LEAP@WPI/QCC: Improving a Service Model Through Customer Discovery and Web-Scraping

An Interactive Qualifying Project
submitted to the faculty of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirement for the
degree of Bachelor of Science

13th May 2020

Authors
Zachary Langlois
Chris Watson
Hamayel Qureshi

Submitted to:
Professor Douglas Petkie
Dr. James Eakin
WPI

Abstract

The Lab for Education and Application Prototypes at Worcester Polytechnic Institute and Quinsigamond Community College (LEAP@WPI/QCC) has been created as a center for advancement in the quickly emerging field of integrated photonics. To effectively achieve the ambitions of LEAP@WPI/QCC, the facility needs actionable data about the industry at large and the individuals or companies that may decide to sign as members to LEAP@WPI/QCC's services. Using customer discovery methods outlined by the National Science Foundation's Innovation Corps Program (I-Corps) as well as a web-scraping methodology to pull informative data from the internet, our project compiled data that can inform LEAP@WPI/QCC's Industry Affiliates Program (IAP). Data was collected in order to provide recommendations for LEAP@WPI/QCC to adapt its IAP and approach to acquiring members according to the needs and requirements of target companies and networking societies. Using collected data from interviews and web-scraping, the project lays out some directions for LEAP@WPI/QCC including industries to focus on and issues in new talent lacking some necessary skills for the integrated photonics space.

Acknowledgments

The group would like to thank Professor Douglas Petkie and Dr. James Eakin for their constant guidance throughout our three terms working on this project. Professor Petkie and Dr. Eakin have been integral in providing structure to our work and constructive feedback every step of the way from the very beginning all the way to the conclusion of our work on the IQP. We would also like to thank our interviewees who were very generous in sparing their time in order to help us with our customer discovery data collection. Finally, we would like to thank those who helped us as we participated in the I-Corps program - Mr. John Marczewski, our industry mentor, and Mr. Todd Keiller, Mrs. Beth Kahn, and Dr. Joe Bush, all of whom work as part of the WPI I-Corps Site Program Teaching Team and provided us with insightful recommendations and analysis for our interviews.

Authorship

This project is an amalgamation of joint effort from each team member. Every individual put in an equal amount of work into the report. The writing for this report was dictated by an outline each one of us contributed to, and after the conclusion of the writing process, we came together to edit and critique each section. Individuals were responsible for their primary focus in methodology: Zachary and Chris worked on the customer discovery portions and Hamayel worked on the web-scraping sections. This project would not have been complete without the full participation of each person in our team.

Table of Contents

Introduction

What Is Photonics?

Over the course of the past few decades, photonics and optics have played a very important part in the development of technology that we have characterized as valuable to our everyday lives. Photonics and optics, specifically fiber optics, are significantly responsible for generating information, displaying or storing information, and communicating this information.[1] Photonics involves the technology that generates and harnesses light, and also produces, detects, transfers, and manipulates information with photons. The applications of this technology range from lasers, fiber optics, electro-optical devices, and optics, and these applications are implemented in a wide variety of fields.[2]

Within the field of photonics, there is currently an important and relatively new branch of technology that is emerging - integrated photonics. The idea behind integrated photonics is relatively straightforward and similar to the concept of an integrated electronic circuit. However, rather than transmitting and processing electronic signals, an integrated photonic circuit would transmit and process light. An integrated photonic circuit consists of photonic devices and instruments which are fabricated in an integrated design on a flat substrate. Using photons to transmit and process information allows information to be transmitted at the speed of light and without expending much energy in comparison to an integrated electronic circuit which would transmit information slower and at the expense of more energy.[3]

What is LEAP@WPI/QCC?

The AIM Photonics Academy Lab for Education & Application Prototypes (LEAP) facility began its development process in 2018 at WPI in collaboration with Quinsigamond Community College. With $4 million in initial funding from the Commonwealth of Massachusetts, the facility set out to “support the development of the integrated photonics

[1] https://www.sciencedirect.com/topics/engineering/photonics
[2] http://www.op-tec.org/what-is-photonics
[3] https://aimphotonics.academy/about/what-integrated-photonics

manufacturing sector in Central Massachusetts[4].” The AIM Photonics academy itself is a member of the AIM Photonics Institute under Manufacturing USA[5] with an overall goal to advance manufacturing processes and technologies to remain a leader in the world economy.

What are LEAP@WPI/QCC’s Goals?

The LEAP@WPI/QCC facility’s fundamental objectives are focused on three concepts - training and education, prototyping, and collaborations on research associated with integrated photonics. These three objectives are essential in supplying the photonics industry with qualified and capable personnel, providing groups involved with or interested in photonics the opportunity to cooperate on research projects, and developing potentially significant and practical prototypes that enable further progress within the photonics industry. All three of LEAP@WPI/QCC’s principal target goals are implemented in the effort to expand the photonics industry and serve a key role in the furtherment of the industry’s technology and capabilities.

By implementing an environment that can be used for specific teaching and firsthand preparation focused on integrated photonics and technology associated with photonics, LEAP@WPI/QCC can help prepare a well trained and educated workforce in the New England area. Workforce development is very important for the field of integrated photonics, since the relatively new industry will adapt its needs according to the direction that the industry moves toward and require technical expertise and experience from many scientific and engineering disciplines. LEAP@WPI/QCC’s focus on training and educating a labor force will be based on the photonics industry’s changing needs. Preparing a capable workforce does not only involve individuals who are at a college-level of education, however. LEAP@WPI/QCC offers “educational programs which will engage students at all levels - from K-12 outreach activities, to community college training programs, to graduate education.[6]” Programs that are offered to students at all levels will allow LEAP@WPI/QCC to prepare and produce a dynamic and efficient workforce for the photonics industry, and possibly industries that work closely with photonics.

---

[4] https://www.wpi.edu/news/laser-focused-photonics-initiatives
[5] http://www.aimphotonics.com/about-us
[6] https://www.wpi.edu/research/core-research-facilities/leap

LEAP@WPI/QCC is also concentrated on creating an environment for members and clients which functions as a facility to construct and work with prototypes that make usage of "photonic integrated circuit technologies essential for manufacturing advancement in fields such as in medical devices, 5G and 6G wireless, datacom, chemical and biological sensors.[7]" This environment would be available to academia, industry, and government with the intent to test and design photonic integrated circuits and devices. Equipped with advanced equipment and tools, the LEAP@WPI/QCC facility has the ability to play a significant role in the New England area concerning the development of prototype technology associated with photonics.

The final primary objective looking to be accomplished at the LEAP@WPI/QCC facility involves opportunities presented with research collaboration. These collaborations on research would be directed towards photonics areas pertinent with "nanoscale and microscale prototyping development, optical and electrical device characterization, fiber-chip interfacing, terahertz sensing, fiber optics, and noninvasive optical metrology for reliability testing." [4] By playing an important part in the collaboration on research projects, LEAP@WPI/QCC can offer capability in pushing the field of integrated photonics further.

