ISSN: 2473-3857
v5 n4925

Developing an Electronic Resume Analyzer Portal (e-RAP): A Natural Language Processing Approach to Enhance College Graduates Job Readiness

Nathan Green
ngreen@marymount.edu

Michelle (Xiang) Liu
xliu@marymount.edu

Diane Murphy
dmurphy@marymount.edu

Department of Information Technology,
Data Science, and Cybersecurity
Marymount University
Arlington, Virginia 22207, USA

Abstract

Finding the first full-time, major-related job is a challenge faced by most college students, particularly those who have not gained much working experience before entering the job market. This challenge is amplified for the students majoring in Information Technology (IT), and cybersecurity in particular, due to the constantly changing technology landscape, intensively competitive markets, and increasingly high expectations from employers on their recruits. This study shows the initial results of a tool called e-RAP which allows the students to submit their current resumes, obtain automatic feedback and a rating report, and consequently take actions to strengthen their portfolio. The authors employ machine learning and natural language processing (NLP) to create a resume analysis and reporting tool. The methodology section provides an overview of the e-RAP analysis process, followed by elaborations on data curation, data collection, and analysis techniques. Several visual examples of the reports generated by e-RAP illustrate the effectiveness of the tool in helping enhance students' resumes and eventually the skills areas they need to work on or highlight. The future direction includes systematic evaluation of the effectiveness of e-RAP and its impact on our student's ability to get high-quality positions. Diving deeper into the various types of cybersecurity positions is also planned.

Keywords: Resume Analysis, Natural Language Processing (NLP), Workforce Readiness, Technology Job Requirements, Major-Related Jobs, Machine Learning

1. INTRODUCTION

Getting that first job, even in a high-demand field such as cybersecurity, can be stressful for students. The 2018 McGraw-Hill Future Workforce Survey disclosed that only four in ten college students feel well-prepared for their careers (McGraw-Hill Education, 2018). According to the report, more than fifty percent of the students surveyed felt that they had not gained the substantial critical skills needed to transition to the workforce. For example, 57% of the students reported feeling a lack of problem-solving skills and 69% a lack of job searching skills. This situation is amplified for the students majoring in Information Technology (IT), and cybersecurity in particular, due to the constantly changing technology landscape, intensively competitive markets, and increasingly high expectations from employers on their recruits: a broad range of specific technical skills, business domain expertise, and highly-refined soft skills. The latest cybersecurity workforce study shows that the global shortage of cybersecurity professionals reaches a worrisome number at 2.93 million ((ISC)², 2018, October 17). Sixty-three percent of organizations surveyed reported that their organizations have a shortage of dedicated cybersecurity workers and 36% of respondents rank the skills gap as their top job concern.

To close the skills gap, different stakeholders including business leaders, government agencies, and policymakers have proposed and experimented with a variety of workforce readiness initiatives and programs. For instance, some companies have developed their own talent pipeline program as the remedy (U.S. Chamber of Commerce Foundation: Center for Education and Workforce, 2014). Another example is the Department of Labor's Registered Apprenticeship Program(https://www.doleta.gov/OA/apprenticeship.cfm), which aims to connect individuals with careers from an early age. Labor-market intermediaries such as employment agencies, employer relationships with technical colleges or other institutions, and employer-provided training are some other initiatives proposed to bridge the gap between supply and demand (Weaver, 2017, August 25). However, most of those initiatives are not subject- or program-specific, require highly motivated personnel from all the involved parties, and have a lack of incentives and visibility. Therefore, those initiatives may not be scalable or sustainable in the long run.

On the other hand, higher education educators and administrators have been tackling these challenges from a different aspect, focusing on academic program renovation and resource realignment (College for Every Student (CFES), 2016; Forshaw et al., 2016). For instance, workforce readiness could be improved through more well-designed and well-planned internships, more guidance for career preparation, and better access to preparation tools (Hanover Research, 2016). Some universities have implemented connected curriculum or programs as a strategic framework to foster student research and internship opportunities as well as enrich the curriculum and student experiences (Fung, 2017). Another study uses a data-driven approach to reflect on the gaps and overlaps between the curriculum and skillsets in latest job postings (Green, Liu, & Murphy, 2019). To address the concern of conspicuously the low number of women professionals in the cybersecurity workforce (Frost & Sullivan, 2015, 2017), some educators have engaged in broadening the participation by females and preparing female students for the cybersecurity career (Liu & Murphy, 2016).

