python**	21.9%	6.0%	15.9%
25	**linux**	21%	5%	16%

As seen in Table 4.19 and Table 4.20, the majority of the attributes that Mech demands more than Fin or IT are related to core Mech skills or work profiles ("AutoCAD", "equipment", "manufacture" etc.). Other insights are as follows:

- Mech mentions design more than other disciplines.
- Parallel to the findings from Section 4.2.3, Mech jobs provide a more tangible and empirical work experience. Besides the mention of "equipment", "machine", "vehicle" etc., Table 4.19 and Table 4.20 highlight the frequency of "plants" and "labs", suggesting more field work in Mech.

- Safety is mentioned almost 18-20% more in Mech than in any other discipline
- MS Office is mentioned in Mech postings 2% and 12% more than in Fin or IT.

Table 4.19 Attributes with higher frequency in Mech than in Fin sorted by their difference

S. No.	Token	Frequency in Mech	Frequency in Fin	Difference in Frequency (↓)
1	**engin**	77.2%	12.4%	64.9%
2	**mechan**	48.6%	0.3%	48.4%
3	**manufactur**	41.9%	4.0%	37.8%
4	**equip**	37.2%	1.8%	35.4%
5	**design**	69.2%	38.4%	30.7%
6	**system**	56.9%	29.2%	27.8%
7	**test**	44.8%	18.4%	26.4%
8	**electr**	26.9%	1.4%	25.5%
9	**product**	61.7%	38.7%	23.0%
10	**assembl**	23.8%	1.2%	22.6%
11	**cad**	21.2%	0.2%	21.0%
12	**draw**	21.6%	1.9%	19.7%
13	**safeti**	20.7%	1.9%	18.9%
14	**autocad**	17.7%	0.1%	17.7%
15	**automot**	18.3%	0.8%	17.5%
16	**machin**	17.8%	0.9%	17.0%
17	**lab**	15.6%	1.1%	14.5%
18	**prototyp**	15.6%	1.1%	14.5%
19	**supplier**	16.4%	2.0%	14.5%
20	**project**	59.0%	44.6%	14.3%
21	**solidwork**	14.2%	0.2%	14.0%
22	**plant**	14.1%	0.5%	13.6%
23	**improv**	29.8%	16.4%	13.4%
24	**troubleshoot**	15.1%	1.9%	13.2%
25	**vehicl**	15%	2%	13%

Table 4.20 Attributes with higher frequency in Mech than in IT sorted by their difference

S. No.	Token	Frequency in Mech	Frequency in IT	Difference in Frequency (↓)
1	**mechan**	48.6%	1.3%	47.3%
2	**manufactur**	41.9%	4.1%	37.8%
3	**equip**	37.2%	5.4%	31.8%
4	**assist**	47.3%	21.9%	25.5%
5	**electr**	26.9%	4.4%	22.5%
6	**assembl**	23.8%	2.2%	21.6%
7	**cad**	21.2%	0.7%	20.5%
8	**engin**	77.2%	57.2%	20.0%
9	**prepar**	25.8%	6.1%	19.7%
10	**safeti**	20.7%	1.3%	19.4%
11	**materi**	20.7%	2.5%	18.2%
12	**draw**	21.6%	3.5%	18.1%
13	**automot**	18.3%	0.7%	17.6%
14	**autocad**	17.7%	0.6%	17.2%
15	**report**	37.8%	22.1%	15.6%
16	**vehicl**	15.4%	0.9%	14.5%
17	**supplier**	16.4%	2.3%	14.1%
18	**solidwork**	14.2%	0.2%	14.0%
19	**studi**	19.5%	5.8%	13.8%
20	**plant**	14.1%	0.8%	13.3%
21	**projectmanag**	22.1%	8.9%	13.2%
22	**control**	28.1%	15.7%	12.4%
23	**msoffic**	17.4%	5.4%	12.0%
24	**construct**	15.2%	3.4%	11.8%
25	**layout**	13%	1%	12%

We conclude that all the disciplines value soft skills and require some software skills. Looking into particular disciplines, we found that Fin places emphasis on grades, demonstrates the greatest need for soft skills, has more client interaction, more assistant positions and a more formal office environment. Mech features more design and field work, while IT includes core technical skills and offers more perks and collaborative work environments.

4.3 Comparison of lower and upper year jobs

In this section, we study the differences between the jobs obtained by lower year and upper year students. We compare the top 100 frequent and representative attributes to find the differences between the lower and upper year jobs of Fin (Section 4.3.1), IT (Section 4.3.2) and Mech (Section 4.3.3). Section 4.3.4 further investigates the common trends of all lower year and all upper year jobs (from all the three disciplines).

4.3.1 Fin Lower Year vs. Fin Upper Year

This section examines if and how Finance jobs obtained by its lower year students are different from the jobs obtained by its upper year students. We compare Finance Lower Year and Finance Upper Year using the Top 100 attributes that are most in demand or representative. Comparisons are made using all the methods listed in Section 3.2.5.

4.3.1.1 Attribute intersection

Using Jaccard Similarity to quantitatively compare the two groups, Figure 4.21 suggests that the frequent attributes of Lower and Upper Year Fin are 75% similar i.e. 75 of the 100 most frequent attributes of the two groups are the same. Figure 4.21 also reflects that Fin Lower and Upper Year are more similar to each other than Fin is to the other disciplines. This suggests that apart from generic attributes and soft skills (which was the main similarity between the frequent attributes of the different disciplines in Section 4.2.4.1), Fin lower year jobs and upper year jobs have more Fin-related attributes in common. This is confirmed by the Venn Diagram in Figure 4.23.

While Fin Lower and Upper Year have many common soft skills (including “team”, “communic”, “relationship”, “learn”, “lead” etc.) and work-related attributes (e.g. “experi”, “business”, “product” etc.), Figure 4.23 suggests that they also have Fin related attributes including “finance”, “account”, “audit”, “tax” etc.

As Figure 4.23 indicates, other frequent attributes common to both Lower and Upper Year Fin jobs include:

- “client”, suggesting that both upper and lower year jobs in Fin revolve around clients.
- “software”, “program” and “data”, suggesting a trend towards IT skills.
- “assist” and “support”, suggesting that Fin students work on assistant positions throughout their academic careers.
- “transcript”, reinforcing the emphasis that Fin places on grades.

Although Figure 4.23 shows that all the above attributes are frequent in both groups, Section 4.3.1.3 will reveal whether an attribute is more frequent in one group than another.

Looking at the Top 100 in-demand attributes that are present in the jobs of the lower year students of Fin but not in the upper year students of Fin (Figure 4.23), we obtain the followings insights:

- Attributes such “file”, “arrange” and “update” being common in the lower year jobs suggests that lower year students do more clerical work.

- Attributes such "database", "test" and "MS Office" in the Lower Year Fin jobs suggest that lower year Fin students have more IT-oriented jobs.

Zooming into Upper Year Fin (Figure 4.23) suggests the following conclusions.

- A high frequency of attributes such as "trade", "insurance", "capital", "invest" and "risk management" suggest that more Finance-specific jobs are available to upper year students.
- "modelling" and "statistics" are frequent in upper year Fin jobs suggesting the use of advanced Fin concepts
- "consult" and "control" appearing in Upper Year Fin jobs suggest that upper year students may be given more autonomy

Figure 4.22 compares Lower and Upper Year Fin using the top 100 representative attributes. Compared to the JS of the representative attributes of Fin and the other disciplines, the JS of the representative attributes of Fin Lower and Fin Upper Year indicates that they are more similar to each other. While Fin Lower and Upper Year share 60 out of the 100 attributes that represent them, each group has 40 attributes that define it more than the other group.

As seen in Figure 4.24, the intersection includes soft skills, some Fin-related skills and administrative components. Comparison of the representative attributes in Figure 4.24 reveals similar findings as suggested by the analysis of the frequent attributes earlier in this section.

- While lower year Fin students appear to have more clerical and assistant placements with less autonomy ("update", "arrange", "review", "maintain", "assist" etc.), Upper year students appear to be involved in "trade", "actuari", "risk management" etc.
    - While "written" and "oral" communication represents lower year students, "modelling" and "strategy" represent upper years.
    - While "English" represents lower year students, "mathematics" and "statistics" represent the upper years.
    - While "listen" represents lower years, "advisory" represents upper years.
- "program" and "vba" are representative skills of Fin's lower year and upper year jobs respectively, suggesting the need for IT skills in Fin.
- "MS Excel" is one of the top 100 representative attributes of both Lower and Upper Year Fin.

Overall, Figure 4.22 and Figure 4.24 suggest that although some Fin related work is done by lower year Fin students, Fin jobs feature more focus and autonomy in upper years.

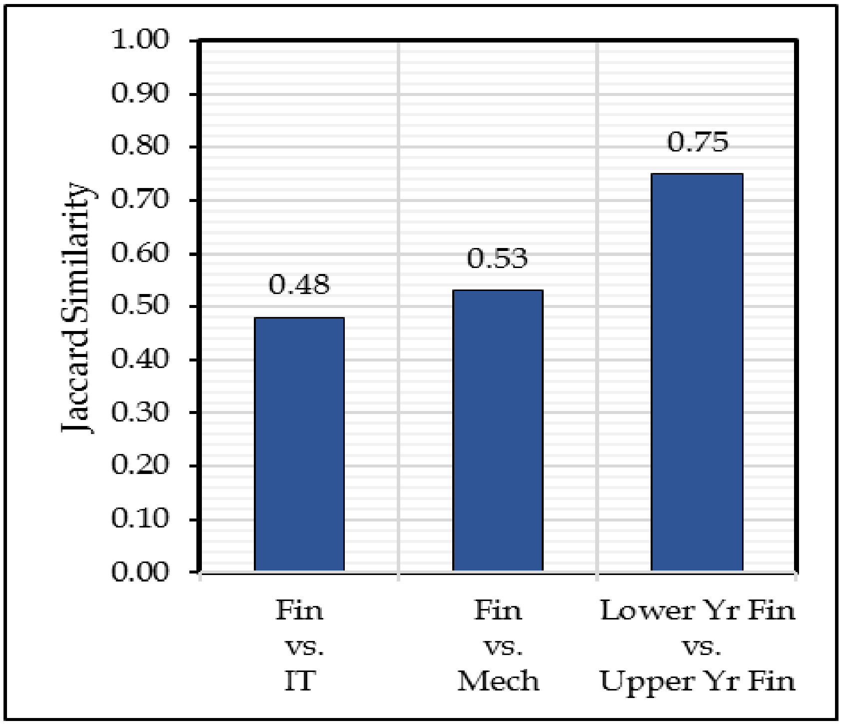

Figure 4.21 Comparison of top 100 most frequent attributes of Lower Year and Upper Year Fin using Jaccard similarity

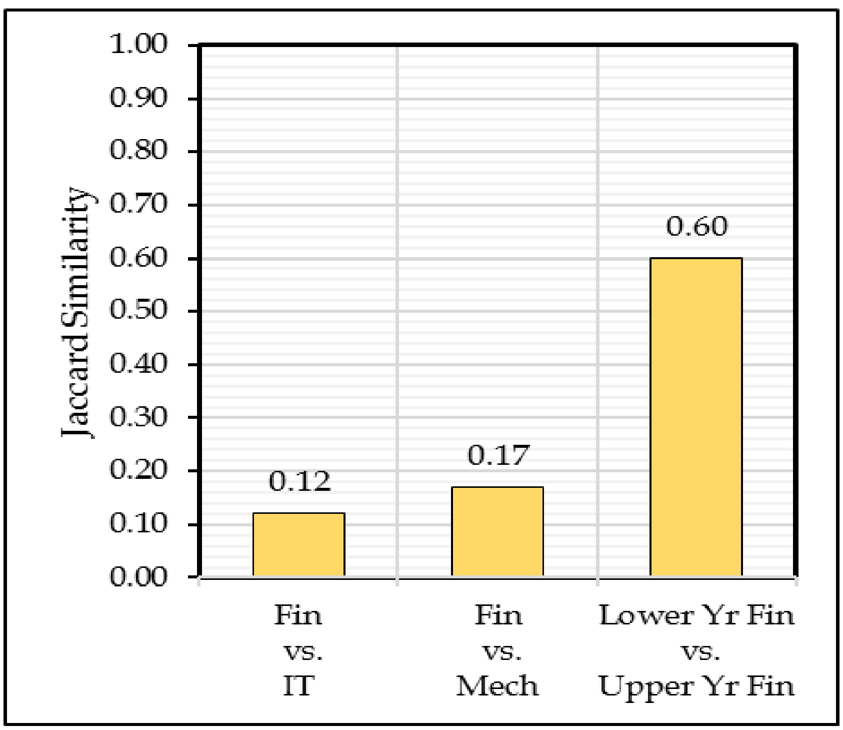

Figure 4.22 Comparison of top 100 most representative attributes of Lower Year and Upper Year Fin using Jaccard similarity

