Bridging the Skills Gap between Post-Secondary Education Outcomes and Employment Opportunities
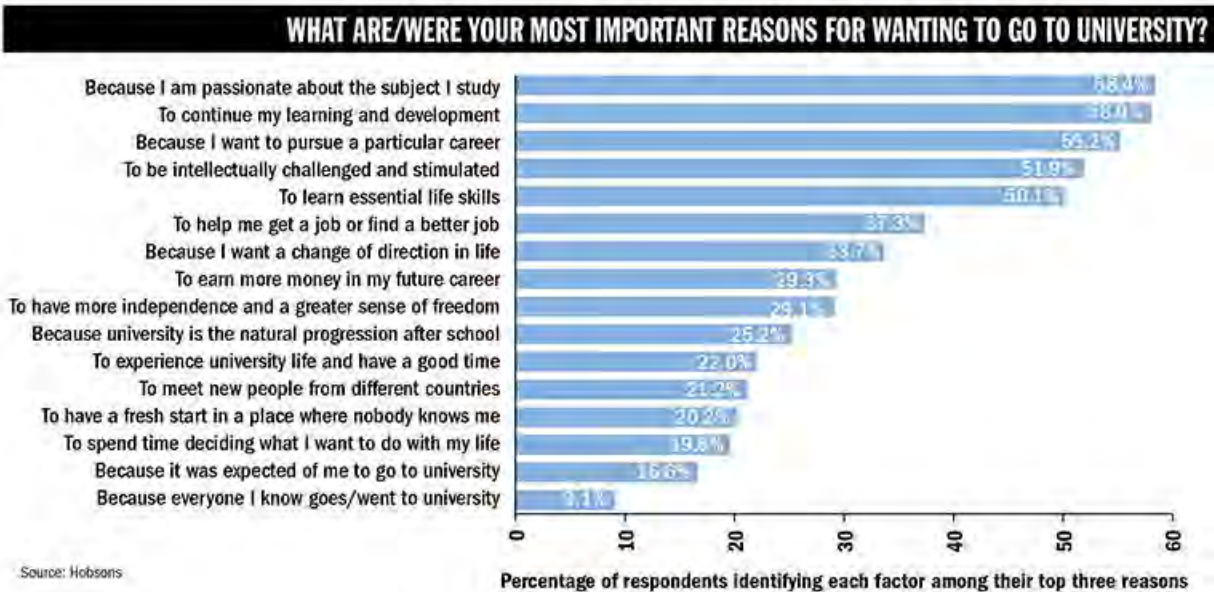
Hong-Yu Xavier Fu, Bhagmattie (Annissa) Rodriguez-Ramdhanie, George Brown College

ABSTRACT

Students are often expected to take a leap of faith in their pursuit of Post-Secondary Education. Their decisions to develop the necessary skillsets that they will need to be successful are often fraught with uncertainty and misinformation. For Post-Secondary Education Institutions, the monitoring of employment qualifications is critical to ensure that post-secondary education supplies graduates with the essential skills required in the industries. Prediction of the changes in industry qualifications is an increasingly difficult task. Colleges often seek the advice of a combination of aggregated labor market data as well as the advisory support of Program Advisory Committees comprised of industry experts and practitioners. The skill classification codes of aggregate labor market data are often too general, and have a tendency to miss emerging skill demands as they enter into the marketplace. In addition, Program Advisory Committees that provide input to College and University program development committees are often subjected to their own biases, or the biases of the faculty members that recruited them into the advisory role. This paper proposes an analytic approach to help better deal with this situation through the development of an ongoing monitoring system that can associate program outcomes to job requirements. This type of approach can be done on a sector specific basis, as well as region specific basis to suit the interests and objectives of the College or University. Through the application of SAS Text Miner, this paper looks at processes to crawl targeted websites for entry and mid-level job postings, and to classify them into topic themes and highlight the underlying skills required. Using a stratified sample of job postings, the development of an automated text profiler for ongoing performance monitoring is explored. With the usage of text cluster, skillsets can be categorized to better match program outcomes, and a visualization of the relationship between post-secondary program outcomes and underlying sought after employment skills can be created. The data can then be used by post-secondary educators to better bridge the skills gap.

INTRODUCTION

According to the Bureau of Labor Statistics, "Of the 3.1 million youth age 16 to 24 who graduated from high school between January and October 2016, about 2.2 million (69.7 percent) were enrolled in college in October." The number continually increase by the rising number of students pursuing further education, a change in career, a new skillset, or increase salary (Bureau of Labor Statistics, 2017). Figure 1 illustrates the reasons students pursue post-secondary education.