The authors have adapted multiple frameworks and best practices based on the above literature. According to the program outcome assessments, a series of job readiness activities embedded in the curriculum and extended over extracurricular have received generally positive feedback from the students. However, quite a few students, especially those who have limited working experiences in the IT and cybersecurity fields, expressed pressing needs to have faculty review their resumes and provide subject matter related suggestions.

This study proposes an innovative as well as systematic approach- an Electronic Resume Analyzer Portal (e-RAP) to analyze students' resume, evaluate them based on job postings, and generate reports with ratings and suggestions on specific skill area(s) the students could or should work on. This tool will complement the ongoing workforce readiness programs and initiatives in the author's department and strengthen the students' portfolio and career readiness in the long run.

2. BACKGROUND

While there is a talent shortage in the IT field, particularly in cybersecurity, organizations receive many applications for each position advertised that come from a variety of sources such as job portals, company web sites, and emails. It is no longer practical for these applications to be screened manually by most human resources departments or hiring managers. Consequently, automated resume scanning tools, also called automated tracking systems, are commonplace, many using NLP techniques to screen resumes as a first pass in the hiring process. In fact, this is not a new practice. Back in 2012, the Wall Street Journal reported that resume screening software was being used by around 90% of companies and it would be exceptionally rare to find a Fortune 500 company not using these systems (Weber, 2012). It is estimated that around 75% of resumes received by a company are never looked at by a human being (Bell, 2018).

Many of these automated tracking systems focus on the keywords in the position description and look for the exact same word in the applicant's resume. In the IT field, however, these words are constantly changing, whether it being new technologies such as DevSecOps or low code; new terms such as scripting or front-end/back-end developer; or new job descriptions such as data engineer or cyber threat hunter. In addition, there is little consistency between companies in how they describe the position requirements. For example, one may say Linux, others may be very specific about the Linux version, such as Ubuntu.

.

Currently many students who submit their resumes to job vacancies advertised on job sites such as Indeed simply do not get any response as their resumes are filtered out by the automated tracking system. Career service centers may help students strengthen their resumes in terms of format, language use, and human resource concerns, but their staff are less likely to answer questions as for what technical skills should be highlighted or which industry certification should be added to make the resumes stand out or match recruiters' criteria. Faculty members may review a student's resume and provide feedback and suggestions on enhancing his or her resume from the subject matter aspect, however, today's students are often reluctant to ask for help (Lammers, 2017). At our institution, we have developed various avenues to help the students prepare for the workplace as discussed in the introduction section. Although helpful for the students, these techniques are often not responsive enough to the students who want immediate responses and expect technology to help them get the answer.

So, we set about developing an automated tool that could help students help themselves and focus their resumes on specific aspects of their job search, based on what employers, and their automated tools, are looking for. It is based on using machine learning from a bank of advertised job positions in certain job categories, learning what and how these job skills are being described by the various organizations, and then matching the contents of a student's resume against them. A visual report of the analysis shows these results for the top job categories related to the student's resume, indicating how weighted the resume is towards each of these job categories.

The range of jobs selected for the study reflects both the focus of their education (such as cybersecurity) but also the background of many of our student population who are career changers or whose background experience is in jobs such as a camp counselor. Showing the weight of these factors is designed to help them focus on their intended career and to enhance their resume to reflect that.

3. METHODOLOGY

Overview

The overall process is shown in Figure 1.

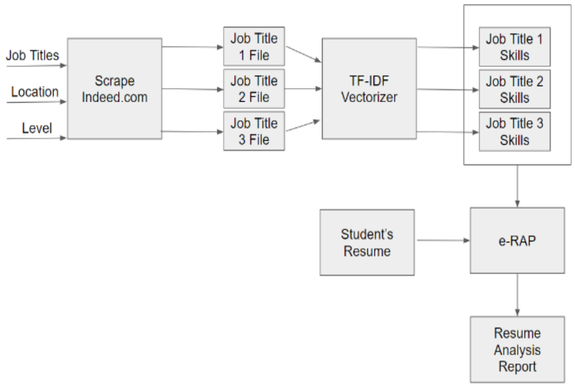


Figure 1 – The e-RAP Analysis Process

We first selected a set of job categories which were reflective of the students currently in the IT program, many specializing in cybersecurity. Some were designed to reflect the various aspects of the IT field and included 'Cybersecurity', 'Data Scientist', 'Software Engineer', 'Cloud Computing', 'Digital Writing', 'Management Analyst', 'Mathematician', and 'IT support'. Others reflected the other disciplines that were common in our career and major changers such as 'Economist', 'Nursing', 'Human Resources'', Financial Analyst', 'Psychologist', and 'Biologist'. The final set was more reflective of the non-technical job experience that students may have had in high school and college such as 'Counselor' and 'Editor'. This initial set of job categories can be easily modified as the effectiveness of e-RAP is evaluated.