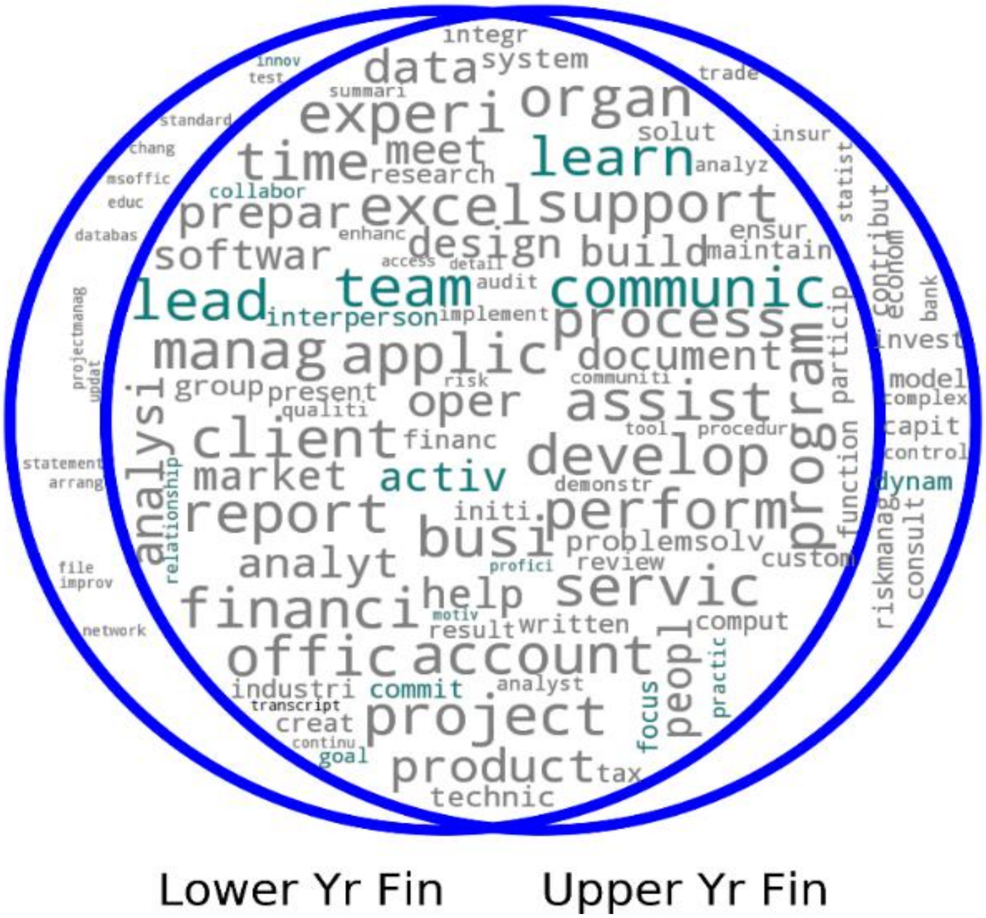

Figure 4.23 Overlap between the top 100 most frequent attributes of Lower Year and Upper Year Fin

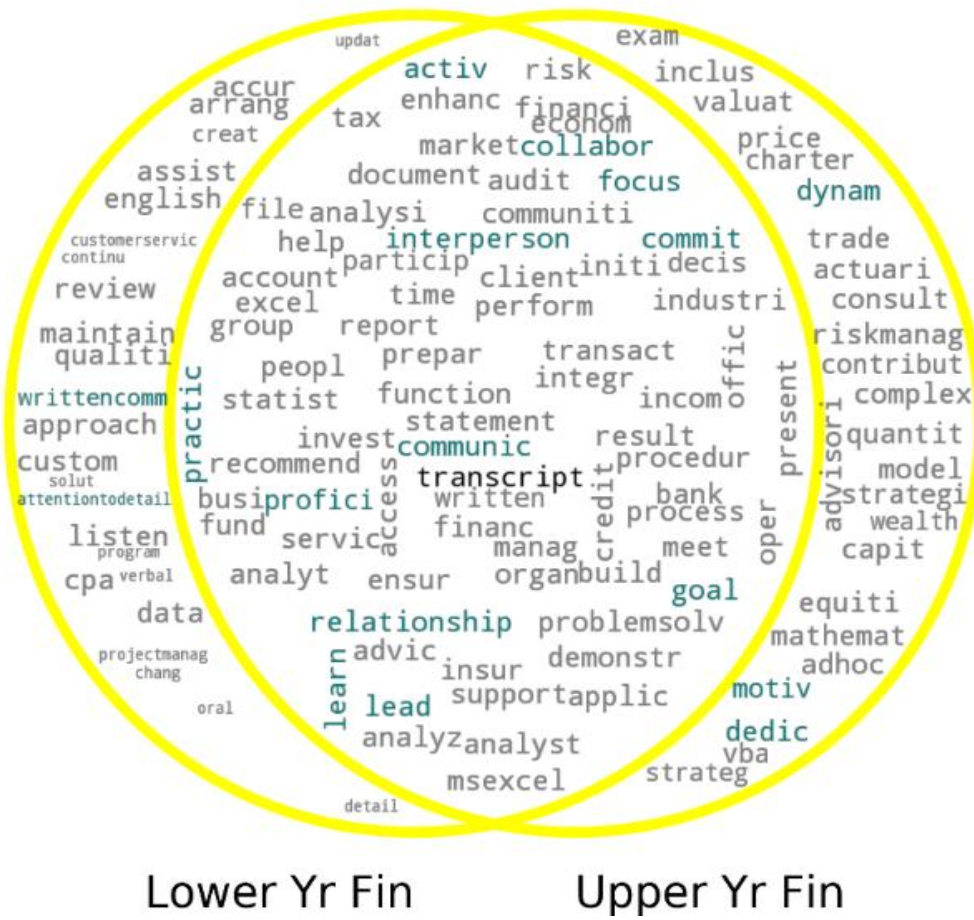

Figure 4.24 Overlap between the top 100 most representative attributes of Lower Year and Upper Year Fin

4.3.1.2 Distribution of types of attributes among the Top 100

Examining the distribution of the types of attributes in Lower and Upper Year Fin jobs in Figure 4.25 and Figure 4.26 reveals that Fin jobs do not offer many perks. Also, both Lower and Upper Year Fin jobs tend to require "transcripts".

Figure 4.25 indicates that soft skills are demanded equally in both Lower Year and Upper Year Fin. Figure 4.26 indicates that even though soft skills are demanded by lower year and upper year jobs, there are slightly more soft skills in the top 100 representative attributes of Upper Year than in Lower Year Fin.

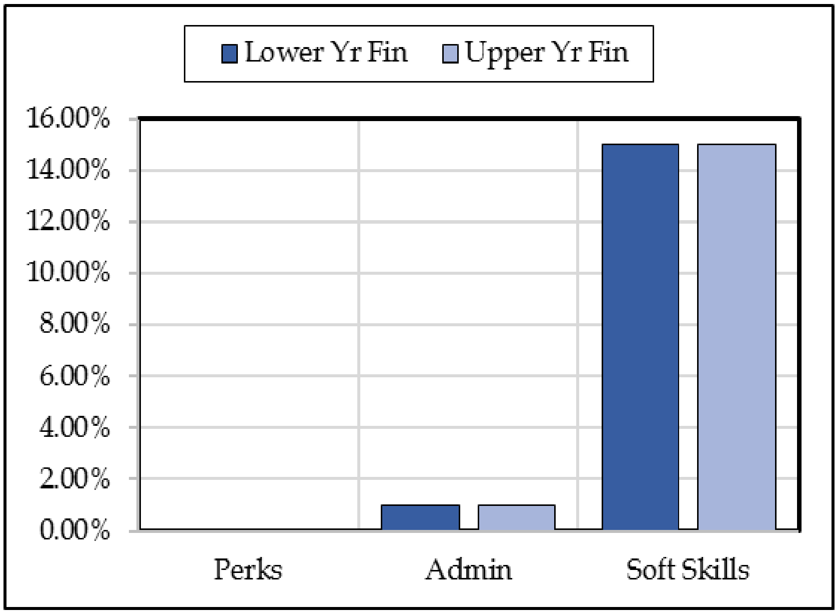

Figure 4.25 Distribution of the various types of attributes in the top 100 most frequent attributes of Lower Year and Upper Year Fin

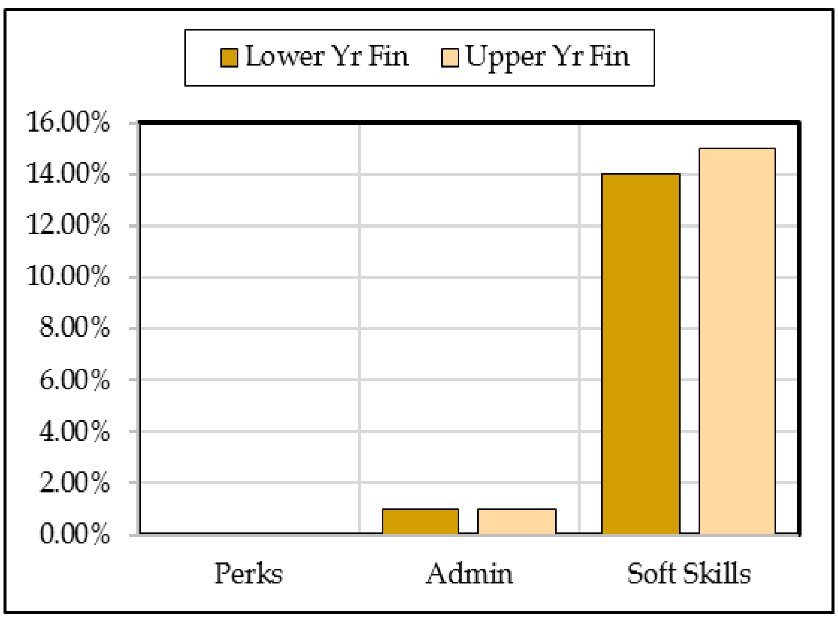

Figure 4.26 Distribution of the various types of attributes in the top 100 most representative attributes of Lower Year and Upper Year Fin

4.3.1.3 Attributes with a higher frequency in one group than the other

Comparing lower year Fin jobs to upper year Fin jobs in terms of the difference in demand they place on different attributes, we reinforce the findings of Section 4.3.1.1 and Section 4.3.1.2 and also draw new insights.

Table 4.21 Attributes with higher frequency in Lower Year Fin than in Upper Year Fin sorted by their difference

S. No.	Token	Frequency in Lower Yr Fin	Frequency in Upper Yr Fin	Difference in Frequency (↓)
1	**arrang**	17.3%	5.7%	11.6%
2	**advertis**	14.0%	3.8%	10.3%
3	**english**	14.3%	4.8%	9.5%
4	**updat**	16.3%	7.8%	8.6%
5	**document**	39.2%	30.9%	8.2%
6	**assist**	56.4%	48.2%	8.1%
7	**system**	33.8%	26.6%	7.3%
8	**accur**	15.0%	7.8%	7.2%
9	**listen**	11.8%	4.8%	6.9%
10	**customerservic**	12.6%	5.9%	6.7%
11	**maintain**	28.3%	22.2%	6.1%
12	**web**	10.2%	4.2%	6.0%
13	**languag**	12.5%	6.7%	5.8%
14	**custom**	27.7%	22.0%	5.7%
15	**softwar**	34.3%	28.6%	5.7%
16	**qualiti**	26.0%	20.5%	5.5%
17	**brand**	8.1%	3.2%	4.8%
18	**cpa**	7.8%	3.0%	4.8%
19	**email**	7.2%	2.8%	4.4%
20	**event**	11.9%	7.6%	4.2%
21	**inquiri**	5.7%	1.6%	4.1%
22	**payment**	7.2%	3.2%	4.0%
23	**projectmanag**	19.5%	15.5%	4.0%
24	**project**	47.1%	43.2%	3.9%
25	**databas**	17%	14%	4%

Table 4.22 Attributes with higher frequency in Upper Year Fin than in Lower Year Fin sorted by their difference

S. No.	Token	Frequency in Upper Yr Fin	Frequency in Lower Yr Fin	Difference in Frequency (↓)
1	**invest**	33.0%	16.1%	16.9%
2	**audit**	34.0%	19.4%	14.6%
3	**group**	41.3%	27.5%	13.7%
4	**analysi**	48.8%	35.1%	13.7%
5	**financ**	36.3%	23.1%	13.3%
6	**commit**	35.1%	22.4%	12.7%
7	**industri**	38.2%	25.6%	12.7%
8	**riskmanag**	21.7%	9.4%	12.4%
9	**dynam**	22.7%	10.3%	12.3%
10	**client**	53.9%	41.8%	12.1%
11	**financi**	65.9%	53.9%	12.0%
12	**model**	26.7%	14.8%	11.9%
13	**develop**	70.8%	59.0%	11.9%
14	**offic**	53.5%	42.5%	11.0%
15	**price**	16.9%	6.1%	10.8%
16	**inclus**	12.5%	2.0%	10.5%
17	**design**	42.1%	31.8%	10.3%
18	**result**	31.4%	21.3%	10.1%
19	**transcript**	27.9%	18.0%	10.0%
20	**risk**	28.0%	18.4%	9.6%
21	**actuari**	17.4%	7.7%	9.6%
22	**capit**	20.7%	11.2%	9.5%
23	**complex**	18.2%	8.8%	9.4%
24	**lead**	50.8%	41.5%	9.4%
25	**servic**	63%	54%	9%

In line with the previous sections, we find that

- Fin students appear to do more clerical work in their lower years than in their upper years. ("arrange", "document", "assist" and "English" appear more frequently in lower years)
- Lower Year Fin students appear to take up more IT oriented jobs than upper year Fin students. ("web" and "software" appear more often in lower year postings)
- Upper year jobs involve core Fin skills. (Table 4.22 contains many Fin related work profiles and skills)
- Upper year Fin students appear to have more autonomy in their upper years. (While "lead" appears 9% more in upper year jobs than in lower years, "listen" appears almost 6% more frequently in lower years than upper)
- Transcripts are demanded almost 10% more often in upper year job postings than in lower year job postings. This suggests that the importance of grades in Fin increases in upper years.