What Can an Emerging Technology Like Integrated Photonics Learn from More Mature Technologies?

In comparison with another technology such as semiconductors/microelectronics and integrated electronics, which has had a steadily growing industry and technology that has improved vastly since the time it emerged, integrated photonics is certainly on the forefront of becoming even more important technology than it currently is.[8] Integrated photonics could be on track to see the same trends that integrated electronics saw as a technology and industry. It should be safe to assume that a discipline such as integrated photonics has the room to grow and become an influential technology that will play an important role in our society's development in the years to come.

[7] https://www.wpi.edu/research/core-research-facilities/leap
[8] https://www.nap.edu/read/13491/chapter/2

Some of the fields where photonics has a considerable impact on include communication, manufacturing, medicine, industrial processes, and aviation. As the field of integrated photonics grows, the technology will move forward and branch out its applications into more fields and have an increasing amount of influence and impact on these fields and more. Integrated photonics has emerged having the potential that a technology such as microelectronics once had, some time ago. The technology involved with microelectronics has grown substantially since its introduction, and its industry has followed the same pattern. If integrated photonics can learn anything from an industry such as the microelectronics industry, it should look to follow similar paths that the microelectronics industry had taken in order to become as immensely influential and successful in today's economic landscape.

What Is the Objective of Our IQP?

As with any venture in its early stages, the primary challenge that LEAP faces in its growth is to refine its range of services to meet the needs of its potential customers enough to attract them to become members. The Industry Affiliates Program (IAP) was developed by hypothesizing about archetypal customers used to represent potential members, such as networking societies, research firms, and manufacturing industries. The objective of our work was to evaluate the accuracy of these archetypes and their predicted needs by conducting interviews and gathering web data concerning the current pain points and focuses of companies and societies in the photonics, optics, and microelectronics industries.

Photonics stands as an attractive long-term alternative or replacement for vital and ubiquitous microelectronics[9]. Integrated photonics have begun to move out of early concept testing and into more rigorous and direct research for applications and manufacturing. As the industry makes this shift, it will be necessary to change accompanying technologies and the training and skills of the workforce ready to work in photonics. Our aim is to provide a summary of the needs which societies and companies are currently facing in their exploration of photonics and some recommendations on which services within the IAP to emphasize based on the current frustrations of our target customers.

[9] https://www.aimphotonics.com/what-is-integrated-photonics

The Industry Affiliates membership consists of several member benefits based on the tier level. Some of these benefits include: "the use of laboratory and cleanroom facilities; opportunities to engage with other experts at the forefront of integrated photonics and related technology sectors; new distribution channels to promote company, products, and services; access to students with various backgrounds (AS/BS/MS/Ph.D.) to fulfill hiring needs[10]."

Currently, there is a three tiered membership system with a certain number of benefits for each tier. The Tier 1 membership, the most inclusive and expensive option, was constructed for a company that wishes to actively pursue exploratory projects using the facility's equipment and personnel. This top-tier membership would also provide maximum networking benefits and access to various communication and hiring resources. Tier 2 is currently viewed as a trial form of LEAP@WPI/QCC's full services for companies that are uncertain about the value of a full membership, but want to get involved and use some of the services in order to decide whether or not to move up to Tier 1 in the future. Tier 3 was created to attract networking societies and non-profits who are not looking for direct research and prototyping opportunities, but are necessary in order to accomplish the ecosystem-building goal of the facility. The below chart highlights the tier system and pricing:

Annual Membership and Benefits	Tier 3*	Tier 2	Tier 1
	$1,500	$7,500	$37,500
Technical Consulting Sessions		✓	✓
Project Feasibility Study and Topic Exploration		✓	✓
Complimentary Registration to LEAP@WPI/QCC Symposium	✓	✓	✓
Direct Access to Student Talent	✓	✓	✓
Job and Internship Posting (through Handshake)		✓	✓
Career Fair Table and Registration			✓
Members-only Networking Events	✓	✓	✓
Promotion through Awareness and Visibility	✓	✓	✓
Laboratory Training		✓	✓
Reduced Rates for Laboratory Usage		✓ Reduced	✓ Internal
Event Hosting Assistance	✓		✓
Student Project Involvement			✓
Advisory Board Participation			✓

Figure 1. Proposed Annual Membership and Benefits Tiers for LEAP@WPI/QCC

[10] https://www.wpi.edu/research/core-research-facilities/leap

A detailed description of some of the benefits is available in the following chart:

Annual Industry Affiliate Membership Details
Technical Consulting (Group) Sessions: Tier 1 & 2 Industry Affiliates can participate in one group consult with up to four faculty, staff for up to one hour during the membership period. Additional group consults and the associated rates and duration will be at the discretion of the LEAP@WPI/QCC faculty. Topics can research, manufacturing, business or workforce development related based on the needs of the Industry Affiliate.
Technical Consulting (Individual) Sessions: Tier 1 Industry Affiliates can participate in two consults with a faculty or staff for up to one hour per session during the membership period. Tier 2 Industry Affiliates can participate in one consult with a faculty or staff for up to one hour during the membership period. Additional consults and associated rates will be at the discretion of the LEAP@WPI/QCC faculty.
Project Feasibly Study Tier 1 & 2 Industry Affiliates can propose topics to explore during the membership and LEAP@WPI/QCC affiliates will dedicate up to 8-person hours of resources to explore potential methods, or processes to accelerate the project if requested.
Registration to annual LEAP@WPI/QCC symposium: Tier 1 Industry Affiliates will three complimentary registrations to the annual LEAP@WPI/QCC symposium. Tier 2 & 3 Industry Affiliates will receive one complimentary registration to to the annual LEAP@WPI/QCC symposium. The symposium will be planned for Spring 2020.
Access to student resumes and student presentations Tier 1 & 2 Industry Affiliates will receive access to student resumes books through the WPI Career Development Center Tier 1, 2, & 3 Industry Affiliates will receive invitations to attend the student poster presentations at LEAP@WPI/QCC events
Job & internship posting through Handshake Tier 1 & 2 Industry Affiliates can participate in job and internship posting through WPI's Handshake system.
Career Fair Table and registration Tier 1 Industry Affiliates receive Bronze level sponsorship at one campus career fair, with preferred table access, registration, and visibility.
Members Only Networking Events Tier 1, 2 & 3 Industry Affiliates can participate in members only networking events with LEAP@WPI/QCC faculty & staff.
Marketing & recognition at LEAP@WPI/QCC LEAP@WPI/QCC will display digital signage of Tier 1, 2, & 3 Industry Affiliates logos at the LEAP@WPI/QCC facility, LEAP@WPI/QCC hosted and sponsored events, and other LEAP@WPI/QCC online distribution channels.
Laboratory Training Tier 1 Industry Affiliates can receive one session of up to four hours of general laboratory training and four sessions up to two hours of specialty laboratory equipment training to obtain access and usage of equipment. Tier 2 Industry Affiliates can receive one session up to four hours of general laboratory training and one session up to two hours of specialty laboratory equipment training to obtain access and usage of equipment. LEAP@WPI/QCC will provide tools for Industry Affiliates to reserve equipment online and offsite in advance. Laboratory training sessions are limited to a group of up to three people.
Reduced Rates for Laboratory Equipment Tier 2 Industry Affiliates will receive 20% discounted rates on laboratory equipment and cleanroom facility usage. Rates will vary based on equipment. LEAP@WPI/QCC will provide a rate sheet for equipment and cleanroom facility usage. Unassisted usage of laboratory equipment can begin after completion of training sessions.
Internal (lowest) Rates for Laboratory Equipment Tier 1 Industry Affiliates will receive internal (lowest) rates on laboratory equipment and cleanroom facility usage. Rates will vary based on equipment. LEAP@WPI/QCC will provide a rate sheet for equipment and cleanroom facility usage. Unassisted usage of laboratory equipment can begin after completion of training sessions.
Event Hosting Tier 1 & Tier 3 Industry Affiliates can host one event at LEAP@WPI/QCC facility to promote company, product, services or professional society/chapter meeting during the membership period. Facility availability will be first come first served. LEAP@WPI/QCC will provide light refreshments and help promote events if requested.
Define & Direct Student Project Tier 1 Industry Affiliates can define and direct one student project. The project scope, complexity, duration and deliverables will be agreed upon the the Industry Affiliate and the LEAP@WPI/QCC Director and related faculty or staff member.