One thousand jobs per job category described above were harvested from Indeed.com (www.indeed.com). Indeed was the initial job board selected based on student's use and the fact that Indeed extracts postings from many job boards, company career sites, associations, and other sources of job postings. The job search was restricted to entry-level jobs in the local area to mirror the expected search criteria that the student would use in looking for an internship or a first job.

Next, the key topics and job skills were automatically extracted for each job category based on the job descriptions, eliminating terms that are common in all job description (e.g., EEO compliance). The initial extraction was based largely on single-word identification (e.g., data) with some compound words based on associations (e.g., quality assurance)

The text of each student's resume was then scraped and compared against these key topics and skills. A visual report of the analysis of the student's resume was produced showing the top 5 job categories represented in the resume. These 5 job categories are displayed on a percentage basis, indicating how weighted the resume is towards each of the job categories. It allows the student to see visually what job categories are most represented in their resume, whether that was their intended job focus or not. For each job category, a competency rating (expert level gets 5 stars and layman level gets 1) is displayed. Further, the report suggests 5 terms that are currently not in the resume that would bolster its relevance for that topic. For example, in a description of courses taken or projects conducted.

Data Curation

To create e-RAP we required two datasets, a set of current jobs and a set of resumes. To harvest current jobs, we created a web scraper for the job site Indeed.com. This scraper was programmed to take 3 parameters: a job category, a location, and a job expertise level. The output of the scraper was a job file with only the text from the relevant results. Each line of the file was a complete job description. We additionally cleaned the data by removing all HTML and Javascript code automatically.

For this study, we restricted the search to jobs in the DC metro area and entry-level jobs only. This is the job profile for most of our undergraduate students as they graduate. For instance, our scraper created the file cybersecurity.txt by pulling the first 100 pages of a search for select parameters such as Entry Level Jobs in Washington DC for 'cybersecurity', capturing 1,000 jobs. Results for each job category scrape was then saved for analysis.

In order to compare these jobs with our student population, we collected an initial set of about one hundred undergraduate student resumes. A data extraction tool was programmed to extract any text in the resume as long as it was a .doc, .docx, or .pdf file. In cases where the student turned in a different format, we converted it to a pdf via a pdf printer.

Data Cleaning

Most job postings on Indeed.com contain basic boilerplate sentences for human resource language and basic company introductions. Generally speaking, this type of information is not of interest to our resume analysis. While a student might care about 401k and paid time off, e-RAP removes such language so that we can focus on attributes of the job posting that may translate well to a resume. To remove these terms from our evaluation, we hand-curated a list of 300+ common employment terms. In addition to human resource language, frequent words (the, is to, a, an) are not particularly useful for our analysis. These words are often called "stop words" and these stop lists come with many NLP kits. We removed all stop words from the job postings using NLTK's stop word list for English (https://www.nltk.org/). This same process was performed for the students' resumes.

After both the stop words and the employment terms were removed, we had a list of terms that mostly relate to job skills needed to fill the position. Our hypothesis is that these are the words useful for evaluation of a resume, and also would be positively reviewed by the automated tracking system if students added them to their resume.

Analysis Techniques

To compare a resume to a job posting, we created a process that would rank the relevancy of each job skill to a specific field. To do this, first we extracted the key terms. We conducted this extraction automatically using the common information retrieval technique of term frequency–inverse document frequency (tf-idf). Tf-idf ranks the importance of a term to a document compared to other documents (Wu et al, 2008). Since all job posting for a job category were stored in a singular file and each job category has its own file, this was a fairly straightforward task.