- Figure 4.26 of Section 4.2.4.2 indicated that Upper Year Fin jobs are represented by soft skills more than Lower year Fin. Table 4.22 reinforces this finding indicating that soft skills are more representative of Upper Year Fin.
- “client” is mentioned 12% more in upper year than in lower year suggesting that client interaction increases in upper years. This might complement the need for more soft skills in upper years.

Overall, the results in this Section suggest that lower year Fin jobs involve more clerical and IT work with less autonomy, whereas upper year Fin jobs focus on analyzing and solving financial problems.

4.3.2 IT Lower Year vs. IT Upper Year

Using all the methods listed in Section 3.2.5, this section examines the differences between lower year and upper year IT jobs in terms of their top 100 most frequent and most representative attributes.

4.3.2.1 Attribute intersection

Similar to Finance, Figure 4.27 shows that the JS of the Top 100 frequent attributes of IT Lower Year vs. IT Upper Year is higher than the JS of IT vs. any other discipline. Figure 4.29 confirms that apart from generic and soft skill attributes, IT Lower and Upper Year jobs also share core IT skills, e.g., “java”, “javascript”, “OOP”, “C”, “Android” etc.

Likewise, Figure 4.28 shows that the top 100 representative attributes of IT Lower and Upper Year are 60% similar while the representative attributes of the different disciplines are <20% similar to IT. This suggests that IT Lower and Upper Year are not as distinct as two different disciplines. The Venn Diagram in Figure 4.30 suggests that even though some core IT skills are more representative of Upper year students, many of them commonly represent both levels.

As expected, the top 100 frequently occurring attributes have more in common than the top 100 representative attributes.

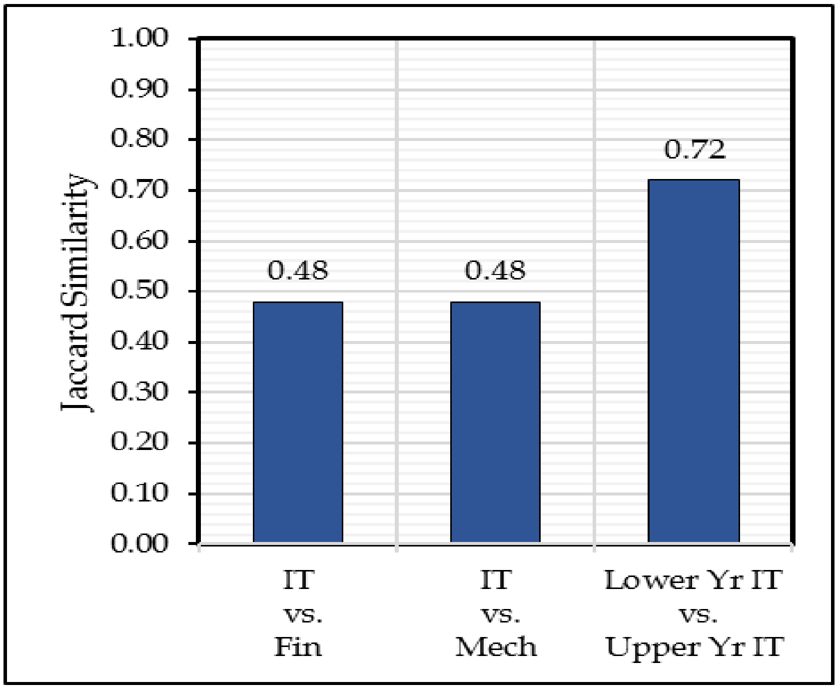

Figure 4.27 Comparison of top 100 most frequent attributes of Lower Year and Upper Year IT using Jaccard similarity

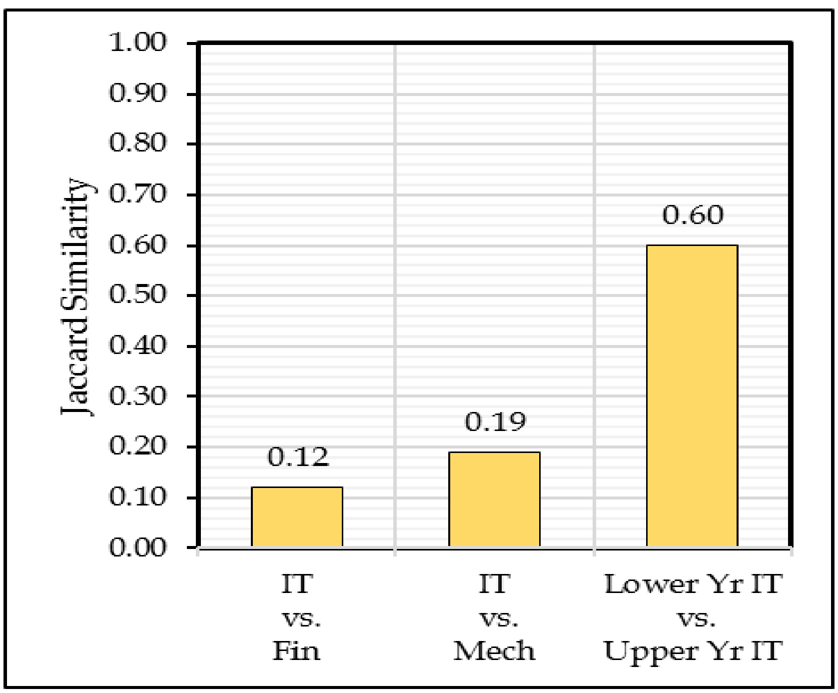

Figure 4.28 Comparison of top 100 most representative attributes of Lower Year and Upper Year IT using Jaccard similarity

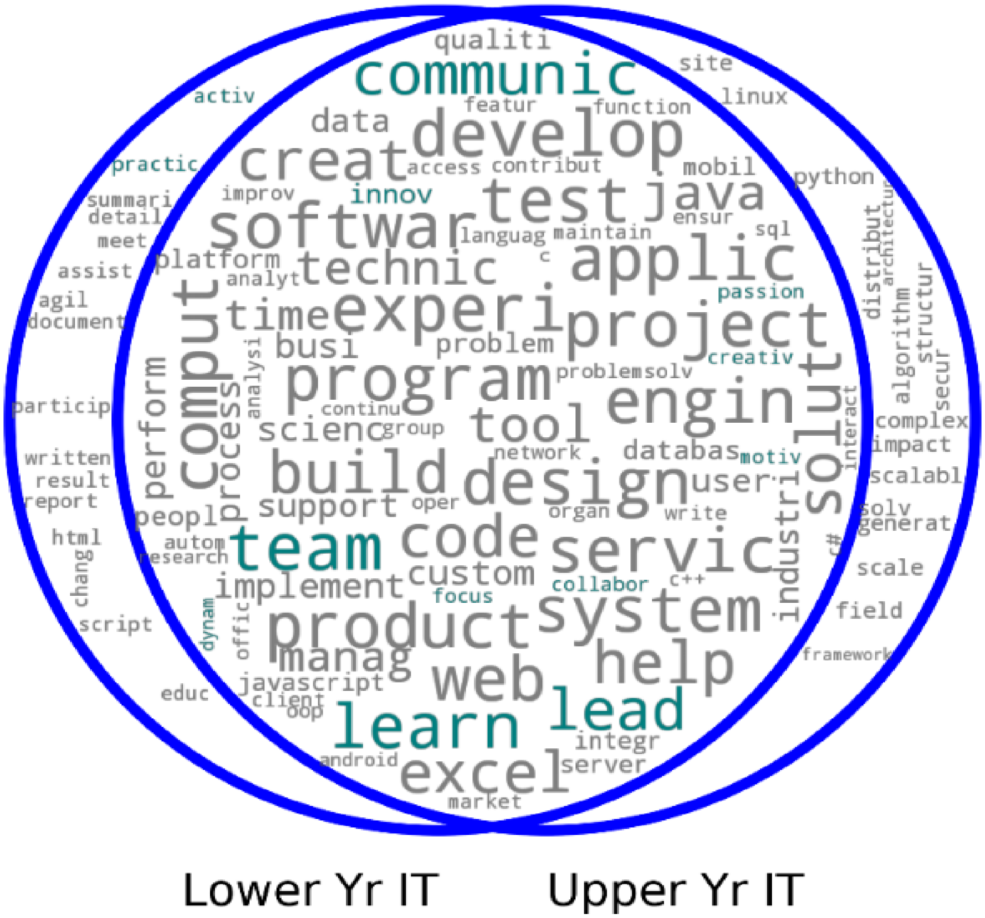

Figure 4.29 Overlap between the top 100 most frequent attributes of Lower Year and Upper Year IT

Zooming into the Venn Diagram of the frequent attributes in the Lower and Upper Year IT (shown in Figure 4.29) suggests the following:

- IT Lower and Upper Year frequently require core IT skills related to development, coding and testing. This suggests that IT jobs are specific to their discipline even in the lower years.
- Both IT Lower and Upper Year emphasize soft skills including "team", "communic", "learning" and mindset related soft skills, e.g., "passion", "creativity", "motivation" and "innovation".
- "html" is more frequent in Lower Year IT than in Upper Year, emphasizing it to be a beginner's skill.
- Lower year IT also mentions "practic" standing for practical experience and/or practice, indicating the emphasis on practical learning in IT.
- Apart from the above, the presence of attributes like "report", "document", "assist", "summarize", "written" etc. emphasizes that Lower Year IT includes more clerical work.
- Apart from offering programming jobs, upper year IT also offers jobs dealing with advanced technologies including "linux", "python", "distributed computing", "security", "architecture", "scalable", "framework", "algorithm" etc.

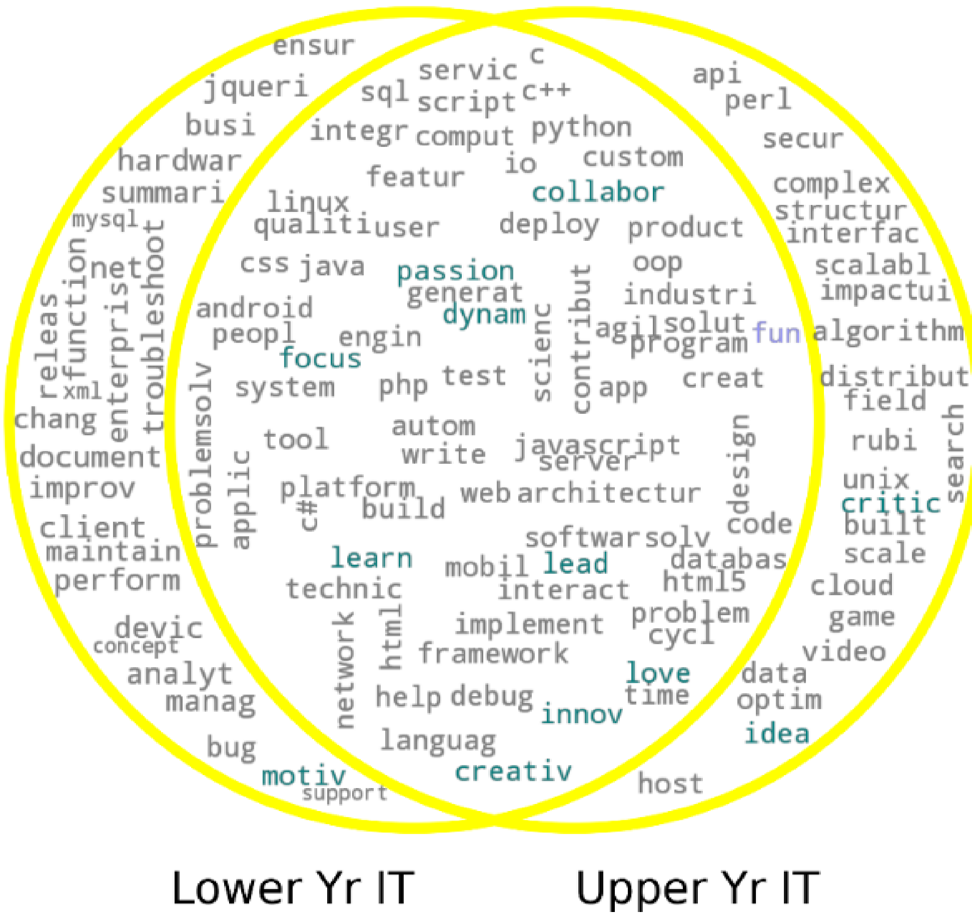

Figure 4.30 Overlap between the top 100 most representative attributes of Lower Year and Upper Year IT

Zooming into the Venn Diagram of the representative attributes of the Lower and Upper Year IT (shown in Figure 4.30) indicates the following:

- Core IT skills of programming languages, web and mobile development equally represent both Lower and Upper IT suggesting that irrespective of level, IT students work on core IT areas from the beginning of their co-op careers.
- IT Lower and Upper Year both require "passion", "love", "focus", "creativity" and "innovation", and offer a "dynamic", "collaborative" and a "fun" environment.
- Lower year IT can be represented by some clerical skills of summarizing and documenting and some technical skills such as "jquery", "XML" and MySQL.
- While Lower Year IT has more testing jobs involving "troubleshooting" and finding "bugs", upper year IT works with more advanced and upcoming concepts, e.g., "algorithm", "cloud", "security", "scalable" etc.
- Perl and Ruby (programming languages) are more representative of upper year IT.
- Lower year job advertisements emphasize "motivation" while upper year jobs place more importance on "critical thinking" and "ideas".