Figure 2. Detailed Description of Membership Benefits[11]

[11] Sourced from LEAP@WPI/QCC

Methodology 1: Customer Discovery

In order to aid LEAP@WPI/QCC in its efforts of finding viable relationships with companies and societies in the area, we employed a technique that pertains to customer discovery, so that we could develop a better understanding of developing a network with practical connections between particular companies and societies and LEAP@WPI/QCC. Customer discovery in the case of LEAP@WPI/QCC involves understanding a particular set of customers (companies associated with microelectronics, optical technologies and photonics in the New England area) and what their needs are, and forming a strategy that would best fulfill these customers' needs. It also involves obtaining a broader view of the electronics, optics, and photonics industries, and what LEAP@WPI/QCCcan do to offer the most optimal services to companies within those industries.

Our Objective: Gather Information on Companies in the Industry (Needs/Insight on Connections and Relationships with Academia)

Our objective was to compose an approach towards customer discovery that garnered a better understanding of how LEAP@WPI/QCC can acquire more members. For us to understand this, we needed to inquire about what type of companies and societies that LEAP@WPI/QCC was seeking to establish relationships with, what type of relationships these targeted companies and societies would consider with a facility like LEAP's, and what type of services and communication on both sides of the relationship would make it as practicable as possible. The central objective we were looking to accomplish through customer discovery was to gather information that could offer some insight on how LEAP@WPI/QCC can acquire more membership partners, as well as what those partners would expect and require from LEAP. By reaching out to companies and societies and looking to contact individuals within those societies and companies of interest, we can pursue information which will hopefully satisfy the customer discovery objective we had anticipated to accomplish.

Our Approach to Customer Discovery

In order for us to work towards accomplishing our objective, we required an approach that allowed us to reach out to electronics, optics, and photonics companies and societies in the New England area and gather data on their current pain points and other needs associated with the services outlined in the IAP. The goal was relatively simple - we would try to make contact with individuals within companies whom we felt could give us some insightful information regarding our objective and set up interviews with these individuals to get that information.

Initially, the IAP was created with minimal direct customer input, though the tiered model resulted from hypothetical customer archetypes. Our primary responsibility in selecting our interviewees was to evaluate whether the services provided in each tier, firstly, served a sizable population (that the hypothetical archetype actually represents a considerable customer base) and secondly, addressed the specific needs that those archetypal customers most need.

In order to inform our method for planning, organizing, and performing our interviews to collect data on prospective members' needs, we participated in the National Science Foundation's Innovation Corps (I-Corps)[12]. This course teaches customer discovery tactics to people with an idea for a product, service, or technology that has not yet been validated as valuable by its market. Several lessons and strategies from this course were directly applied to our approach to gathering and conducting interviews.

The I-Corps Site Program in which we participated set a goal for thirty interviews over the course of ten weeks. While interviewing, I-Corps participants are encouraged to meet with interviewees face-to-face. In our work, what interviews were not conducted in-person were conducted over the phone, which was posed as a next-best option by Todd Keiller in his September 30, 2019 presentation for the WPI I-Corps Site Program (as opposed to email interviews or surveys).

Interviewees are selected under the I-Corps strategy by defining a hypothesis about the "customer archetype," a person or group of people who can be seen as representative of a wider set of individuals with similar technical backgrounds, professional experiences, or interests in a given market. In our case, customers were selected for their proximity to microelectronics,

[12] https://www.nsf.gov/news/special_reports/i-corps/

optical research, and early developing photonics initiatives in the New England area. In particular, we focused on engineers and project managers who we believed could speak to the technical and hiring needs of projects being conducted in the microelectronics/photonics ecosystem.

Under the I-Corps Model, interviews are not supposed to include "sales pitches" or details about the product or service being researched. Instead, interviewers are encouraged to ask insightful questions that require thoughtful, complex answers about the interviewee's experiences relevant to the market space. Since we were gathering data to validate the Industry Affiliate Program, this criteria for a good interview was particularly important since it focused on listening to the pain points of the interviewee. This provided insightful data about the parts of the program worth highlighting as the most impactful on the needs of prospective members.

Although genuine and thoughtful answers were desired from interviewees, there is also the possibility that they will not elicit every detail about company information, because of the competitive nature of the industry, and a desire to keep their company's unique practices safe. Understanding this as a necessary behavior within a highly competitive and evolving industry, our interviews focused primarily on the interviewees' perceptions of the success of previous strategies and interactions with other companies and institutions that were already public-facing and were unlikely to elicit a defensive response.

Contact With Networking Societies

In order to get a full understanding of all portions of the photonics ecosystem and test some of the assumptions that went into the formation of the Tier 3 membership benefits discussed in the Industry Affiliates Program, we corresponded with the International Microelectronics Assembly and Packaging Society New England Chapter (iMAPS NE[13]) and the New England Section of the Optical Society of America (NES/OSA[14]). While these two societies may not represent the full breadth of all local networking societies and Tier-3 target organizations, their connection to, respectively, microelectronics and optics allowed us to view

[13] http://imapsne.org/
[14] https://www.nesosa.org/

the current attitude toward photonics advancement in its adjacent industries. As societies dedicated to the dissemination and discussion of their respective technologies, each society aims to attract student and industry members who want to present research to other scientists in their field, to cooperate with people in other sections of the industry, and to network with different contacts in companies for future hiring or employment opportunities.