WHAT ARE/WERE YOUR MOST IMPORTANT REASONS FOR WANTING TO GO TO UNIVERSITY?

Reason	Percentage
Because I am passionate about the subject I study	58.4%
To continue my learning and development	58.0%
Because I want to pursue a particular career	55.2%
To be intellectually challenged and stimulated	51.9%
To learn essential life skills	50.1%
To help me get a job or find a better job	37.3%
Because I want a change of direction in life	33.7%
To earn more money in my future career	29.3%
To have more independence and a greater sense of freedom	29.1%
Because university is the natural progression after school	25.2%
To experience university life and have a good time	22.0%
To meet new people from different countries	21.2%
To have a fresh start in a place where nobody knows me	20.2%
To spend time deciding what I want to do with my life	19.8%
Because it was expected of me to go to university	16.6%
Because everyone I know goes/went to university	9.1%

Source: Hobsons

Percentage of respondents identifying each factor among their top three reasons

Figure 1: Why do student Pursue Post-Secondary Education (Bhardwa, 2017)

A student's passion for a subject area, combined with their strong motivation to continue their learning and development are key drivers for pursuing post-secondary education (Bhardwa, 2017). One of the primary outcomes of that passion is for the institution to direct their energies into developing skillsets that will set them up for success in their future career aspirations.

Post-secondary education is continually advertised in society as a gateway to success defined by the basis that upon graduation, meaningful skills are learned. The success is further realized through the graduate embarking on a meaningful career in a job market that values their acquired skills. Programs are constantly challenged to develop comprehensive monitoring systems that can validate the claims that they make. They often rely on selective anecdotal evidence from key alumni.

A potential student often decides if post-secondary education is the correct step in their life's path through evaluating the skillsets learned against the skillsets needed in their desired career. This hectic procedure translates into exploring websites similar but not limited to Indeed, Monster, or Workopolis to analyze data from job postings. The seeker then concludes if the skillsets required for employment are already possessed, and if not, makes a plan to acquire the suggested skillsets.

However, for a prospective student, simply being aware of the skills is not sufficient to be successful in the job market. Students must also be able to evaluate the capabilities of the programs that they select to apply to in order to support their goals. To gain the skills required by the job market, the seeker then must evaluate if post-secondary education will lead to learning the missing skills. This can be done through reading the program outcomes listed on university or college's websites. In doing this, the future student hopes that the program outcomes matches the skillsets needed for employment. This is usually when the student makes the leap of faith.

To make this leap of faith more manageable, institutions make it known though advertisement that programs are specifically designed and tailored to meet the needs of the recruiters in industries. This is done through focus groups between industry leaders, and program developers. The process, of which the outcomes are derived, includes the two parties mapping desired employable skills and creating a course outcome that teaches these employable skills.

The created course outcome is believed to be successful because both parties have an understanding and a part in the design of the success criteria. Although periodic program reviews are required to ensure the currency of program outcomes to skills that are demanded in the marketplace, there's still a significant amount of discretion and judgement that is placed in the hands of a select panel of industry and education experts. In the end, every side makes a leap of faith through the trust that the course designed will meet the desires of every party.

The underlying goal for a student is to learn the skills needed, while the underlying goal for the post-secondary education is for the course to be able to provide those skillsets needed by the industry. Regardless, without a large enough sample, it cannot be confirmed if the skills taught are the skills needed by the majority of the industry and future skillsets cannot be predicted.

Our program is built to bridge this gap so that both students and post-secondary education can achieve the common goal.

BUILDING THE DATASET

The dataset gathered for the text mining program contains job data from 340 different jobs encompassing business analysts, accountants, financial advisors, medical receptionist, fashion management, management, and culinary positions. Different jobs were gathered through Boolean searches on Indeed.ca and the qualifications and requirement were gathered manually and imported into SAS Enterprise Miner.

Similarly, for the program outcomes dataset, George Brown College's resources were used and the mandatory course syllabuses for B412 (Business Analytics), F112 (Fashion Management), and H112 (Chef Training) were gathered. The course outcomes were then inputted manually and imported into SAS Enterprise Miner.

PROGRAM STRUCTURE

The datasets gathered contain raw unfiltered text. In order to filter and correlate the two datasets, the program needs to do the following:

1. Parse the raw text data to contain only nouns, verbs, and abbreviations.
2. Filter the text data for important words by frequencies and weight.
3. Using the words remaining, group important words into clusters.
4. Score the clusters to the dataset through numerical meaning.