4.3.2.2 Distribution of types of attributes among the Top 100

As Figure 4.31 and Figure 4.32 suggest, *Perks* although not in the top 100 frequent attributes of IT, find equal representation in both Lower and Upper Year IT jobs. Additionally, IT jobs do not appear to emphasize any administrative requirements.

Figure 4.31 shows that out of the 100 most frequent attributes of the two groups, soft skills are more demanded in Lower Year jobs than in Upper Year. However, Figure 4.32 shows that there are more soft skills found in the top 100 attributes that represent Upper Year IT than in the top 100 attributes that represent Lower Year IT. This suggests that soft skills, even though less frequently mentioned in the Upper Year IT, represent it more closely.

The explicit mention of soft skills in Lower Year jobs versus its absence in Upper Year could stem from the employers' notion that such skills would not exist in all lower year applicants to the same degree as they can be assumed to exist in upper year students. Thus, even though upper year jobs place more importance on soft skills (Figure 4.32), perhaps they are not explicitly mentioned in their postings (Figure 4.31).

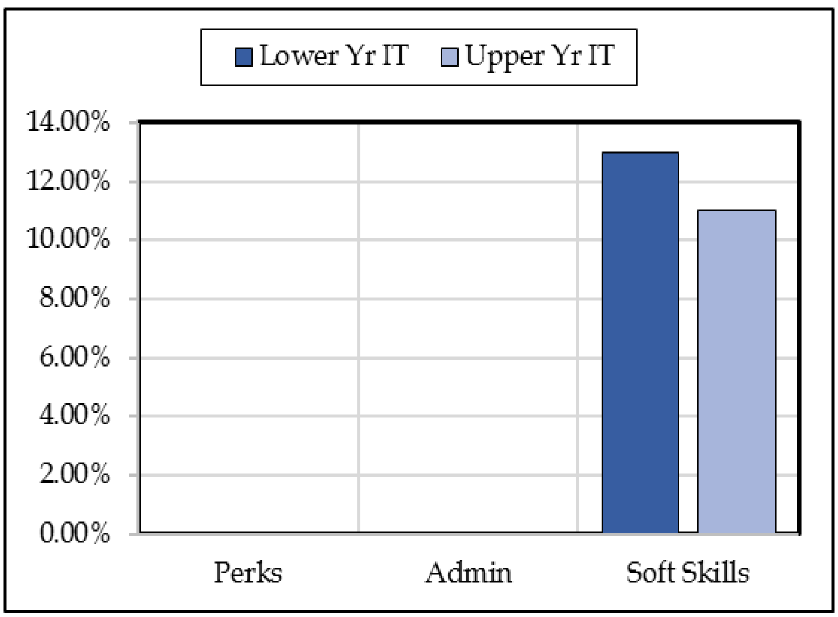

Figure 4.31 Distribution of the various types of attributes in the top 100 most frequent attributes of Lower Year and Upper Year IT

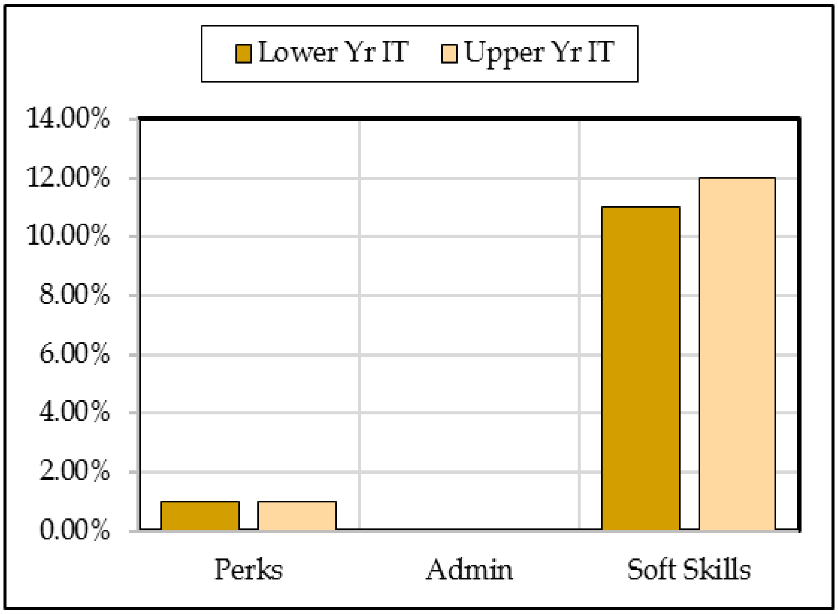

Figure 4.32 Distribution of the various types of attributes in the top 100 most representative attributes of Lower Year and Upper Year IT

4.3.2.3 Attributes with a higher frequency in one group than in the other

Table 4.23 and Table 4.24 confirm the findings of the previous sections.

- As seen in Table 4.23 and Table 4.24, Lower Year IT includes more soft skills while Upper Year IT emphasize technical skills.
- While Lower Year IT have jobs that require more clerical work and junior positions including "documenting", "assisting", "reporting", "writing" etc., core IT skills are required by upper Year IT.
- Lower year IT includes "troubleshooting", "installing", "MS Office", "testing" while Upper Year IT focuses on user-centred development using various programming languages. This is suggested by >9% frequency of "user", "features" and "design" in Upper Year IT than in Lower Year IT (Table 4.24).
- "html" and "SQL" are found 5% and 4% more frequently in Lower Year IT than in Upper Year, suggesting that these skills are beginner skills.

Table 4.23 Attributes with higher frequency in Lower Year IT than in Upper Year IT sorted by their difference

S. No.	Token	Frequency in Lower Yr IT	Frequency in Upper Yr IT	Difference in Frequency (↓)
1	document	28.8%	15.9%	13.0%
2	support	42.5%	31.4%	11.1%
3	assist	26.8%	15.9%	10.8%
4	manag	47.5%	37.6%	10.0%
5	communic	52.7%	42.8%	10.0%
6	test	53.9%	44.8%	9.0%
7	report	26.1%	17.4%	8.7%
8	busi	42.5%	33.8%	8.6%
9	written	21.1%	13.4%	7.8%
10	activ	22.9%	15.3%	7.6%
11	educ	17.0%	9.6%	7.4%
12	standard	15.3%	8.3%	7.1%
13	interperson	13.3%	6.5%	6.8%
14	instal	9.3%	2.8%	6.5%
15	troubleshoot	15.1%	8.7%	6.5%
16	msoffic	8.3%	2.0%	6.3%
17	attentiontodetail	10.7%	4.7%	6.0%
18	summari	23.7%	17.7%	6.0%
19	execut	15.2%	9.2%	6.0%
20	account	12.3%	6.5%	5.9%
21	track	14.5%	8.7%	5.8%
22	updat	10.6%	5.1%	5.5%
23	organ	28.8%	23.4%	5.4%
24	resolut	7.6%	2.2%	5.3%
25	prepar	8%	3%	5%

Table 4.24 Attributes with higher frequency in Upper Year IT than in Lower Year IT sorted by their difference

S. No.	Token	Frequency in Upper Yr IT	Frequency in Lower Yr IT	Difference in Frequency (↓)
1	c++	45.9%	21.4%	24.5%
2	engin	70.5%	46.1%	24.4%
3	algorithm	28.4%	8.6%	19.8%
4	scale	27.6%	8.7%	18.8%
5	scienc	48.9%	31.4%	17.5%
6	featur	38.9%	21.9%	17.0%
7	python	30.8%	14.4%	16.4%
8	scalabl	22.9%	7.2%	15.7%
9	data	46.1%	30.5%	15.6%
10	build	56.8%	41.5%	15.4%
11	code	54.5%	39.8%	14.7%
12	complex	26.5%	13.0%	13.5%
13	comput	59.0%	45.8%	13.1%
14	c	31.1%	18.3%	12.8%
15	product	69.2%	56.6%	12.6%
16	field	23.2%	10.8%	12.4%
17	structur	21.2%	8.8%	12.4%
18	java	50.0%	37.7%	12.3%
19	distribut	23.3%	12.0%	11.3%
20	search	16.4%	6.1%	10.4%
21	problem	39.6%	29.3%	10.3%
22	system	63.2%	53.2%	10.0%
23	help	50.7%	41.1%	9.6%
24	design	70.3%	60.7%	9.6%
25	user	40%	31%	9%

To summarize, while jobs obtained by Lower Year IT students involve some technical skills such as HTML and SQL, working with advanced software and platforms (cloud, scale, security) is more common in Upper Years. Furthermore, Lower Year IT jobs involve more troubleshooting, testing and documenting.

4.3.3 Mech Lower Year vs. Mech Upper Year

This section compares the attributes of the Mech Lower Year and Mech Upper Year jobs. Similar to the previous sections, it uses all the methods listed in Section 3.2.5 to compare the Top 100 frequent and representative attributes to find major differences between the two groups.

4.3.3.1 Attribute intersection

As seen in the other disciplines, Figure 4.33 and Figure 4.34 suggest that Mech Lower and Upper Year are more similar to each other than Mech is to any other discipline. This is because as seen in Figure 4.35 and Figure 4.36, the top 100 frequent and representative attributes of Mech Lower and Upper Year share some typical Mech skills that are absent in any other discipline.

Looking at Figure 4.35 reveals the following insight about the Top 100 frequent attributes of Mech Lower Year and Mech Upper Year:

- Apart from core Mech skills, e.g., “assemble”, “prototype”, “cad” etc., Mech Lower and Upper Year have common soft skills including “learn”, “communic”, “team”, “innovation” etc.
- Mech Upper and Lower Year both mention “practic” standing for practical experience or practice.
- Attributes like “assist”, “report” and “supervise” are found in both Mech Upper and Lower Year top 100 frequent attributes. We will investigate this further in Section 4.2.4.3.
- “software”, “program” and “MS Office” are mentioned in both Mech Upper and Lower Year. Section 4.2.4.3 will further investigate whether Mech Lower Year has a greater demand for IT oriented skills.
- Apart from the core Mech skills frequent in both the groups, additional attributes contained in only Lower Year Mech exemplify clerical work, e.g., “write”, “update”, “change” and “procedure” and IT related work (suggested by “website”).
- Lower Mech jobs also mention “field” and “client” more often.
- Upper Year Mech specify “troubleshooting”, “costing”, “packaging” and “transport” more often.
- Upper Year Mech jobs also specify the demand of more soft skills including handling a “dynamic” environment, “interpersonal” skills and “commitment”.
- While Lower year Mech jobs mention “labs”, Upper Year Mech jobs mention “plants”. This might suggest a work profile shift.