Being involved with the LEAP facility at WPI also gave us the opportunity to host the NES/OSA chapter meeting which was held at the LEAP facility[15]. While attending this NES/OSA chapter meeting, we were able to speak with those who came to tour the facility, and then conduct interviews with some of these individuals. The interviews conducted at this and other society chapter meetings were somewhat informal, but were constructed to allow us to collect data from the interviewees that was consistent with what we gathered from our other approaches. Furthermore, questions asked of society representatives were aimed at understanding the priorities of a networking organization to understand what value a program like the IAP could deliver to them.

Interviewing Format and Process

After successfully making contact with an individual from a particular company, an interview was scheduled and conducted depending on the contacted individual's availability. In order to ensure fruitful discussion directed toward our objectives of understanding the interviewee's pain points, interviews consisted of prepared conversation-starting questions followed by improvised follow-up questions directed at better understanding the specifics.

Three areas of specific interest we had entering the interview process were:

1. How companies currently engage with academic institutions and the quality of those interactions
2. Pain points in hiring and their perception on the current talent landscape

15 https://www.nesosa.org/meetings/2019-2020-calendar/225-november-21-2019-doug-petkie-james-eakin-at-wpi

3. What new technologies or initiatives companies are focusing on to remain relevant as technology progresses

In order to identify each interviewee's perspective on these three topics, a list of General Interview Questions (on pages 15 - 16) was prepared to serve as a jumping-off point for further discussion. Conversations were intentionally allowed to evolve naturally, with the interviewer(s) honing in on specific topics mentioned by the interviewee in an attempt to approach new information from an original angle that may lead to previously unknown pain points or key takeaways.

Example General Interview Questions

Section 1: Questions to identify previous involvements with academic institutions and determine most valuable aspects and pain points.

1. (Mention known interactions of the specific company with academic institutions.) What was the extent of that/those engagement(s)? Are there any other interactions that I have not mentioned?
    a. (If no prior engagements are known or reported,) Can you think of any needs you currently have that such an engagement could at least partially meet?
2. Did any aspect of your engagement experience detract from the overall value? If so, what specifically made your experience more difficult?
    a. Were there any issues specifically relating to intellectual property? To pacing of the interaction (i.e. time taken to solve problems, time taken to communicate/organize, etc.)?
3. What part of your engagement experience did you find most valuable? Were there any specific interactions or services that convinced you the engagement was "worth it"?
4. Of needs or services that you believe COULD be met by an academic institution, are there any you are currently solving via an engagement with another type of organization (e.g. government, partner company, other industry contacts)?
5. Is there anybody else in the organization who has had a similar or even drastically different experience with these kinds of interactions?

Section 2: Questions about current hiring process and difficulties surrounding the current talent pool.

1. Talk to me a little bit about your hiring process, from searching for qualified candidates with the necessary technical knowledge to the kinds of qualities you look for once you have your list of prospective new hires.
2. Would you consider finding qualified employees a pain point for your department or company generally?
    a. If so, would you say that the issue is a matter of quantity or quality of candidates?
    b. Where do you currently search for talent in these positions? How effective are these sources in relationship to each other?
3. Have you ever stopped pursuing a new business direction because you didn't believe the workforce was developed enough in that area? If so, tell me a little bit about the direction and about the information you used to determine that the current workforce was unsatisfactory.

Section 3: Questions about the company's short-term outlook and willingness to participate in new technologies and ventures; emphasis on a company's need to grow and remain relevant.

1. Are there any areas of your current business model that you have identified as particularly important in maintaining your relevance and growth? In other words, what topics, methods, or technologies have you identified as "high potential"?
    a. Have you considered contacting an academic institution for any of your needs in that area? Why or why not?
2. What are some of the new topics, methods, or technologies that you have not yet explored but are watching closely before deciding to get involved?
    a. What are some of the specific reasons you have not yet committed to pursuing or investigating these items?
3. Are there any topics, methods, or technologies which you have identified as a "high competition item" between your company and other companies?

4. Where would you normally go for help solving the kinds of problems you may face while exploring new avenues or working to remain relevant within your industry? Have you thought about going to an academic institution?
5. Are there any customers in your supply chain who I should be talking to about this so I can gain a greater understanding of this ecosystem?

Methodology 2: Web-scraping

To approach the issue of developing a membership model to encourage participation in the LEAP@WPI/QCC facility, one of the methodologies we applied is 'web-scraping.' Web-scraping is a method for data acquisition that allows a user to set certain parameters to acquire information about a subject from the internet. The benefits of web-scraping come from the sheer volume of data available to a user on the internet. Whereas it might take an individual hours or even days to extract hundreds or thousands of points of information from a website to develop a useful database, a well-written web-scraping script can do this in a matter of minutes[16]. Moreover, using web-scraping to acquire data can prove to be more reliable than interviews, though this comes at the cost of loss of nuance that a human interviewee can provide.

Objective: Gain Insight on Photonics/Optics Industry Job Market

One of the ways we believe LEAP@WPI/QCC can encourage greater membership is by developing and advertising certain skills that the photonics industry needs from hires. Furthermore, it is imperative to understand the current 'industry' sphere that photonics occupies in order to better posit LEAP@WPI/QCC to the field. Some of the questions that need answering include:

- What companies are hiring for photonics/optics related jobs?
    - Are any of these companies that we would not expect?
- What locations in the New England area are hotspots for photonics?

16 https://books.google.com/books?hl=en&lr=&id=TYtSDwAAQBAJ&oi=fnd&pg=PT8&dq=web+scraping&ots=y0B3xHspfh&sig=e7vFmj1h-hqauzQ2U2xNrKpWF7I#v=onepage&q=web%20scraping&f=false

- How do other industries like semiconductors and robotics compare to photonics in the area?
- What are the most common skill sets job postings are asking for from applicants?
- What industry spheres do most of these jobs occupy?
    - Eg. research/electrical & electronic manufacturing/defense and space?
    - Hardware vs. software needs?

While interviews with industry insiders are helpful in answering some of these questions, there is a lot more data that can be immediately accessed online. This is why we have adopted web-scraping as one of the primary methodologies. With the help of some Python web-scraping scripts and simple data analysis, we will be able to make fruitful recommendations for LEAP@WPI/QCC.

Our Approach to Web-scraping

To procure the necessary data, we chose to write our web-scraping scripts in Python due to its ease of use and several available modules (libraries of pre-existing code specific to certain tasks/functions) tailored towards web-scraping. We chose to use LinkedIn as the source for our job postings due to our familiarity with it as a job-hosting website and its significance as a social media platform for professionals. With over 660,000,000 users[17], there is an abundance of data to be found.