The program achieved these criteria through the usage of the Text Parsing Node, Text Filter Node, Text Cluster Node, and Score Node, as shown in Figure 2. Each Node will be further analyzed.
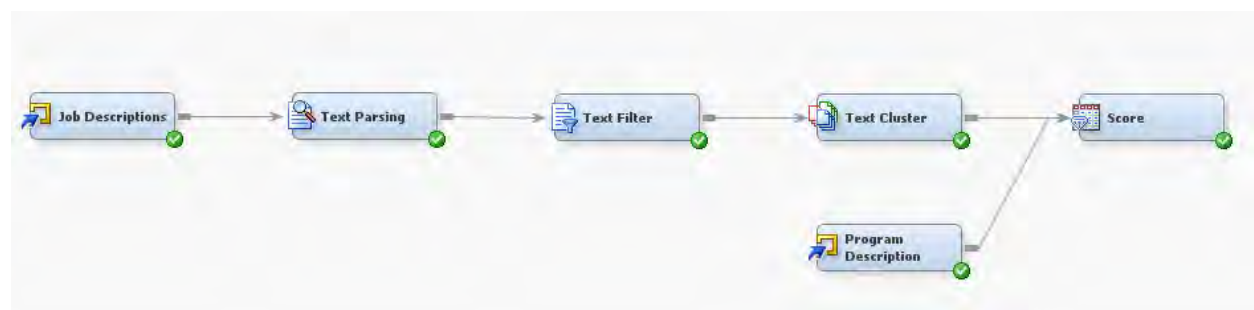
Figure 2: Node structure of the program.

TEXT PARSING NODE

After the import of the dataset through the Import Node, a Text Parsing node was connected. The node offers the ability to ignore certain words and punctuations which was much needed to structure the raw data. The settings were left to default and the Job Description data was parsed to only include nouns, verbs, and abbreviations.

Only the Job Description data was used in the bulk of the program for the purpose of the Score Node.

TEXT FILTER

Once parsing was completed, a Text Filter Node was used to filter the synonyms in the dataset. The Text Filter Node was switched to Log frequency weighing and Inverse Document Frequency term weighing to allow for the document to ignore certain key words that occur frequently (such as Team Player). To further the analysis, the filter viewer was used to classify certain words into synonyms.

TEXT CLUSTER

After text filtering, the output data was then connected into the Text Cluster Node. The properties of the node were set to 5 terms per cluster for a maximum of 40 clusters. This was to allow the skillsets to not contain overwhelming amounts of keywords, making it easier to identify the proposed skill. Furthermore, having a maximum of 40 clusters allow SAS Enterprise Miner to determine the important skills. The validity of the skillset is heavily dependent on the filter node, which contains the most human input.

SCORE

The final node of the program contains the score node. To make sense of the data and clusters, the score node needs to numerically relate the clusters from job descriptions to the program outcomes. By setting the program outcomes dataset to SCORE properties, and then connecting it into the score node with the cluster, the probability of the program outcome inside each cluster is found in Figure 3.

TextCluster_pr... ▼	TextCluster_prob2	TextCluster_prob3	TextCluster_prob4	TextCluster_prob5	TextCluster_prob6
0.999996	1.76E-18	1.845E-9	3.534E-6	3.68E-79	3.381E-9
0.999993	1.01E-16	1.01E-14	2.61E-20	1.71E-54	6.887E-6
0.992882	4.87E-20	8.827E-6	2.2E-20	3.79E-22	0.007109
0.937857	.0000012	3.167E-5	4.496E-5	1.66E-12	0.062065
0.906314	6.14E-12	0.002421	8.31E-10	3.3E-86	0.091265
0.749191	6.06E-24	.0000043	0.250805	2.22E-7	1.61E-12
0.680479	1.19E-15	8.31E-13	7.63E-34	1.2E-117	0.319521
0.437514	6.94E-22	0.555472	6.1E-16	1.75E-10	0.007013
0.144284	7.03E-20	0.855448	1.157E-9	1.14E-72	0.000268
0.088792	1.36E-7	0.910586	.0005313	1.2E-224	9.036E-5
0.054147	4.199E-9	0.012422	6.84E-12	0.933354	7.755E-5

Figure 3: Part of the probability output of the score node.

As seen in Figure 3, each row contains a course from a program, each column contains the probability of that row being related to a founded cluster/skillset. In other words, the score node relates the chance of learning the employable skillset text mined from job postings in the course offered by a specific program.