Having a value for each term, we could turn each document into a vector of values representing the terms in the document. To find out if a resume is similar to a job category, we took the cosine similarity measure between the industry vector and a tf-idf vector for the resume. If the cosine value is equal to 1, they are the same document, if the cosine value is 0, the documents share no terms in common. (Tata and Patel, 2007). We compared a resume against each field tf-idf vector and reported the top 5 fields by similarity. These scores allowed us to show the distribution of the student's resume. For example, the top 5 job categories for a student's resume may be: Cybersecurity, Data Scientist, Natural Language

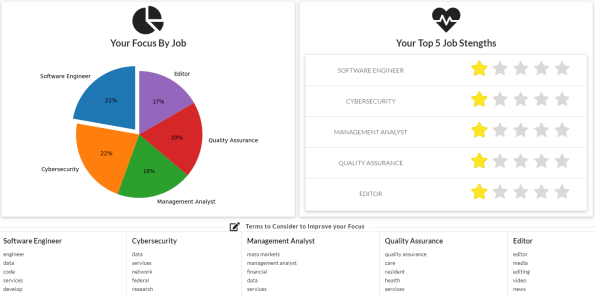
Processing, Software Engineer, and IT Support. We went a step further and told the user the distribution such as 70% Cybersecurity, 20% Data Scientist, 5% Natural Language Processing, 4% Software Engineer, and 1% IT Support.

The student's weighted distribution will always equal 100% but just because 70% of the student's resume is focused toward a field does not make the student qualified. To provide this additional information we also give a star rating 1-5 based on how high the particular job similarity is to their resume. This helps adjust for situations that often happen in undergraduate programs where the student's main experience is still in high school level jobs, so while they focus on one job area, they do not have the coverage in that job area to be considered qualified.

The final part of our analysis gives suggestions for how to improve the resume. This is simply the top 5 terms for a given job category per tf-idf score that is not currently in the student's resume. This allows the student to make iterative changes to their resume and see the overall effect on our analysis. For example, if they see data, they can think back to their database class, reflecting on the tools that they used and any related project work.

4. INITIAL RESULTS

In this section, we illustrate the e-RAP results with three examples of the visual report produced and given to the student. In Figure 2, we first illustrate a student whose resume is very broad.



Your Top 5 Job Stengths

SOFTWARE ENGINEER	★☆☆☆☆
CYBERSECURITY	★☆☆☆☆
MANAGEMENT ANALYST	★☆☆☆☆
QUALITY ASSURANCE	★☆☆☆☆
EDITOR	★☆☆☆☆

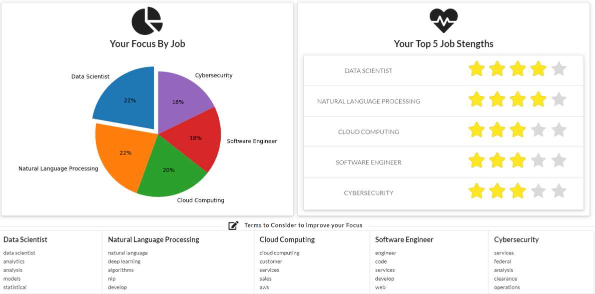
Terms to Consider to Improve your Focus

Software Engineer	Cybersecurity	Management Analyst	Quality Assurance	Editor
engineer	data	mass markets	quality assurance	editor
data	services	management analyst	care	media
code	network	financial	resident	editing
services	federal	data	health	video
develop	research	services	services	news

Figure 2: e-RAP Results for a Typical Early Undergraduate IT/Cybersecurity Student (see appendix)

In this example, the resume text is split across 5 areas fairly evenly, but the student is not an expert in any of them. This tells us that the student mentioned a few terms related to each job category but not specific enough to show any support for those skills.

In contrast, Figure 3 illustrates an IT student in the data science specialty that has some industry experience gained through a series of internships while in college.



Your Top 5 Job Stengths

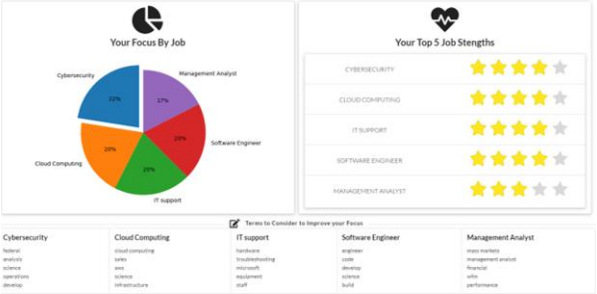
DATA SCIENTIST	★★★★☆
NATURAL LANGUAGE PROCESSING	★★★★☆
CLOUD COMPUTING	★★★☆☆
SOFTWARE ENGINEER	★★★☆☆
CYBERSECURITY	★★★☆☆