The image below serves as an example for the amount of potential information we can collect on a website like LinkedIn alone. The keywords ‘Photonics’ and ‘United States’ procure 762 results which would take several days to manually go through and extract relevant information from. Our ‘control’ manual data-scraping from LinkedIn took approximately 2 hours to fill a CSV (comma-separated values) file with around only 35 results. Nonetheless, by using manual scraping as a starting point for our methodology, we were able to identify the type of information we could gather through an automated web-scraping process

[17] https://about.linkedin.com/

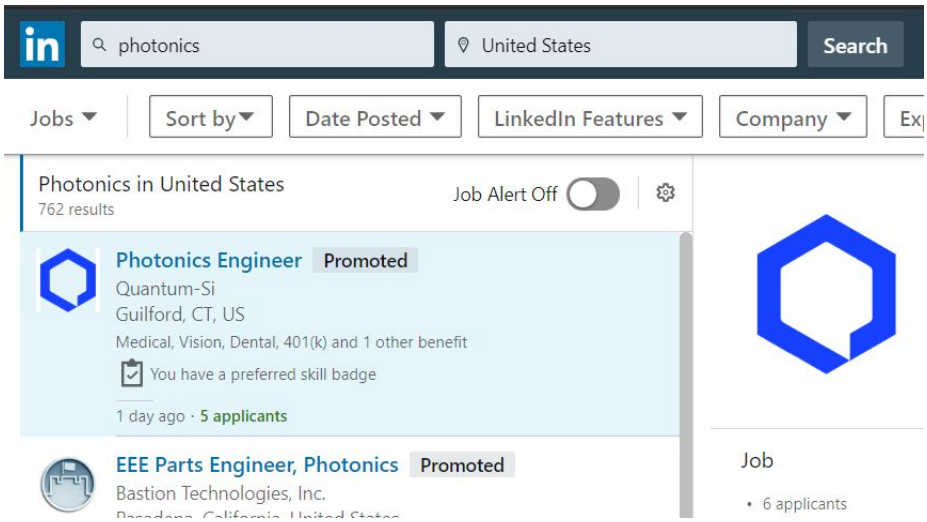

Figure 3. No. of Current Jobs for Photonics in USA

For the purposes of our web-scraping code, we used the browser-automating module 'Selenium' to automate the process of handling the entire web-scraping process. The web-scraper begins by logging into LinkedIn then proceeding to google with a LinkedIn specific search: **site:linkedin.com/jobs and "JOB PARAMETER" and "LOCATION PARAMETER".** Here, 'JOB PARAMETER' and 'LOCATION PARAMETER' are placeholders for the specific search terms we would like to gather data from. For example, we may replace these placeholders with 'Software Engineering' and 'Silicon Valley' to get job postings for LinkedIn pertaining to those parameters. It then goes through a user specified number of Google search results pages and fills a Python array with job links to scrape data from. However, an occasional issue in data collected from our web-scraping method is the acquisition of extraneous data, i.e. data from parameters outside of those specified. This is due to the web-scraper eventually running out of appropriate data to find within the specified search parameters. This is, fortunately, easy to remedy by excluding this data when conducting our data analysis (graphically and otherwise) which we will discuss shortly.

The scraper then loads each page and pulls data from specific parts on the site. The below image serves a visual to help describe where each piece of data comes from:

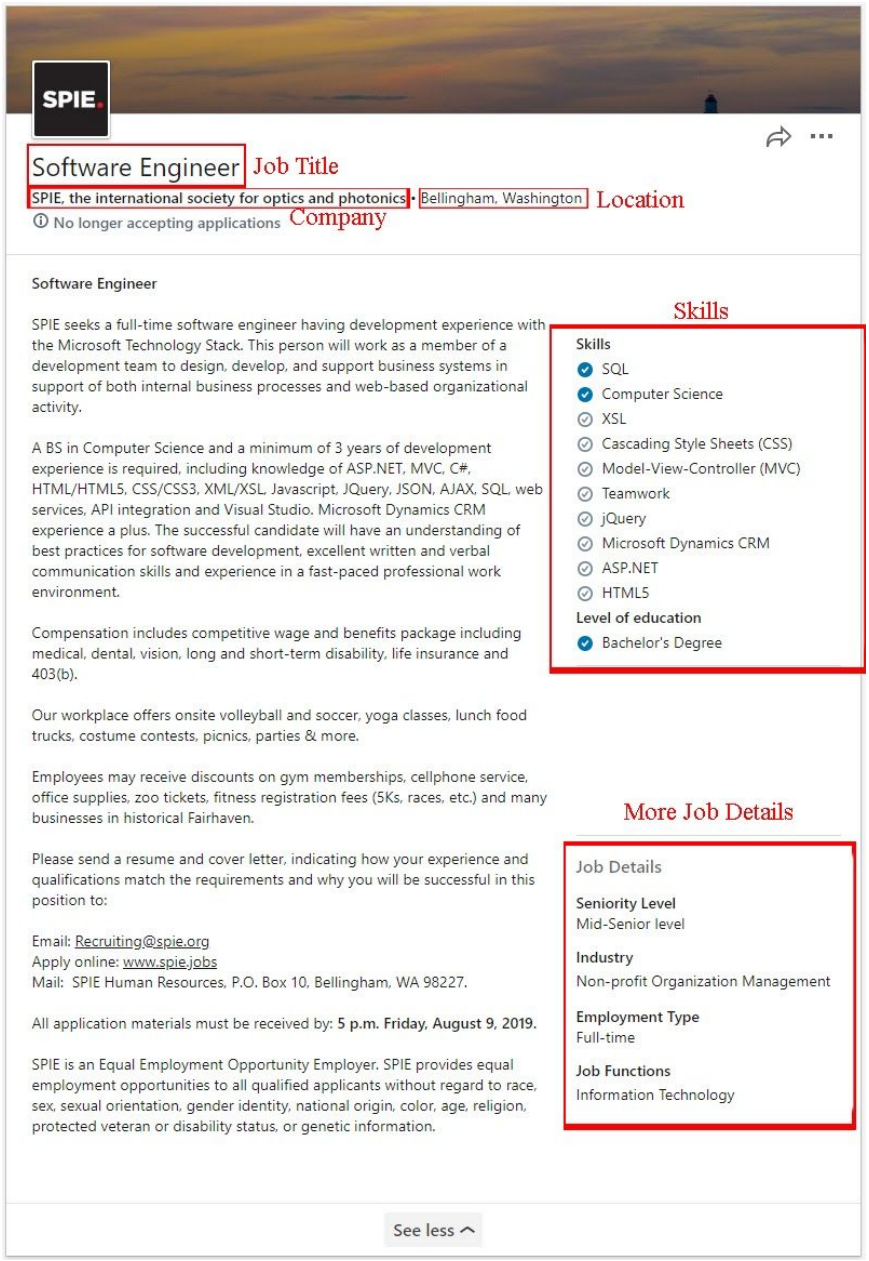

Figure 4. Highlighting Web-Scraping Fields

The web-scraper attempts to find the fields highlighted in red on each LinkedIn job posting based on its xpath[18] (a unique identifier in HTML DOM). If present, it fills out appropriate columns in a CSV (comma separated value, similar to an excel sheet) with the information. For example, the skills for each job go under the column titled 'Skills' and the company's name goes under 'Company.' This is done by using the default Python module 'csv.' Pulling data from the

[18] https://www.w3schools.com/xml/xml_xpath.asp

description through an automated process would be extremely difficult to control for with all the different ways someone may choose to organize their description.