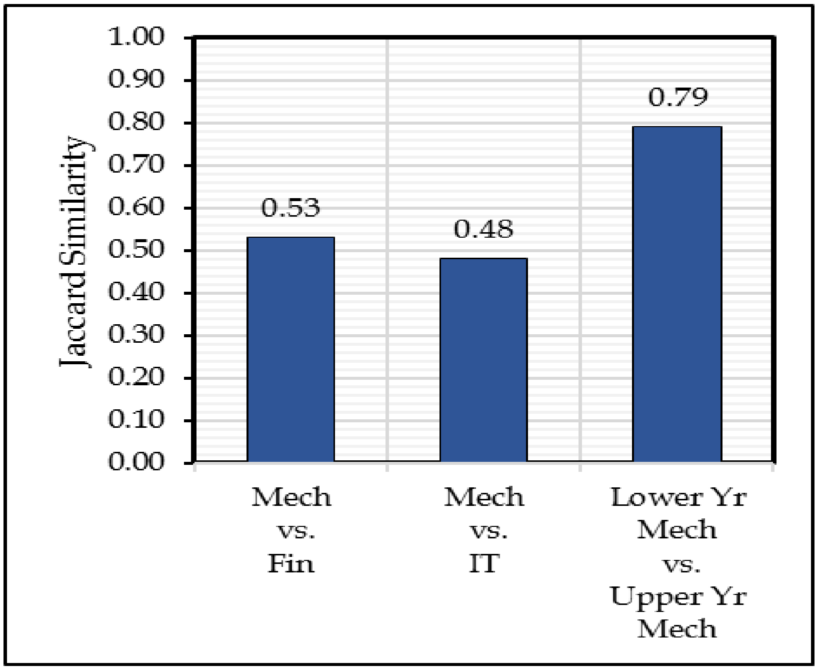

Figure 4.33 Comparison of top 100 most frequent attributes of Lower Year and Upper Year Mech using Jaccard similarity

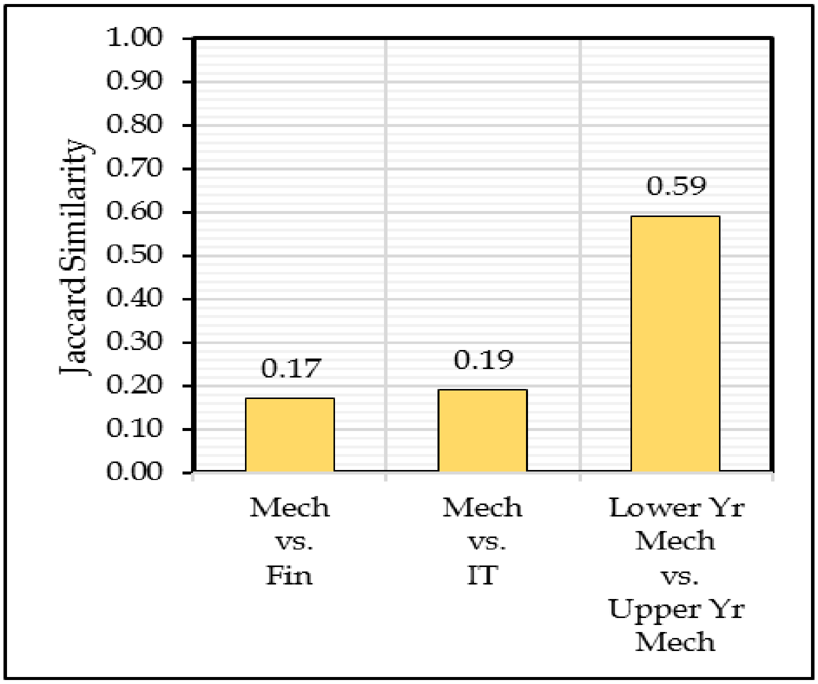

Figure 4.34 Comparison of top 100 most representative attributes of Lower Year and Upper Year Mech using Jaccard similarity

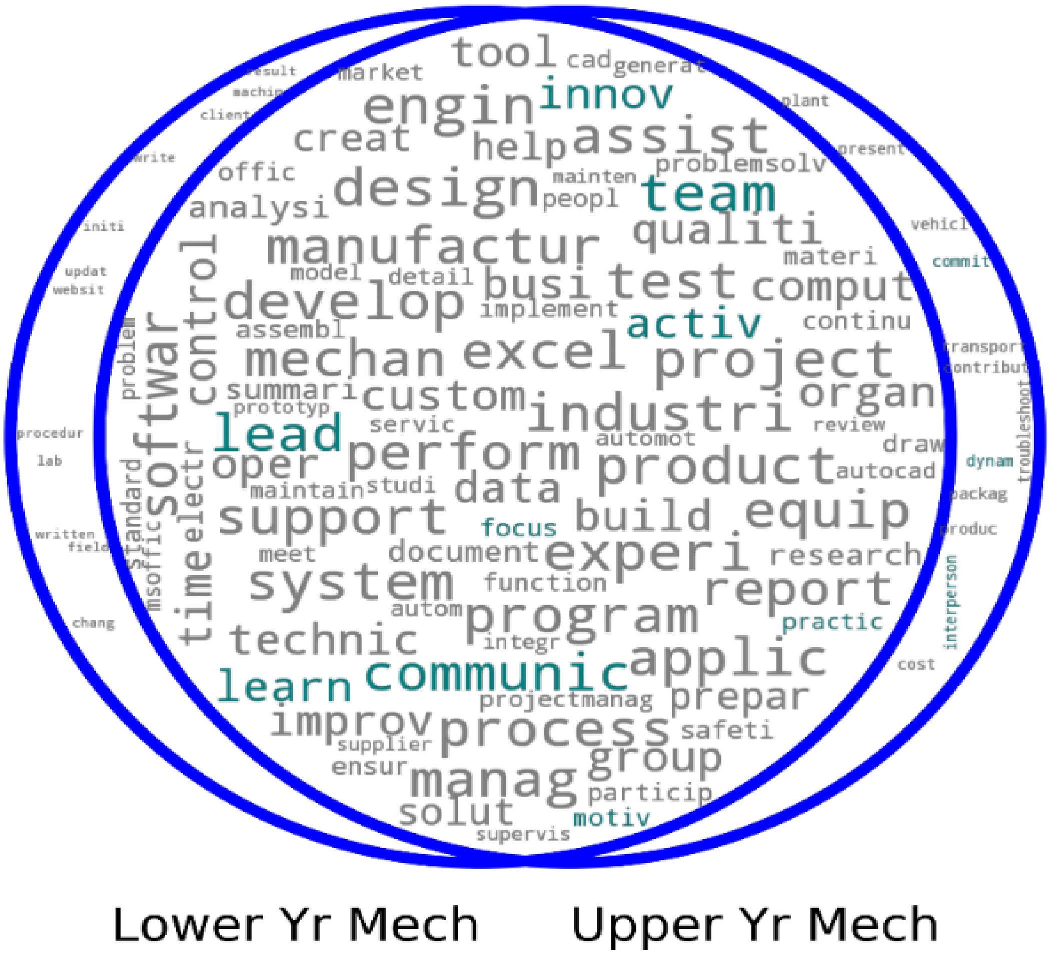

Figure 4.35 Overlap between the top 100 most frequent attributes of Lower Year and Upper Year Mech

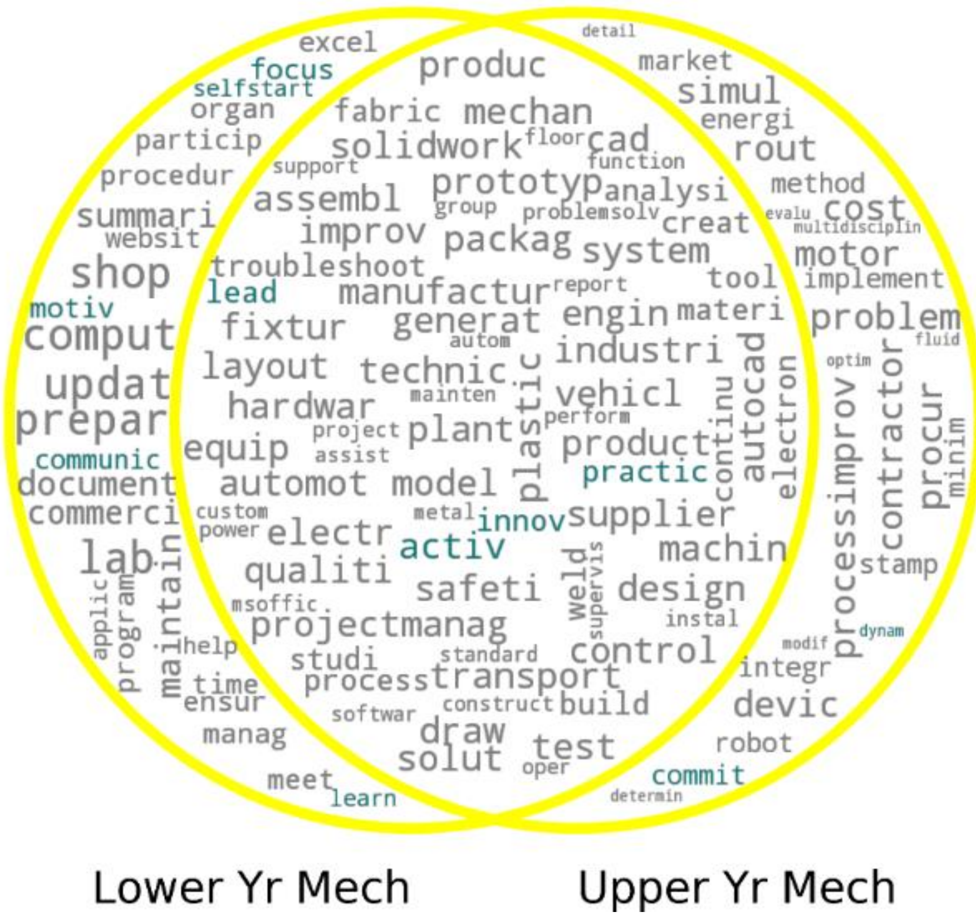

Figure 4.36 Overlap between the top 100 most representative attributes of Lower Year and Upper Year Mech

Zooming into the top 100 most representative attributes of Lower and Upper Year Mech, we observe the following:

- Besides core Mech skills of “power”, “solidwork”, “weld”, “autocad”, “manufacture”, etc. and soft skills like “innovative”, both Mech Upper and Lower Year mention “software” and “MS Office”.
- Lower year Mech’s representation is dominated by soft skills including “communic”, “motivation”, “self-starter” and “focus”.
- While Lower Year Mech contain clerical attributes including “prepare”, “update” etc., IT skills are also representative of Lower Year Mech (“prorgram”, “website” etc.).
- Upper Year Mech contains many core Mech skills including “simulation”, “processs improvement”, “robot”, “energy”, “fluid” etc.

4.3.3.2 Distribution of types of attributes among the Top 100

Figure 4.37 and Figure 4.38 indicate that Mech Lower and Upper Year do not include attributes related to Perks or Administrative requirements. Furthermore, it appears that soft skills are more frequent in upper years whereas specific technical skills become more important in upper years.

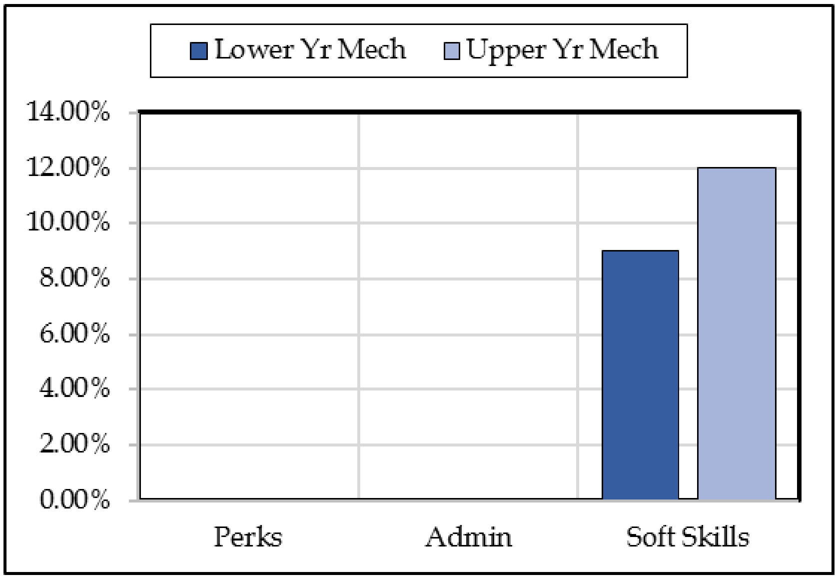

Figure 4.37 Distribution of the types of attributes in the top 100 most frequent attributes of Lower Year and Upper Year Mech

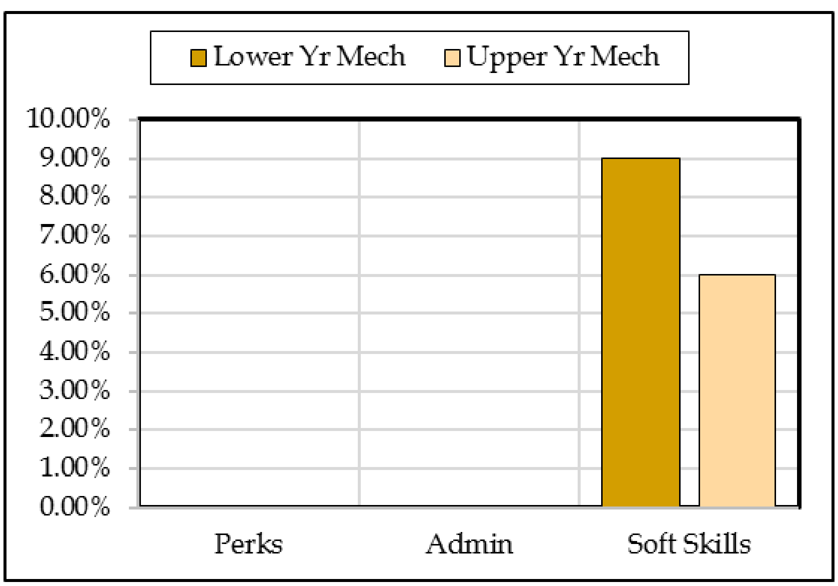

Figure 4.38 Distribution of the types of attributes in the top 100 most representative attributes of Lower Year and Upper Year Mech