Terms to Consider to Improve your Focus

Data Scientist	Natural Language Processing	Cloud Computing	Software Engineer	Cybersecurity
data scientist	natural language	cloud computing	engineer	services
analytics	deep learning	customer	code	federal
analysis	algorithms	services	services	analysis
models	nlp	sales	develop	clearance
statistical	develop	aws	web	operations

Figure 3 – e-RAP Results for an IT Student with Relevant Internship Experience (see appendix)

We see the resume still shows five areas that are all applicable to the student but the star ranking shows their true expertise. The student gets 4 stars for Data Science and for NLP (a subdiscipline of Data Science). If this particular student was hoping to get into cybersecurity, however, the student could see the list in the bottom right and see that it might be helpful to add any experience that dealt with services, federal, analysis, clearance, and operations. We can run pre- and post-tests where students update their resume based on the feedback report, submit to the e-RAP tool, for a second report, and compare the differences and changes.



Your Top 5 Job Stengths

CYBERSECURITY	★★★★☆
CLOUD COMPUTING	★★★★☆
IT SUPPORT	★★★★☆
SOFTWARE ENGINEER	★★★★☆
MANAGEMENT ANALYST	★★★☆☆

Terms to Consider to Improve your Focus

Cybersecurity	Cloud Computing	IT support	Software Engineer	Management Analyst
federal	cloud computing	hardware	engineer	mass markets
analysis	sales	troubleshooting	code	management analyst
science	aws	microsoft	develop	financial
operations	science	equipment	science	wfm
develop	infrastructure	staff	build	performance

Figure 4 – e-RAP Results for an IT Student in the Networking and Cybersecurity Specialty (see appendix)

Finally, Figure 4 illustrates an IT student in the networking and cybersecurity specialty that has some experience gained through a series of research projects and an internship while in college. Her job strengths are strong in cybersecurity and in the related fields of software engineering, IT support, and data science. While management analyst represents a high percentage, it is at a lower skill level. The suggested words (e.g. federal) reflect the jobs in our region with a high percentage of jobs in the federal sector, government and government contractors.

5. Further Work

To date, the analysis has been performed in the background, and faculty have reviewed the results. They are providing feedback to the system so that the e-RAP system can receive annotations on its analysis and thus improve the results. Once this has been concluded, a front-end will be developed so that the students can run the analysis themselves. Each time when they change their resume, the students would also obtain feedback provided by the e-RAP. Another future direction will be to delve into the individual cybersecurity jobs themselves as reflected in the NICE Cybersecurity Workforce Framework (Newhouse et al, 2017) and the Cyberseek cybersecurity pathways (https://www.cyberseek.org/pathway.html).

To our best knowledge, no such tool as e-RAP is made and employed in a university setting to help students sharpen their resumes from the skills angle. Some similar tools might exist in the market, but most likely they are proprietary and not easy to access for college students. We will make it accessible to all of our IT and cybersecurity students. The ultimate goal is to extend this tool beyond these programs so that students from other programs can also access and have their resumes analyzed.

Finally, the authors recognize that they are in the early stages of success with this project and that the next steps will be a systematic and rigorous evaluation process to assess the effectiveness of the tool. For example, we can have some human annotators rate the accuracy of the tool. We can also run a test where students look at the report generated by the e-RAP, change their resumes based on the suggestions, and run the e-RAP on the updated resume to generate a second report. We will have a review panel including our alumni working in the IT field or doctoral students with recruiting experiences to review pre- and post-versions of a student's resume and assess if the changes would increase their chances of being recruited.

6. REFERENCES

Bell, T. (2018). The secrets for beating an applicant tracking system, CIO, April 17, 2018, retrieved from https://www.cio.com/article/2398753/careers-staffing-5-insider-secrets-for-beating-applicant-tracking-systems.html

(ISC)². (2018, October 17). Cybersecurity Professionals Focus on Developing New Skills as Workforce Gap Widens: (ISC)² CYBERSECURITY WORKFORCE STUDY, 2018. Retrieved from https://www.isc2.org/-/media/ISC2/Research/2018-ISC2-Cybersecurity-Workforce-Study.ashx?la=en&hash=4E09681D0FB51698D9BA6BF13EEABFA48BD17DB0

College for Every Student (CFES). (2016). New Dimensions of College and Career Readiness: Implications for Low-Income Students. Retrieved from http://www.collegeaccess.org/SD08092016Article3

Forshaw, M., Solaiman, E., McGee, O., Firth, H., Robinson, P., & Emerson, R. (2016). *Meeting Graduate Employability Needs through Open-source Collaboration with Industry.* Paper presented at the Proceedings of the 47th ACM Technical Symposium on Computing Science Education, Memphis, Tennessee, USA.