VISUALIZATION

To see the results in a simplified manner, visualizations were created by importing the score output data into Tableau. Looking specifically at three job skills for culinary positions (Team Supervision, Preparation, and Customer Value) in the example provided in Figure 4, it can be seen that for the Program B412 (Business Analytics Program), the course outcomes do not teach much skills needed for team supervision or preparation. Whereas for Customer Value, there appears to be more correlation.
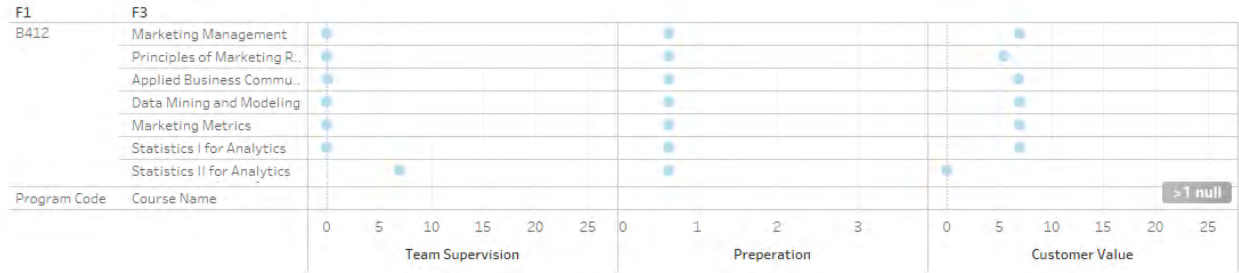
Figure 4: Example of Tableau visualization of Chef Job Skills to B412.

When comparing the same three job skills against F112 (Fashion Management Program) in Figure 5, there was an increase in skills hits based on the course outlines, suggesting more correlation.
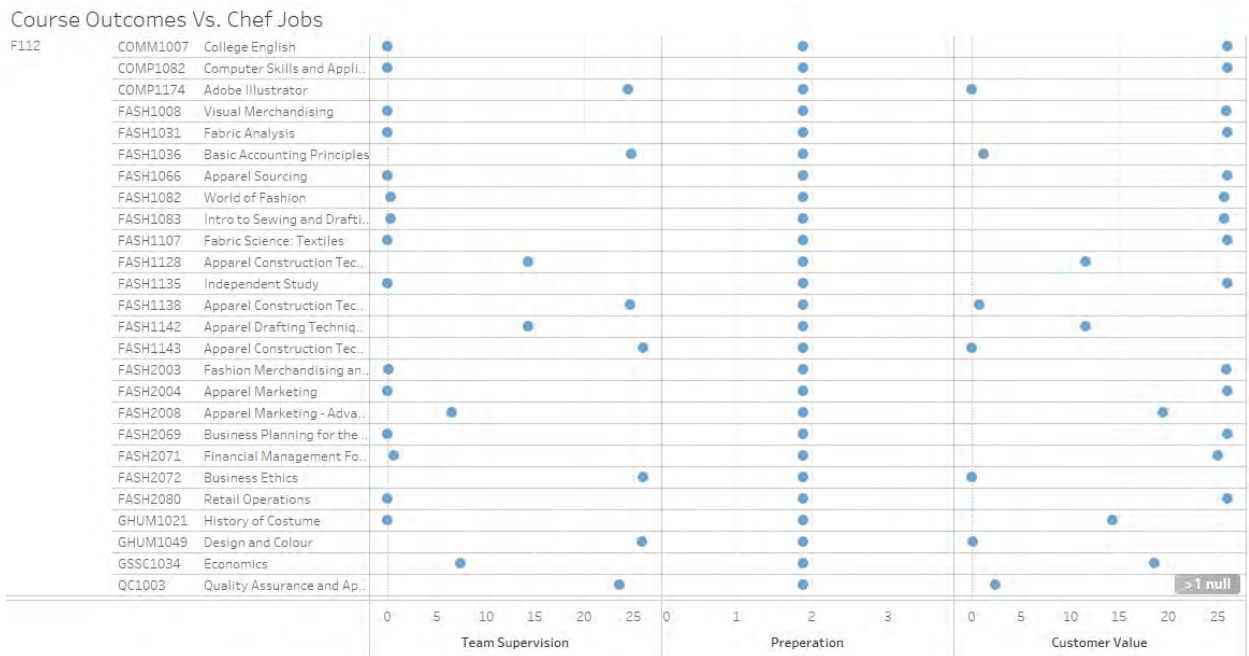
Figure 5: Chef Job Skills to F112