4.3.3.3 Attributes with a higher frequency in one group than in the other

Table 4.25 and Table 4.26 summarize how the frequency of attributes of the lower year Mech jobs differ from that of the upper year Mech jobs. The main findings include:

- Clerical work, e.g., “update”, “maintain”, “arrange”, “email”, “written” etc. are mentioned more frequently in lower year Mech jobs.
- IT related skills, marked by “database”, “compute”, “server”, “platform”, “language”, “sql”, “web” etc., have a higher frequency in Lower Year Mech jobs, suggesting that Lower Year Mech students take up IT jobs. “MS Office” is mentioned equally in both groups.
- Upper Year Mech contains more core Mech skills including “mechan”, “cad”, “manufacture” etc.
- While “project management” and “supervise” is 10% and 3% more frequent in Upper Year than in Lower year, attributes like “assist”, “support”, “report” etc. appear equally in both the groups.
- Attributes including “client”, “custom” and “meet” are more frequent in Lower Year Mech while “create”, “implement”, “design”, “analysis”, “evaluate” etc. have a higher frequency in Upper Year Mech. This could suggest that Lower Year students make field visits to collect requirements from clients while Upper Year Mech design and implement solutions.
- While “adapt” is mentioned more in Lower year Mech, “team” is mentioned more Upper Year.

Table 4.25 Attributes with higher frequency in Lower Year Mech than in Upper Year Mech sorted by their difference

S. No.	Token	Frequency in Lower Yr Mech	Frequency in Upper Yr Mech	Difference in Frequency (↓)
1	**client**	16.4%	8.4%	8.0%
2	**custom**	35.7%	28.1%	7.7%
3	**databas**	13.4%	5.9%	7.5%
4	**servic**	24.6%	17.3%	7.3%
5	**updat**	17.2%	10.1%	7.1%
6	**advertis**	11.7%	5.1%	6.7%
7	**comput**	36.9%	30.4%	6.5%
8	**meet**	23.8%	17.4%	6.4%
9	**maintain**	23.0%	16.8%	6.2%
10	**arrang**	12.7%	6.7%	6.1%
11	**server**	8.2%	2.4%	5.9%
12	**platform**	12.2%	7.1%	5.0%
13	**languag**	10.5%	5.7%	4.8%
14	**written**	17.1%	12.8%	4.2%
15	**experi**	72.2%	68.0%	4.2%
16	**web**	13.9%	9.8%	4.1%
17	**deliveri**	8.1%	4.1%	4.0%
18	**framework**	4.7%	0.8%	4.0%
19	**sql**	7.2%	3.3%	3.9%
20	**email**	5.7%	1.7%	3.9%
21	**check**	8.1%	4.3%	3.8%
22	**excel**	41.1%	37.4%	3.7%
23	**shop**	10.2%	6.5%	3.7%
24	**manual**	6.0%	2.5%	3.5%
25	**adapt**	6%	2%	3%

Table 4.26 Attributes with higher frequency in Upper Year Mech than in Lower Year Mech sorted by their difference

S. No.	Token	Frequency in Upper Yr Mech	Frequency in Lower Yr Mech	Difference in Frequency (↓)
1	**mechan**	60.9%	42.2%	18.6%
2	**design**	80.3%	63.3%	17.1%
3	**electr**	35.5%	22.4%	13.1%
4	**manufactur**	50.1%	37.6%	12.5%
5	**engin**	84.6%	73.4%	11.3%
6	**activ**	36.5%	25.5%	11.0%
7	**projectmanag**	29.0%	18.5%	10.5%
8	**problem**	26.1%	16.1%	10.1%
9	**model**	27.4%	17.7%	9.8%
10	**technic**	42.5%	32.7%	9.8%
11	**improv**	36.0%	26.6%	9.4%
12	**assembl**	30.0%	20.6%	9.4%
13	**analysi**	32.8%	23.7%	9.1%
14	**solut**	32.5%	23.8%	8.7%
15	**system**	62.6%	54.0%	8.6%
16	**transport**	21.4%	13.0%	8.4%
17	**creat**	33.0%	24.6%	8.3%
18	**practic**	25.7%	17.8%	7.9%
19	**team**	72.6%	64.9%	7.7%
20	**cad**	25.8%	18.7%	7.1%
21	**plant**	18.7%	11.7%	7.0%
22	**implement**	26.8%	19.9%	6.9%
23	**evalu**	15.8%	9.2%	6.7%
24	**integr**	22.0%	15.4%	6.6%
25	**build**	37%	31%	7%

This section suggests that Lower Year Mech students are involved in more clerical and IT related work, while Upper Year Mech jobs focus on designing and implementing solutions. However, "assistant" positions are common in lower and upper years.

4.3.4 Lower year vs. upper year analysis across disciplines

After comparing the lower and upper year attributes of each discipline independently, we now compare lower and upper year jobs across all disciplines.

4.3.4.1 Similarities between lower year students across the disciplines

Examining Figure 4.39 which shows the frequent attributes mentioned in lower year jobs across all disciplines, we observe that many of the top 100 frequently mentioned attributes are common.

- While many generic attributes, e.g., “experi”, “busi”, “organ” are found, so are attributes, e.g., “document”, “report” and “summary”. This indicates the clerical nature of lower-year jobs.
- While leadership (“lead”) appears in lower year jobs, attributes like “assist” and “support” are common, suggesting that lower year students work in junior positions regardless of discipline.
- The frequency of IT related skills including “software”, “comput” and “program” suggests that lower year students of all disciplines obtain IT jobs early in their careers. While Lower Year Fin and Mech shares “MS Office”, Lower Year Fin and IT share “database”, suggesting a trend towards IT in the lower year jobs of all disciplines.
- Soft Skills including teamwork, motivation, communication and life-long learning appear frequently in all lower year jobs (“team”, “communic”, “learn”, “motivation”).
- While the above soft skills are emphasized by all Lower Year jobs, Lower Year IT additionally includes “passion” and “creativity” and Lower Year Fin includes “commitment”, “goal” and “interpersonal” skills. No special soft skill is mentioned by Mech.
- As seen in Figure 4.39, some core attributes of each discipline are frequent among each discipline’s lower year jobs as well.

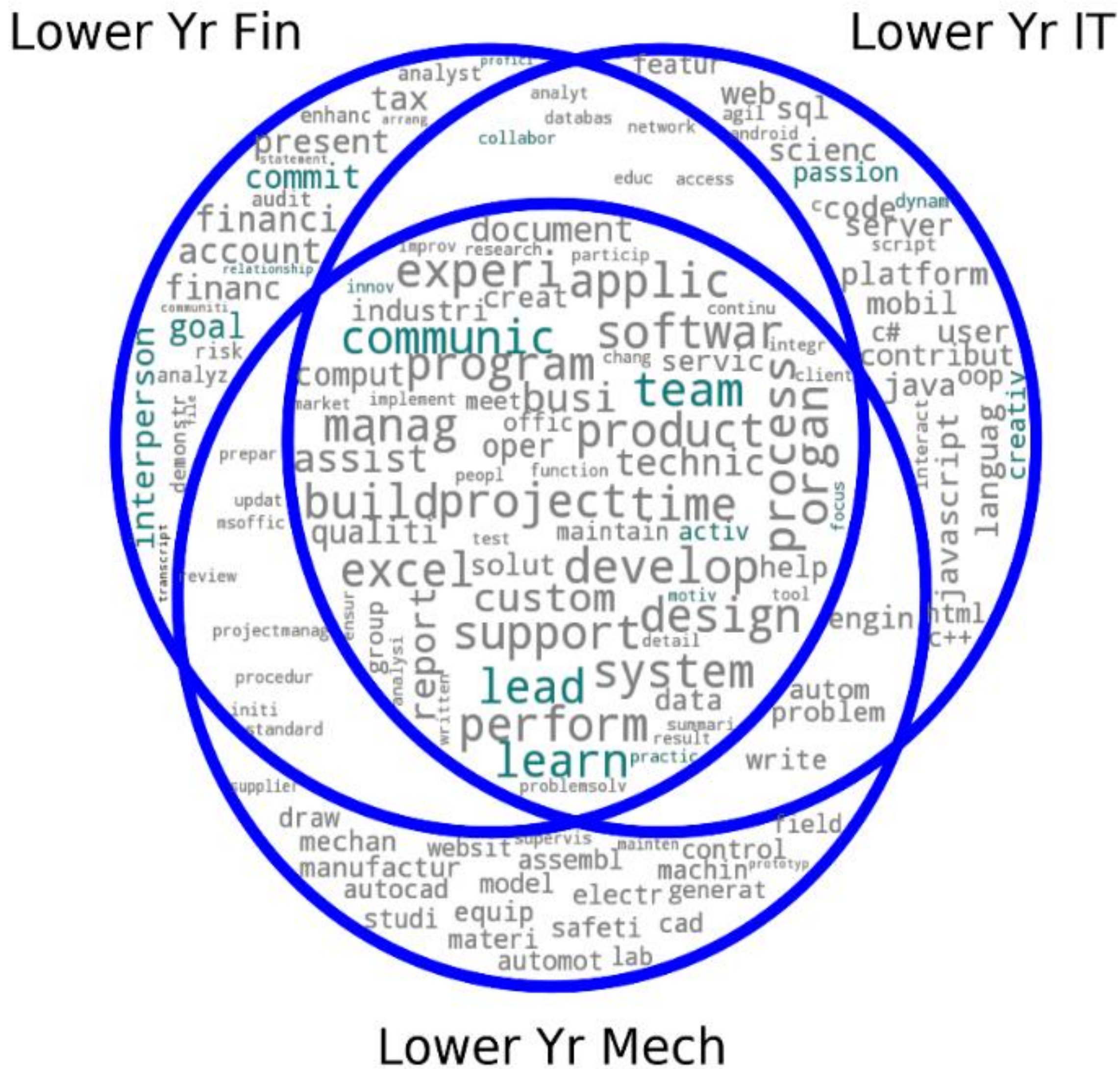

Figure 4.39 Overlap between the top 100 most frequent attributes of the Lower year jobs of Fin, IT and Mech

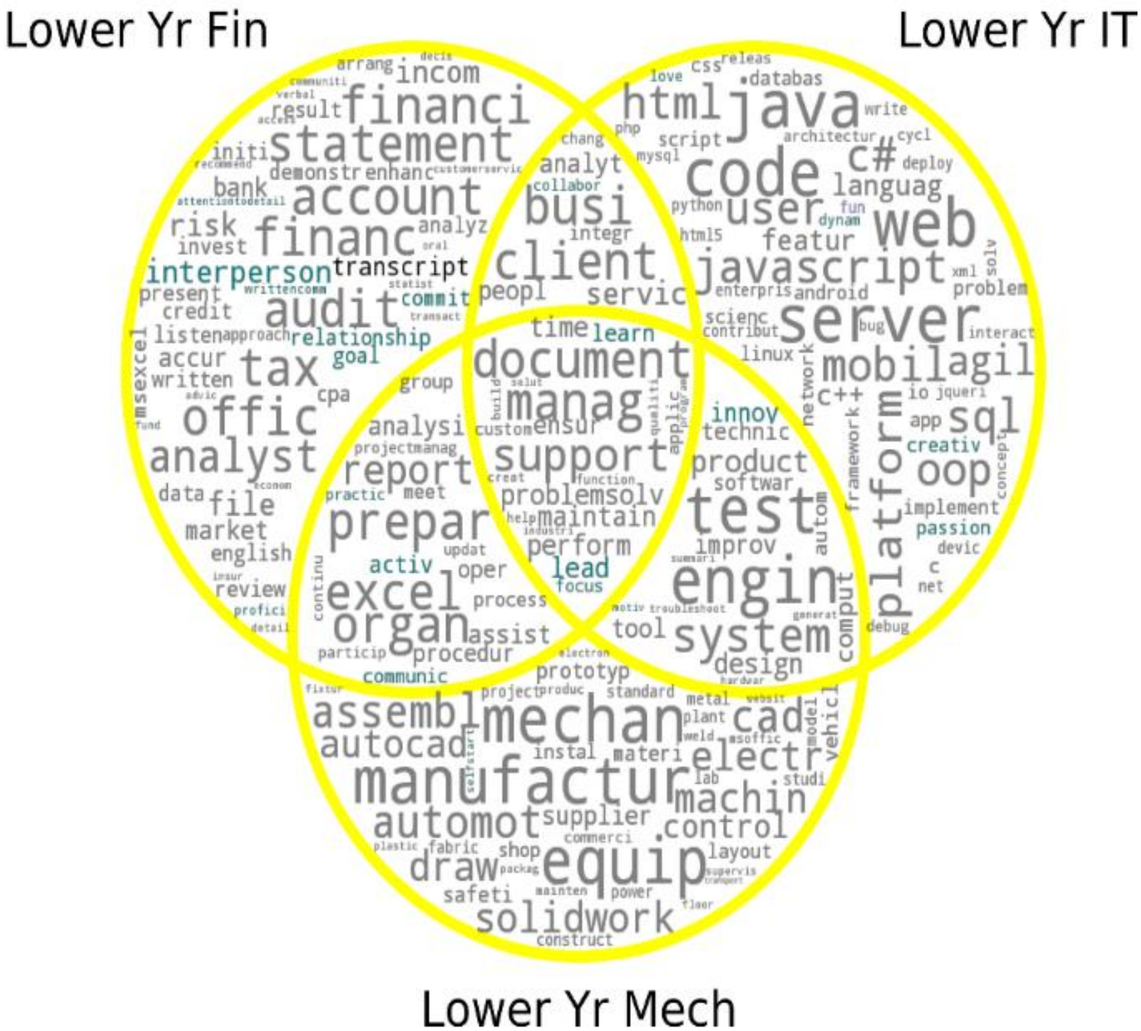

Figure 4.40 Overlap between the top 100 most representative attributes of the Lower year jobs of Fin, IT and Mech