Frost & Sullivan. (2015). The 2015 (ISC)² Global Information Security Workforce Study. Retrieved from https://www.isc2.org/-/media/Files/Research/GISWS-Archive/GISWS-2015.ashx?la=en&hash=01D5BD45477FB7B45EF773366CF7D1D9BB6A6753

Frost & Sullivan. (2017). 2017 Global Information Security Workforce Study: Benchmarking Workforce Capacity and Response to Cyber Risk. Retrieved from https://iamcybersafe.org/wp-content/uploads/2017/06/europe-gisws-report.pdf

Fung, D. (2017). *Connected Curriculum for Higher Education* (1st ed.). London, England: UCL press.

Green, N., Liu, X., & Murphy, D. (2019). Revisiting an Educator's Dilemma: Using Natural Language Processing to Analyze the Needs of Employers and Inform Curriculum Development. *Journal of Computing Sciences in Colleges, 34*(3), 97-107.

Hanover Research. (2016). McGraw-Hill Education 2016 Workforce Readiness Survey. Retrieved from https://s3.amazonaws.com/ecommerce-prod.mheducation.com/unitas/corporate/ideas/2016-student-workforce-readiness-survey-expanded-results.pdf

Lammers, W. (2017) Why Won’t They Ask Us for Help, Faculty Focus, march 24, 2017, retrieved from https://www.facultyfocus.com/articles/edtech-news-and-trends/why-students-dont-ask-for-help-and-what-you-can-do-about-it/

Liu, X., & Murphy, D. (2016, 28-30 Sept. 2016). *Engaging females in cybersecurity: K through Gray.* Paper presented at the 2016 IEEE Conference on Intelligence and Security Informatics (ISI).

Newhouse, W., Keith, S., Scribner, B. and Witte, G. (2017).National Initiative for Cybersecurity (NICE) Cybersecurity Workforce Framework, NIST< August 2017

McGraw-Hill Education. (2018). 2018 Future Workforce Survey. Retrieved from https://s3.amazonaws.com/ecommerce-prod.mheducation.com/unitas/corporate/promotions/2018-future-workforce-survey-analysis.pdf

Tata, S. and Patel, J. (2007). Estimating the selectivity of tf-idf based cosine similarity predicates. SIGMOD Rec. 36, 4 (December 2007), 75-80.

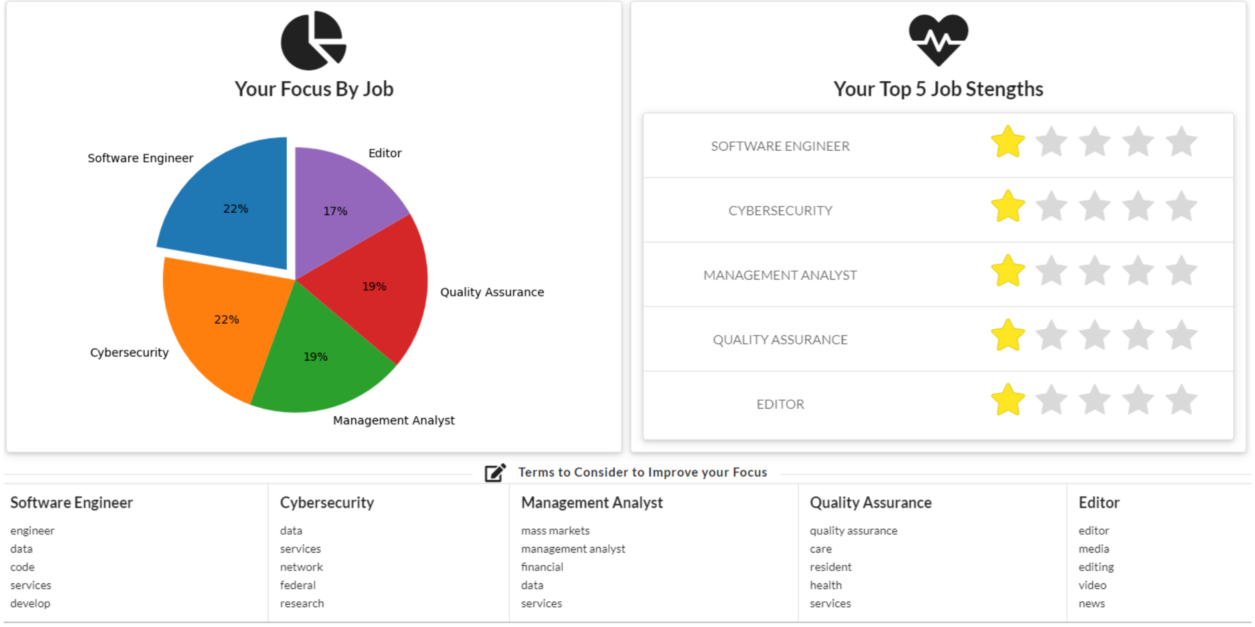
U.S. Chamber of Commerce Foundation: Center for Education and Workforce. (2014). Managing the Talent Pipeline: A New Approach to Closing the Skills Gap. Retrieved from https://www.uschamberfoundation.org/sites/default/files/Managing%20the%20Talent%20Pipeline.pdf