	A	B	C	D	E	F	G	H
1	**Job Name**	**Company**	**Job Type**	**Seniority**	**Industry**	**Skills**	**Location**	**URL**
2	Laser Engi	Analog Phc	Full-time	Mid-Senio	Semicondu	Customer I	Greater Bc	https://wv
3	Opto-Mec	The Photo	Full-time	Mid-Senio	Aviation &	Optical Eng	Greater Bc	https://wv
4	Laser Engi	IPG Photor	Full-time	Mid-Senio	Electrical &	Research, C	Greater Bc	https://wv
5	Laser Scier	IPG Photor	Full-time	Mid-Senio	Electrical &	Research, C	Greater Bc	https://wv

Figure 5. Example of CSV Output from Web-Scraping Script

Not every LinkedIn job posting has each of the highlighted fields filled out which makes it difficult to extract data, but due to the volume of information being found this is a relatively inconsequential problem. We will also manually fill out a 'control' CSV based on similar search parameters to extract data from job descriptions where such data may not be present in the highlighted sections in Figure (insert figure no. here).

Using this web-scraping script, we will be able to find an abundance of information with which we can make beneficial recommendations for LEAP@WPI/QCC. Please refer to Appendix () for the full commented Python script.

Analysing Our Data

After collecting our data, we must analyze it in order to make meaningful recommendations. Our approach to analyzing the scraped data also utilizes a Python script due to the volume and nature of the data we will collect. This second data analysis script uses a number of external modules in order to graph and organize the data. We will be using the 'pandas' module[19] to read our CSVs and to create graphable data frames, the 'numpy' module[20] to handle inconsistencies in our data that may otherwise break the code, and the 'matplotlib' module[21] to graph our findings.

Our data analysis script goes through each column in a CSV and graphs elements out based on certain specifications. For example, for some columns, there may be too many elements

[19] https://pandas.pydata.org/
[20] https://numpy.org/
[21] https://matplotlib.org/

to visually represent, in which case we can specify that we only want graphs of those elements which have a certain number of occurrences or greater. These graphs will help us better understand the data we have collected, though the script also outputs text files with precise numbers in the form of a Python dictionary for further analysis. Please refer to Appendix () for the full commented Python script.

Results and Analysis: Customer Discovery

Our interview data may be used to create recommendations for LEAP's focuses and strategies in its Industry Affiliates Program as well as for methods of contacting target individuals and companies. During and following our interview process, the data (type and quantity of information gathered, interview/contact success rate, etc.) was assessed for quality in order to improve our communication and questioning methods for more high quality data and interviews, and we elaborate on this in detail in the subsection titled "Redesigning Our Approach." The content of each discussion was documented through audio recording or paraphrased transcripts that summarized the content and character of topics discussed. For the written notes, careful attention was paid to listing mentions of keywords and specific topics of interest. During our time trying to find contacts within companies to conduct interviews with, we realized we could alter our approach, based off of the initial results we were getting (lack of interviews through inquiry emails). After completing our interviewing process, we needed to assess the data that we had collected in order to determine how this particular data could help us accomplish our objective. As we conducted our interviews, whether it was free-flowing and face-to-face or over the phone, we documented the answers to each of the questions we asked from our interview template, and also made notes about themes and key points we picked up on from our interviewees. The data collected from our interviews offered us satisfactory data to understand the industry needs and pain points, how LEAP@WPI/QCC could satisfy those needs, and the services that LEAP@WPI/QCC can offer to its members.

Redesigning Our Approach

As we were conducting interviews, we determined which strategies were best in terms of successfully initiating contact with a target company/individual. Our initial approach involved online research into photonics, electronics, and optics companies in the New England area and referring to those companies' contact page and sending inquiry emails regarding the LEAP@WPI/QCC facility and our association with the facility through our IQP. At the same time, we tried meeting with representatives from companies and societies at networking events, and we were able to circumvent the contact information gathering step. In our attempts to

establish contact in these ways, we sent out 11 inquiry emails and heard back from 4 contacts, and met through social meeting events with 12 contacts, which yielded 8 completed interviews. Although we were able to schedule and complete a few interviews using this approach, we needed to redesign our approach in order to successfully make new contacts and schedule interviews with our contacts.

As we moved forward to remodel our approach, we had one idea in mind, which was to try a new way of reaching out to the companies we thought could produce viable contacts. The methods we used to locate companies in the microelectronics, photonics, and optics industries seemed consistent, so the primary focus of this remodeling would be on our technique employed for contact. Rather than sending an inquiry email to these companies, we decided to take a more personal approach and use the company's provided contact in a different way - we would call in to a company phone number which was provided on the company website's contact page. Our inquiry emails did not always receive replies that would lead to further discussion. We hypothesized that the more personal nature of a phone call that requires two-way communication would encourage a more active dialogue with our target contacts.

Once we had altered our method of contact, we saw an improvement in the amount of contacts made, as well as the amount of interviews scheduled and conducted. We called 19 companies, made contact with 7 individuals within the companies, and scheduled and conducted 6 interviews. The number of interviews scheduled and achieved through cold calling was more than the number of interviews we had tried to schedule and collect by submitting inquiry emails. The cold calling interviews were an improvement on the results we had seen, and served as supplements to the contacts and interviews we were making through attendance at social events such as iMAPS and NES/OSA.

(next page)

Customer Needs: Important Focuses for LEAP and its Industry Affiliates Program

Workforce as a Primary Focus

Throughout the interviewing process, a consistent point of frustration was hiring and internship resources. Four of our interviews were conducted with individuals who hold leadership positions within local networking societies. Of these four interviews (Interviewees 1-4), three interviewees focused heavily on seeking out and developing students in terms of hire-ability in order to deliver value to their members in industry. The connection between students who are in search of jobs and companies in search of qualified students and academic researchers is a distinct objective of the societies with whom we spoke. According to the interviewees associated with iMAPS, forging connections with students in order to increase interest amongst a younger crowd is an important aspect of their charter on which they intend to focus in the upcoming years.

One of the most common complaints during our interviews was that hands-on experience and technical expertise are invaluable in the industry and somewhat rare amongst applicants and interns. Of the thirteen interviewees speaking towards their industry perspectives, seven spoke directly about frustrations to do with hiring. Seven interviewees we spoke with were in engineer positions or a position that required a degree and multiple years of experience. In an interview conducted with Mr. Douglas Naylor of Headwall Photonics, Mr. Naylor indicated that "it can be somewhat 'tough' to find help in this field (Optical engineering/photonics) with reliable experience," and that "teaching skills that go along with the theory is very important today." The skills desired by companies in new employees involved practical training with specialized equipment, as well as project and organizational skills. Our initial presumptions about workforce development was that some entry-level positions, such as assemblers, and some more advanced positions, such as technicians, would be the individuals who required more training, but it seems that this concept is also applicable to engineers and perhaps some higher positions as well. Comparing this to the web-scraping data results solidifies this need for a particular set of skills, knowledge, and experience that is desired by companies across different industries. Although a degree earned from an academic university is most certainly valuable to employers, the practical

skills and experience that should be associated with a degree could be even more important in some cases.