As expected, there is less overlap in the top 100 representative skills than in the top 100 frequent skills. Zooming into Figure 4.40, the following can be inferred:

- Lower year jobs, irrespective of their discipline, are defined by “document”, “support”, “maintain” etc. showing that all lower year positions involve clerical work.
- “program” defines all lower year jobs suggesting a trend towards IT.
- “problem solving” represents all the lower year jobs suggesting the need of some analytical work.
- Fin and IT appear to offer a more “collaborative” environment.

Overall, this Section reveals that with the exception of some analytical skills specific to their discipline, lower year jobs of any discipline include clerical and IT related work.

4.3.4.2 Similarities between upper year students across the disciplines

Next, we examine the top 100 frequent attributes of upper year positions of all disciplines (Figure 4.41).

- Apart from generic attributes such as "experi", "project" and "product", attributes indicating more autonomy and application of knowledge appear in upper year jobs (e.g. "build", "create", "analysis", "ensure").
- "software" is still frequent in all upper year jobs indicating basic IT knowledge required by all disciplines irrespective of level.
- Similar to all lower year jobs, all upper year jobs include Soft Skills such as teamwork, communication, life-long learning and leadership. "dynamic" is mentioned in all upper year jobs but was not mentioned in lower year jobs (Figure 4.39).
- Many core attributes of each discipline are more frequent among each discipline's upper year jobs (Figure 4.41) than in their lower year jobs (Figure 4.39).

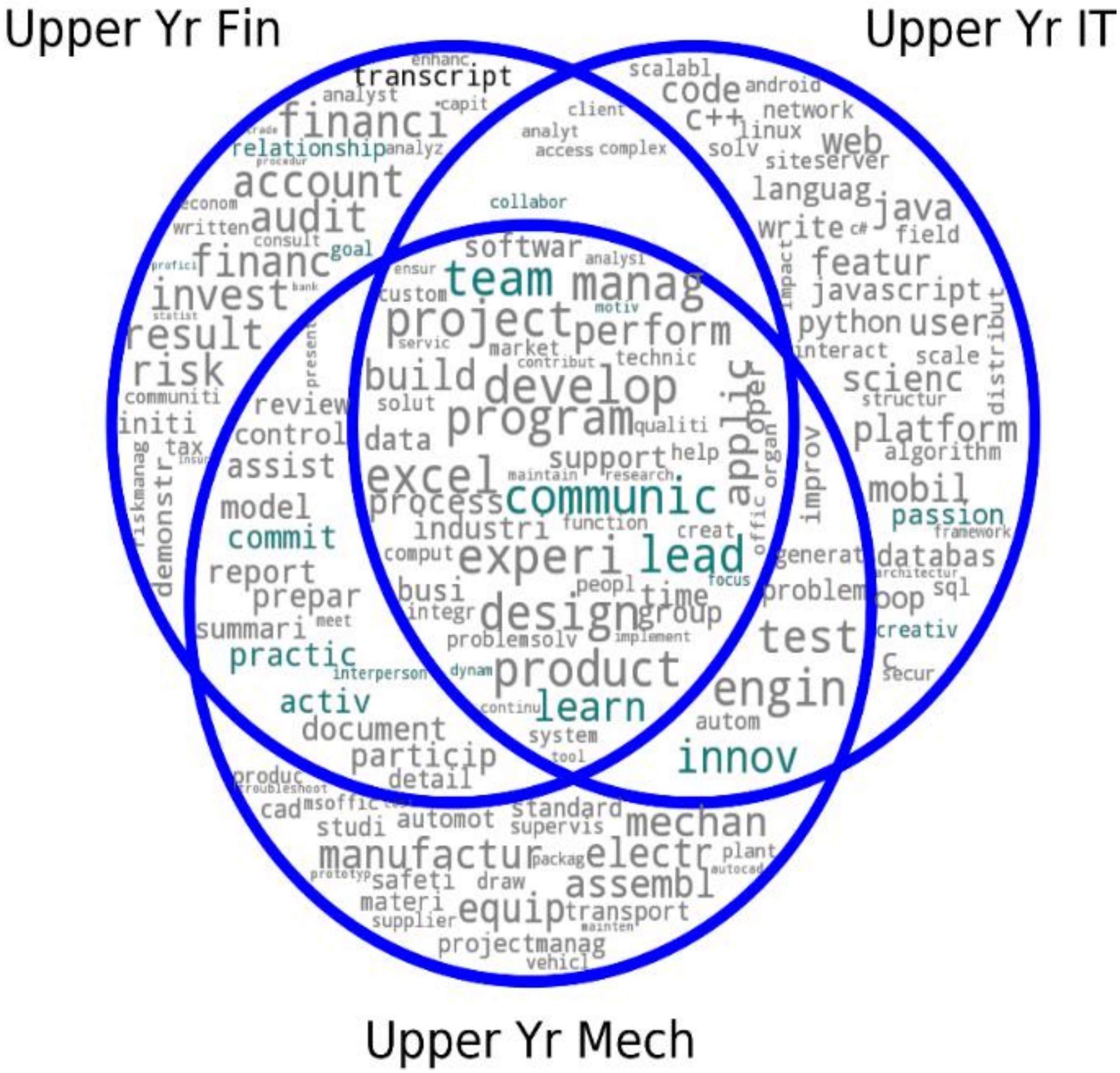

Figure 4.41 Overlap between the top 100 most frequent attributes of the Upper year jobs of Fin, IT and Mech

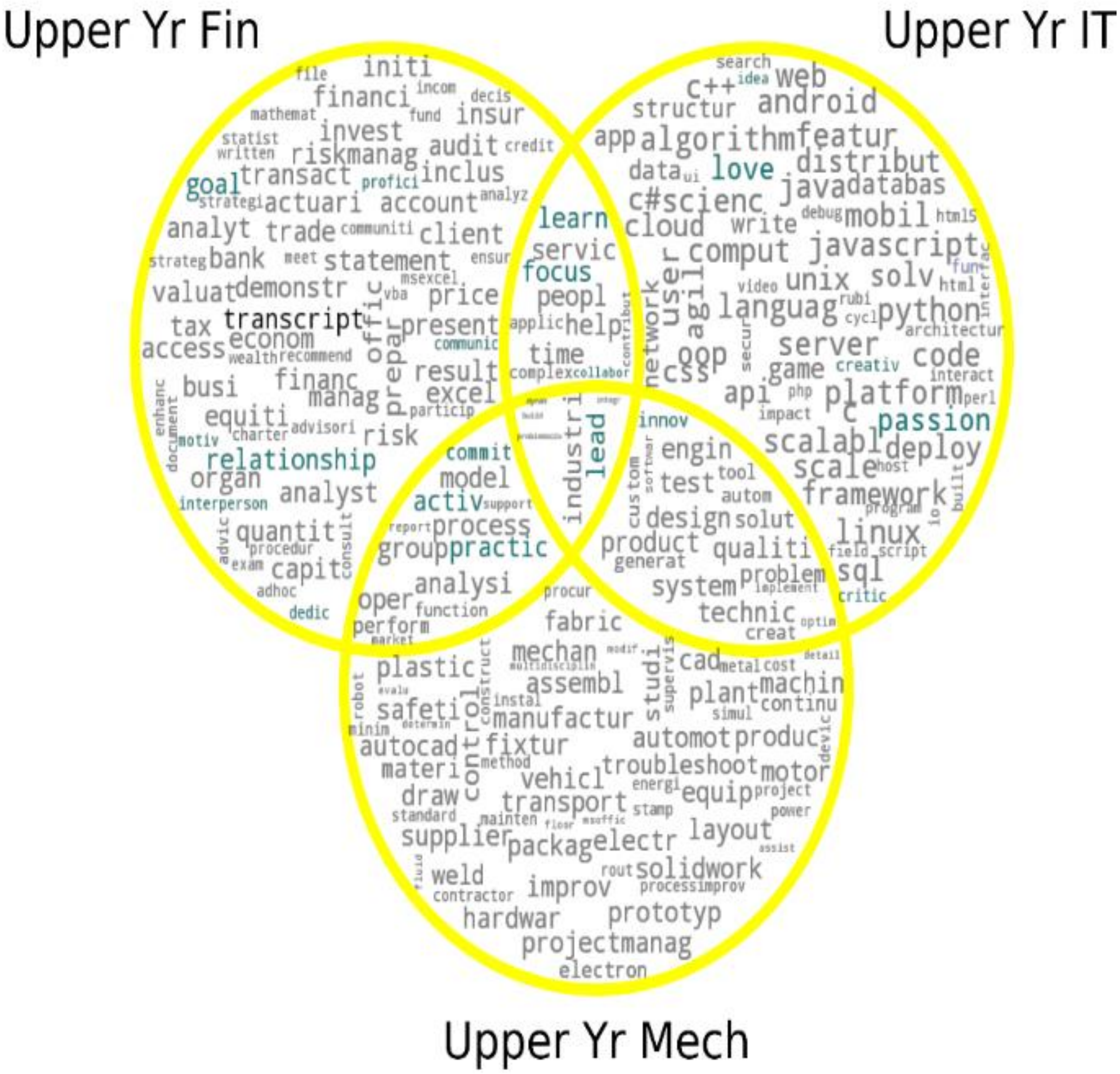

Figure 4.42 Overlap between the top 100 most representative attributes of the Upper year jobs of Fin, IT and Mech

As the top 100 representative attributes of the upper year jobs of each discipline represent the advanced skills of each discipline, they have less overlap, as shown in Figure 4.42. We draw the following conclusions.

- An Upper Year job profile of any discipline contains leadership roles.
- Upper Year jobs offer "dynamic" environments.
- "building" and "problem solving" represent all upper year jobs.

Overall, this Section reveals that irrespective of discipline, upper year jobs appear to seek dynamic individuals to fill leadership roles and use advanced concepts to build new things.

Chapter 5

Conclusions and Future Work

In this thesis, we presented a text-mining study of a co-op market at a large post-secondary institution. Using a large dataset of job postings, we developed a methodology to extract and compare the main attributes of jobs filled by students from various disciplines and with different seniority levels. Our main findings are as follows.

- As expected in an undergraduate co-op marketplace, there are many "assistant" and "junior" positions.
- Regardless of discipline, soft skills (teamwork, communication) were frequently mentioned in job postings, with IT postings additionally mentioning mindset (passion and love for the work) and Fin emphasizing interpersonal relationships.
- Non-IT fields such as finance and mechanical engineering appear to be trending towards IT and software, especially in their junior-level positions.
- Job postings from different disciplines suggest different working environments: labs and manufacturing plants in Mech, office environments in Fin, and casual, fun and collaborative environments in IT. In particular, "teamwork" appeared most frequently in IT postings, followed by Finance and Mechanical.
- Regardless of discipline, lower-year positions are more clerical while upper year positions tend to mention advanced concepts and solution development.

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

a) Diversity in size and age of companies, e.g., IT has many modern companies that emphasize a fun work culture while Fin has more traditional companies which might emphasize relationships.

b) 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.