Weaver, A. (2017, August 25). The Myth of the Skills Gap. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/608707/the-myth-of-the-skills-gap/

Weber, L. (2012). Your Resume vs Oblivion, The Wall Street Journal, January 24, 2012), retrieved from https://www.wsj.com/articles/SB10001424052970204624204577178941034941330

Wu, H. C., Luk, R. W. P., Wong, K. F., & Kwok, K. L. (2008). Interpreting tf-idf term weights as making relevance decisions. ACM Transactions on Information Systems (TOIS), 26(3), 13.

Appendices and Annexures



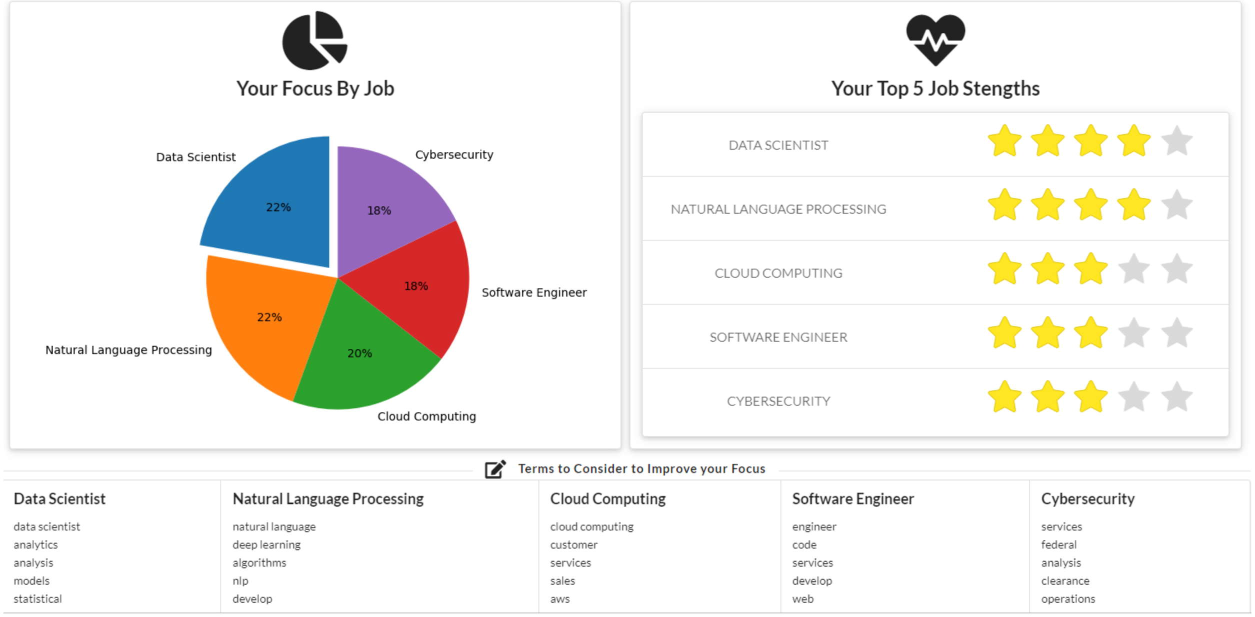
Your Top 5 Job Stengths

SOFTWARE ENGINEER	★☆☆☆☆
CYBERSECURITY	★☆☆☆☆
MANAGEMENT ANALYST	★☆☆☆☆
QUALITY ASSURANCE	★☆☆☆☆
EDITOR	★☆☆☆☆

Terms to Consider to Improve your Focus

Software Engineer	Cybersecurity	Management Analyst	Quality Assurance	Editor
engineer	data	mass markets	quality assurance	editor
data	services	management analyst	care	media
code	network	financial	resident	editing
services	federal	data	health	video
develop	research	services	services	news