The following table captures some recurring themes and topics discussed throughout our interviewing process. We selected the items on the left according to ideas we thought to be relevant and actionable in relation to LEAP's role as an ecosystem facilitator. Since networking societies and companies differ in their goals for this kind of relationship with LEAP, we separated the table accordingly, with some topics considered only relevant to one set of interviews or the other.

(next page)

Interviewee #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Mentioned knowledge/ skills gaps in current talent pool		✓	✓		✓	✓	✓	✓	✓		✓			✓	✓	✓	✓
Mentioned importance of hands-on project experience		✓		✓	✓	✓	✓	✓	✓		✓				✓		✓
Mentioned hireability and student development as a focus	✓	✓	✓														
Mentioned previous connections with academia for employment/ internships						✓				✓		✓	✓		✓		
Generally expressed frustration with hiring					✓	✓	✓				✓	✓			✓	✓	

Table 1. A visualization of the topics mentioned by each of our 17 interviewees. A list of the interviewees, their roles/positions, and their corresponding number is given in Appendix C. Interviewees 1-4 (highlighted) spoke in regard to a networking society while the remainder spoke toward their roles in a company.

Relationships with Academia

Considering that the New England area is a serious source for "talent" and growth for companies to be involved with, we expected that companies would be considering relationships and interactions with academic institutions. These relationships and interactions were expected to be focused on recruiting talent from universities, sponsoring or participating in research or projects, or stimulating interest in a field of an industry. Two particular interviews, conducted with Mr. Edward Ackerman of Photonic Systems Inc.and Edward Freniere of Lambda Research Corporation, provided data on this subject. Mr. Ackerman explained how he participated in seminars regarding traditional (analog) photonics at some universities. These seminars were not specifically designed to recruit students from the universities that the seminars took place at, rather they were held "in order to garner interest in the field of analog photonics and plant the 'seeds' of interest within students." His participation at a seminar "was not for Photonics System Inc. and was focused on hopefully attracting the students attending towards my field of work (photonics)." This type of interaction between a company and academic institution was specifically compelling to us because Mr. Ackerman and Photonics Systems Inc. did not expressly want anything to come out of the seminars that would outright benefit their company, instead they simply wanted to contribute to the photonics industry as a whole by hoping to attract students towards photonics.

Connections to the Industry Affiliates Program

A few of our interviewees gave specific insight into some situations addressed or improved by points in the proposed Industry Affiliates Program. Richard Dumont, Ed Ackerman, and Eugene Sokolov expressed concerns related to applicant specificity.

Richard Dumont, who works with silicon carbide structures and deformable mirrors for Northrop-Grumman, told us about how he and others in his department had to spend a lot of time and resources looking for people with even slightly related experiences. When they did find someone with a somewhat related experience, it took longer to "manage their expectations" for the role than they anticipated. Edward Ackerman reinforced this idea by adding that, as an engineer at Photonics Systems, Inc., he's noticing that the skills of the talent pool is shifting from

hardware- to software-focused, which is leaving a vacuum for hiring in hardware positions. These interviews showed us the value that is added by a candidate or new hire with specific experience in a cutting-edge application.

Alternatively, Eugene Sokolov from VPI Photonics told us that when some companies are unable to find candidates with enough related experience, they abandon hiring and search within their current roster for someone who can cover the role. Although we don't have specific data concerning the methods they utilized before turning within, this is a specific instance in which whatever methods connecting VPI Photonics to its required candidates were insufficient and had a direct impact on the expansion of their work or headcount.

Another portion of our interviews that we recognized were companies whose only "interaction" with universities was in hiring graduates. For example, Michael Foley mentioned that WPI students have a reputation of being "hard-working and intuitive," but he said that his company had no other connections to universities. This is an important customer segment that will need to be addressed in the final IAP to expand the customer base.

How LEAP@WPI/QCC Can Use The Analysis of This Data

Following our analysis of the interview data that we collected, we were able to connect common keywords and concepts in the answers we found from interviews with different individuals in different companies, and sometimes different industries altogether. This commonality between interviewees of concepts discussed, specific needs and desires specified, and workforce topics only served to solidify our understanding of how LEAP@WPI/QCC can appropriately make use of our collected data. As we noticed how important our interviewees indicated that hiring individuals with honed technical skill sets and experience was, we began to question why this was, and how LEAP@WPI/QCC can assess this particular situation and step in to improve workforce development.

The experience and skills desired by companies could be a result of a company's motive to avoid having to spend both time and resources in order to allow a new hiree to have to be "caught up to speed" by becoming properly acquainted with instruments and procedures. Some of these instruments/equipment that could be implemented in a training program for instrument

training and procedure could include optical electronic measuring systems, die place-and-attach bonder, automatic organic dispenser, gold (wire) ball bonder, vacuum oven, plasma cleaner, X-ray and CT, and optical microscopes.[22] After careful consideration, we understood that because the photonics industry is still growing in terms of its impact and influence in comparison to well established industries such as the microelectronics industries, there will be a potential disconnect in how available and prepared the industry's workforce will be as the industry continues to grow. This circumstance provides LEAP@WPI/QCC the ability to move towards training a workforce for the continuously changing industry landscape.

In some cases, it could be possible that an applicant for a specific job may have seen on the job posting that there were skills desired for the job, and they may not know that they lack the actual experience utilizing or even developing those skills since they feel as though their corresponding theoretical knowledge satisfies the skill requirement. This appears to be a result of a lack of proper training for individuals to go with their theoretical background, as well as the possibility of being unprepared to enter into a relatively new field such as integrated photonics. LEAP@WPI/QCC could play an immense role in ensuring that candidates are experienced in hands-on application and do not rely too heavily on theoretical understanding while applying for positions that require specific practical skills. The LEAP@WPI/QCC facility could assess the web-scraping and interview data we have collected in an effort to establish programs and training courses for those interested in entering into the photonics industry's workforce. Some of these programs could potentially be focused on training individuals for positions which might not require a four year degree to be hired for, such as technicians, assemblers, and other entry level positions. With training for these positions, companies could potentially be much more satisfied with the state of their workforce, since they could override the need to train the hirees themselves and effectively consume their time, money, and resources doing so.

Training programs could also apply to positions requiring BS or higher degrees, since some positions require prior relative experience and technical skills that might not have been fully developed from studying at an academic university. With a relatively new and still developing field like integrated photonics, it can sometimes be hard to enter into the workforce

[22] https://aimphotonics.academy/education/lab-education-application-prototypes

for the first time without relative experience. LEAP@WPI/QCC could potentially offer students interested in photonics the opportunity to acquire experience working with photonics and the technology associated with it. LEAP@WPI/QCC could base the training programs and courses off of which positions appear to need serious skills and experience, whether they are for entry level positions or higher positions in the engineering, R&D, or management departments. The industry's needs in developing an apt workforce coincides with all levels of positions, and these needs will only serve to push the industry as a whole forward at all levels.