Evaluating the same three job skills against H112 (Chef Training Program) in Figure 6 shows the largest fluctuations amongst the three courses, and had the highest amount of hits against the three employable skills.
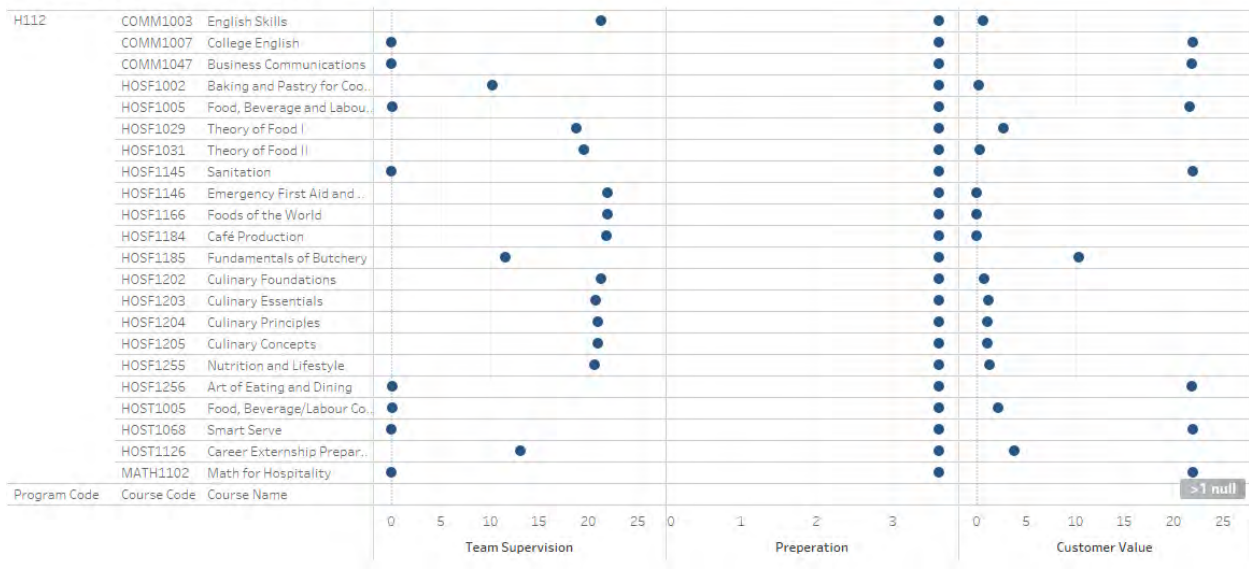
Course Outcomes Vs. Chef Jobs

Program Code	Course Code	Course Name	Team Supervision	Preperation	Customer Value

Figure 6: Chef Jobs skills vs H112

CONCLUSION

The objective of our program was to improve the post-secondary program outlines to better cater to the current job market. In order to achieve this objective, the relationship was studied. Compiling the raw data into structured data and running through the program and visualization programs, the steps to the traditional methods can be simplified for more meaningful insights. Furthermore, this analysis can be used to predict the job market when using larger samples.

From the analysis of the score node, which shows the probabilities of employable skills of a specific job within the program outlines, it can be found that the employable skills can be related or unrelated to the courses based on the jobs. For example, during our analysis of the culinary positions employability data, it showed a massive relationship to the Chef Training program, some relationship to the Fashion Management program, and little relationship to the Business Analytics program. It is important to note that these finding illustrates the program's ability to show relationship down to the course level, based on the original design of the course. The most important criteria of the program is that it translated words into numbers that can be easily visualized through BI tools such as Tableau.

Ideally the process can be further improved by mass scraping job engines, and targeting specific jobs to specific programs. These results can tell institutions how to design their programs to meet marketplace and students' needs. It can also help post-secondary education provide efficient solutions to meet market demand for more skilled workers, ultimately bridging the gap between post-secondary outcomes and employment opportunities.

REFERENCES

Bhardwa, Seeta. "Why do students go to university and how do they choose which one?" June 7, 2017. https://www.timeshighereducation.com/student/news/why-do-students-go-university-and-how-do-they-choose-which-one

Bureau of Labor Statistics. "College Enrollment and Work Activity of 2016 High School Graduates" November 17, 2017. https://www.bls.gov/news.release/hsgec.nr0.htm

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Xavier Fu
fu.xavier@gmail.com

Annissa Rodriguez-Ramdhanie
annissa@canada-storage.com