Figure 2: e-RAP Results for a Typical Early Undergraduate IT/Cybersecurity Student



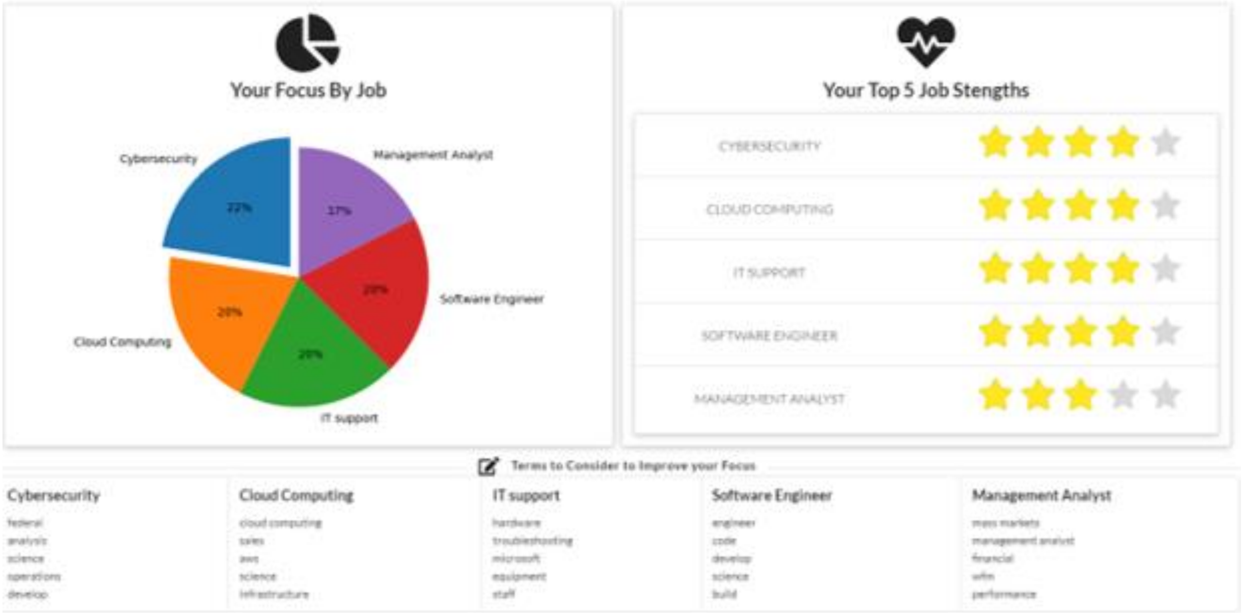
Your Top 5 Job Stengths

DATA SCIENTIST	★★★★☆
NATURAL LANGUAGE PROCESSING	★★★★☆
CLOUD COMPUTING	★★★☆☆
SOFTWARE ENGINEER	★★★☆☆
CYBERSECURITY	★★★☆☆

Terms to Consider to Improve your Focus

Data Scientist	Natural Language Processing	Cloud Computing	Software Engineer	Cybersecurity
data scientist	natural language	cloud computing	engineer	services
analytics	deep learning	customer	code	federal
analysis	algorithms	services	services	analysis
models	nlp	sales	develop	clearance
statistical	develop	aws	web	operations

Figure 3 – e-RAP Results for an IT Student with Relevant Internship Experience



Your Top 5 Job Stengths

CYBERSECURITY	★★★★☆
CLOUD COMPUTING	★★★★☆
IT SUPPORT	★★★★☆
SOFTWARE ENGINEER	★★★★☆
MANAGEMENT ANALYST	★★★☆☆

Terms to Consider to Improve your Focus

Cybersecurity	Cloud Computing	IT support	Software Engineer	Management Analyst
federal	cloud computing	hardware	engineer	mass markets
analysis	sales	troubleshooting	code	management analyst
science	aws	microsoft	develop	financial
operations	science	equipment	science	wfm
develop	infrastructure	staff	build	performance

Figure 4 – e-RAP Results for an IT Student in the Networking and Cybersecurity Specialty