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

- We can provide students with a better understanding of the co-op opportunities in various disciplines and therefore help them select the right academic program.
- 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, and help retain, junior students. As we showed, it may take until senior years to obtain a co-op position that fully utilizes discipline-specific skills.
- The institution may use frequently appearing job attributes in various disciplines to produce more

effective promotional material 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 the disciplines seem to emphasize teamwork skills, the institution can incorporate more team exercises in their course curriculum that can help students hone this skill. Hackathons and other competitions could be made part of the curriculum to foster passion and other mindset related skills in IT students while mock client meetings could be arranged for Fin students to give them a chance to hone their interpersonal skills.
- Employers may examine our findings to understand which skills are in high demand and therefore to understand the extent of competition in the co-op market.
- Our lists of frequently appearing and representative attributes may be used to re-design the way employers submit job postings. For instance, a separate field (outside the job description) may be added for required skills, with a drop-down list populated with frequent and representative skills. Similarly, dropdown lists for popular administrative requirements, perks, company culture, salary etc. could guide employers to express their needs more appropriately. Collecting structured job descriptions and resumes could help students as well as employers to find appropriate matches efficiently.

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 help minimize the gap between employers' needs and students' talents. In this thesis, we focused on job descriptions, which provide an indication of what co-op employers are looking for. In future work, we will characterize what students have to offer by mining resumes and what students are good at by analyzing work term evaluations. We also plan to design a recommender system that will identify suitable students for a given job posting.

References

[1] Abbreviations and acronyms. Accessed on 25 Aug 2016, at https://www.gcflearnfree.org/grammar/abbreviations-and-acronyms/1/

[2] Academic Degrees. Accessed on 25 Aug 2016, at https://en.wiktionary.org/wiki/Appendix:Academic_degrees

[3] Accessing Text Corpora and Lexical Resources. Accessed on 30 Jun 2017, at http://www.nltk.org/book/ch02.html

[4] ASCII Control Character. Accessed on 30 Jun 2017, at https://web.archive.org/web/20091028135111/http://geocities.com/dtmcbride/tech/charsets/ascii.html

[5] Barber, A. E. (1998). *Recruiting employees: Individual and organizational perspectives* (Vol. 8). Sage Publications.

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

[7] Bauer, L. (2003). Introducing linguistic morphology.

[8] Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). paper recommender systems: a literature survey. *International Journal on Digital Libraries*, *17*(4), 305-338.

[9] Belt, J. A., & Paolillo, J. G. (1982). The influence of corporate image and specificity of candidate qualifications on response to recruitment advertisement. *Journal of Management*, *8*(1), 105-112.

[10] Blair, B. F., Millea, M., & Hammer, J. (2004). The impact of cooperative education on academic performance and compensation of engineering majors. *Journal of Engineering Education*, *93*(4), 333-338.

[11] Breaugh, J. A. (1992). Recruitment: Science and practice. Boston: PWS-Kent. Breaugh, JA 2008. Employee recruitment: Current knowledge and important areas for future research. *Human Resource Management Review*, *18*, 103-118.

[12] Breaugh, J. A., & Starke, M. (2000). Research on employee recruitment: So many studies, so many remaining questions. *Journal of management*, *26*(3), 405-434.

[13] Cable, D. M., & Turban, D. B. (2003). The value of organizational reputation in the recruitment context: A brand-equity perspective. *Journal of Applied Social Psychology*, *33*(11), 2244-2266.

[14] Canadian Association For Co-Operative Education CAFCE. Accessed on August 20 2017. http://www.cafce.ca/accreditation.html

[15] Cerf, V. G. (1969). *ASCII format for network interchange* (No. RFC 20).

[16] Chatman, J. A. (1991). Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly*, 36(3), 459-484.

[17] Co-operative Education & Career Services Employment Process Review Summary Report. Accessed on 30 Jun 2017. https://uwaterloo.ca/co-operative-education/sites/ca.co-operative-education/files/uploads/files/process_review.pdf

[18] Cober, R. T., Brown, D. J., Levy, P. E., Cober, A. B., & Keeping, L. M. (2003). Organizational web sites: Web site content and style as determinants of organizational attraction. *International Journal of Selection and Assessment*, *11*(2-3), 158-169.

[19] Coll, R. K., & Chapman, R. (2000). Evaluating service quality for cooperative education programs. *Asia-Pacific Journal of Cooperative Education*, *1*(2), 1-12.

[20] Coll, R. K., & Zegwaard, K. E. (2006). Perceptions of desirable graduate competencies for science and technology new graduates. *Research in Science & Technological Education*, *24*(1), 29-58.

[21] Coll, R. K., Zegwaard, K., & Hodges, D. (2002). Science and technology stakeholders' ranking of graduate competencies Part 1: Employer perspective. *Asia-Pacific Journal of cooperative education*, *3*(2), 19-28.

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

[23] Collins, C. J. (2007). The interactive effects of recruitment practices and product awareness on job seekers' employer knowledge and application behaviors. *Journal of applied psychology*, *92*(1), 180.

[24] Collins, C. J., & Han, J. (2004). Exploring applicant pool quantity and quality: The effects of early recruitment practice strategies, corporate advertising, and firm reputation. *Personnel Psychology*, *57*(3), 685-717.

[25] Common Abbreviations. Accessed on 25 Aug 2016, at https://media.gcflearnfree.org/ctassets/modules/48/common_abbr.png

[26] Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search engines: Information retrieval in practice* (Vol. 283). Reading: Addison-Wesley.

[27] Diaby, M., Viennet, E., & Launay, T. (2013, August). 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* (pp. 821-828). ACM.

[28] Eames, C., & Coll, R. K. (2010). Cooperative education: Integrating classroom and workplace learning. *Learning through practice*, 180-196.

[29] Examples of the best skills to include on Resumes. Accessed on 25 Aug 2016, at https://www.thebalance.com/list-of-the-best-skills-for-resumes-2062422

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

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

[32] Gatewood, R. D., Gowan, M. A., & Lautenschlager, G. J. (1993). Corporate image, recruitment image and initial job choice decisions. *Academy of Management journal*, *36*(2), 414-427.

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

[34] Harris, K. S., & Rogers, G. E. (2008). Soft skills in the technology education classroom: What do students need?. *The technology teacher*, *68*(3), 19-25.

[35] Hesketh, A. J. (2000). Recruiting an elite? Employers' perceptions of graduate education and training. *Journal of education and work*, *13*(3), 245-271.

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

[37] HTML element reference. Accessed on 30 Jun 2017, at https://developer.mozilla.org/en/docs/Web/HTML/Element

[38] Jiang, Y. H., & Golab, L. (2016). On Competition for Undergraduate Co-op Placements: A Graph Mining Approach. In *EDM* (pp. 394-399).

[39] Jiang, Y. H., Lee, S. W. Y., & Golab, L. (2015). Analyzing Student and Employer Satisfaction with Cooperative Education through Multiple Data Sources. *Asia-Pacific Journal of Cooperative Education*, *16*(4), 225-240.

[40] Job Code Title Abbreviation List. Accessed on 25 Aug 2016, at http://hr.iu.edu/SOS/Job_Code_Title_Abbreviation_List.htm

[41] JobMine, University of Waterloo Student Career Management Database. Accessed on 30 Jun 2017, at http://www.jobmine.uwaterloo.ca.

[42] Kim, S. S., & Gelfand, M. J. (2003). The influence of ethnic identity on perceptions of organizational recruitment. *Journal of Vocational Behavior*, *63*(3), 396-416.

[43] Leung, H. (2007). *The Impact of Information Quality of Job Descriptions on an Applicant's Decision to Pursue a Job* (Master's thesis, University of Waterloo).

[44] Li, M., & Lam, B. (2005). Cooperative learning. *The Active Classroom, The Hong Kong Institute of Education. Hong Kong.*

[45] Lievens, F., Decaesteker, C., Coetsier, P., & Geirnaert, J. (2001). Organizational attractiveness for prospective applicants: A person–organisation fit perspective. *Applied Psychology*, *50*(1), 30-51.

[46] List of Common misspellings. Accessed on 25 Aug 2016, at https://en.wikipedia.org/wiki/Wikipedia:Lists_of_common_misspellings/For_machines

[47] List of social networking websites. Accessed on 25 Aug 2016 at, https://en.wikipedia.org/wiki/List_of_social_networking_websites

[48] List of English contractions. Accessed on 25 Aug 2016, at https://en.wikipedia.org/wiki/Wikipedia:List_of_English_contractions

[49] Longman Communication 3000. Accessed on 25 Aug 2016, at http://www.lextutor.ca/freq/lists_download/longman_3000_list.pdf

[50] Malherbe, E., Diaby, M., Cataldi, M., Viennet, E., & Aufaure, M. A. (2014, August). Field selection for job categorization and recommendation to social network users. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on* (pp. 588-595). IEEE.

[51] McCurdy, S. U. S. A. N., & Zegwaard, K. E. (2009). Faculty voices: What faculty think about work-integrated learning. *Journal of Cooperative Education and Internships*, *43*(1), 36-53.

[52] Monster. Accessed on 30 Jun 2017, at http://www.monster.ca.

[53] Most Common Prefixes. Accessed on 25 Aug 2016, at http://teacher.scholastic.com/reading/bestpractices/vocabulary/pdf/prefixes_suffixes.pdf

[54] NOC Occupational Classification 2016. Accessed on 25th Aug 2016, at http://noc.esdc.gc.ca/English/noc/welcome.aspx?ver=16

[55] Ontario Ministry of Education. Earning more credits through workplace experience. Accessed on 30 Jun 2017, at https://www.edu.gov.on.ca/morestudentsuccess/coop.html.

[56] Porter, M. F. (2001). Snowball: A language for stemming algorithms. October 2001. *Retrieved March*, *1*, 2014.

[57] Raelin, J. A., Bailey, M. B., Hamann, J., Pendleton, L. K., Raelin, J., Reisberg, R., & Whitman, D. (2011). The effect of cooperative education on change in self-efficacy among undergraduate students: Introducing work self-efficacy.

[58] Rainsbury, E., Hodges, D. L., Burchell, N., & Lay, M. C. (2002). Ranking workplace competencies: Student and graduate perceptions.

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

[60] Roberson, Q. M., Collins, C. J., & Oreg, S. (2005). The effects of recruitment message specificity on applicant attraction to organizations. *Journal of Business and Psychology*, *19*(3), 319-339.

[61] Rynes, S. L., & Barber, A. E. (1990). Applicant attraction strategies: An organizational perspective. *Academy of management review*, *15*(2), 286-310.

[62] Slaughter, J. E., & Highhouse, S. (2003). Does matching up features mess up job choice? Boundary conditions on attribute-salience effects. *Journal of behavioral decision making*, *16*(1), 1-15.

[63] Smith, B. N., Benson, P. G., & Hornsby, J. S. (1990). The effects of job description content on job evaluation judgments. *Journal of applied psychology*, *75*(3), 301.

[64] Stengel, D., Young, D. R., Chaffe-Stengel, P., & Harper, R. M. (2010). Assessing the academic and workplace skills of undergraduate business interns. *Journal of Cooperative Education & Internships*, *44*(18), 13-22.

[65] Stéphane, L. E., & Seguela, J. (2012). *U.S. Patent Application No. 13/543,616*.

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

[67] Thompson, K. (1968). Programming techniques: Regular expression search algorithm. *Communications of the ACM*, *11*(6), 419-422.

[68] Toulis, A., & Golab, L. (2017, May). Graph Mining to Characterize Competition for Employment. In *Proceedings of the 2nd International Workshop on Network Data Analytics* (p. 3). ACM.

[69] University of Waterloo. Communications and Public Affairs. University of Waterloo. Accessed on 30 Jun 2017, at http://www.uwaterloo.ca.

[70] Workopolis.com. Accessed on 30 Jun 2017, at http://www.workopolis.com.

[71] World Association for Cooperative & Work-integrated Education. Wace global institutions. Accessed on 30 Jun 2017, at http://www.waceinc.org/global_ institutions.html

[72] Yüce, P., & Highhouse, S. (1998). Effects of attribute set size and pay ambiguity on reactions to 'Help wanted' advertisements. *Journal of Organizational Behavior*, 337-352.

[73] Zakaria, E. (2009). Promoting cooperative learning in science and mathematics education: A Malaysian perspective. *Coleccio□ n Digital Eudoxus*, (22).

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

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

[76] Zegwaard, K., Coll, R. K., & Hodges, D. (2003). Assessment of workplace learning: A framework. *Asia-Pacific Journal of Cooperative Education*, *4*(1), 9-18.

[77] 21 Employee Perks That Attract The Best Talent, Forbes Magazine. Accessed on 25 Aug 2016 at, www.forbes.com/sites/brockblake/2015/02/24/21-employee-perks-that-attract-the-best-talent/

[78] 15 of the Most Popular Employee Perks, Glassdoor Blog. Accessed on 10 Jan 2017 at, https://www.glassdoor.com/blog/15-of-the-most-popular-employee-perks/

[79] 40 Business Acronyms and Business Abbreviations You Should Know. Accessed on 25 Aug 2016, at https://www.altalang.com/beyond-words/2014/02/17/40-business-acronyms-know/

[80] 2500 Chat Acronymns and Text Shorthand. Accessed on 25 Aug 2016, at http://www.netlingo.com/acronyms.php