Results and Analysis: Data from Scraping

After developing our web-scraping methodology, we have to put our web-scraping script to use and find data relevant to our goals. Furthermore, once we find all of our data, we must then analyze it to make constructive suggestions for LEAP@WPI/QCC to develop a better membership model that encourages more participation in the programs it offers and to meet the needs of the industry based on results from customer discovery. Through each iteration of data our web-scraping script finds and analyzes, we are able to fine-tune our search parameters and look for more data that would be germane to our project. It is also important that we point out actionable data for LEAP@WPI/QCC and discover if our methodology is effective and may be used in the future for further improvements. This section will be discussing our findings and presenting informed recommendations from these findings. We will primarily be looking at our scraped 'Industry' and 'Skills' data as we believe it is most pertinent to the objective of our project, though a full breakdown of all our collected data is available in the appendix (Appendix B).

Our Findings from Data Scraping

As stated in the 'Approaches and Methods' section, we began our process of data-scraping (and subsequently web-scraping) by manually going through a number of LinkedIn job postings. This allowed us to identify the best fields we could automate the process of data-scraping from (refer to Figure 4.) while also acting as a control to compare to if any of our web-scraped data seems peculiar. Once we have a number of CSVs with relevant information filled out, we will start our process of analysis.

Before we do this, however, we must decide on the parameters we will be using to collect data from. For our manual scrape, we chose 'Photonics' and 'Greater Boston Area', and for our automated web-scrape we chose the following: 'Optics' and 'Greater Boston Area', 'Photonics' and 'Greater Boston Area', and 'Optics' and 'New England'. We also chose the parameters 'Robotics' and 'New England' alongside 'Semiconductor' and 'New England' to serve as comparisons from more mature industries that adopt photonics technologies. Although the focus

of the IQP is on the Optics/Photonics industry, it is important to put the industry in context with some of its closest industry relatives.

The first piece of data we will look at pertains to the industries the photonics and optics spheres occupy. We will begin by first looking at the 'control' manual scrape, then at our web-scraped data. Our presentation of this data will follow roughly this same format throughout.

Industry Findings and Recommendations

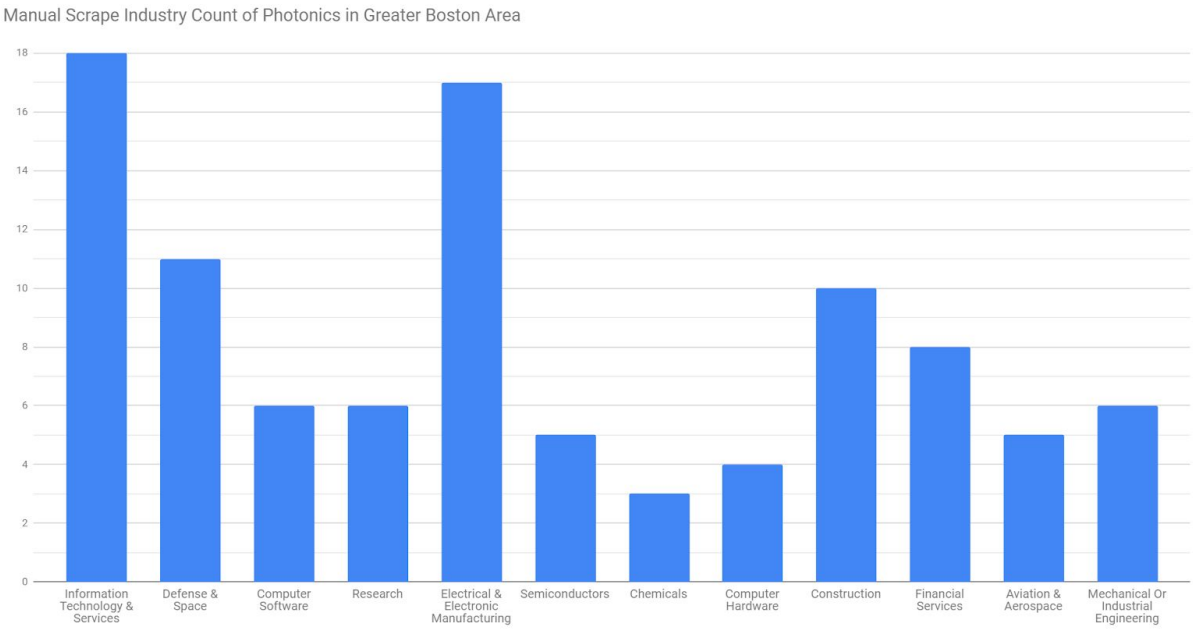

Figure 6. Manual Scrape of Industries Hiring for Photonics in Greater Boston Area

Job Name	Company	Industry	Location
Photonics Engineer	The Charles Stark Draper Laboratory, Inc.	Information Technology & Services, Defense & Space, Computer Software, Engineering, Information Technology	Cambridge, MA
Student Technical Assistant--Electronic-Photonic Test and Evaluation	MIT Lincoln Laboratory	Information Technology & Services, Defense & Space, Research	Lexington, MA
Integrated Photonics Design Engineer	Analog	Electrical & Electronic Manufacturing,	Boston, MA

	Photonics	Semiconductors	
Scientist - Laser Diode Products	IPG Photonics	Electrical & Electronic Manufacturing, Research, Chemicals	North Oxford, MA
Photonics Design Engineer	Lightelligence	Electrical & Electronic Manufacturing, Computer Hardware, Semiconductors	Boston, MA
Graduate Summer Intern - Nanophotonics Researcher	Raytheon Technologies	Construction, Information Technology & Services, Financial Services	Cambridge, MA

Figure 6.1 Sample of Data Found from Manual Scrape (Appendix B for All CSVs)

Looking at the graph above, we see that the three most common industries hiring for Photonics in the Greater Boston Area (that our manual scraping methodology found) are 'Information Technology & Services', 'Defense & Space', and 'Electrical & Electronic Manufacturing.' It is important to note here that some descriptors in the above graph describe different industries in context. 'Information Technology & Services' is an umbrella term for the technology industry, and 'Construction' here can be interchangeable with 'Manufacturing.' If you refer to Appendix B for the full breakdown of the CSV the graph is pulling data from, you will notice that Raytheon Technologies is consistently associated with the 'Construction' industry. The company is not actually involved in the traditional 'Construction' industry but rather with the 'construction and manufacturing of various technologies.' We suspect that some of these inconsistencies in descriptors versus most likely meaning are due to a company's Human Resources team appointing a non-technical employee to post jobs to LinkedIn.

We will continue our analysis by presenting a few more graphs then discussing our findings and making recommendations based on our analyzed data.

(next page)

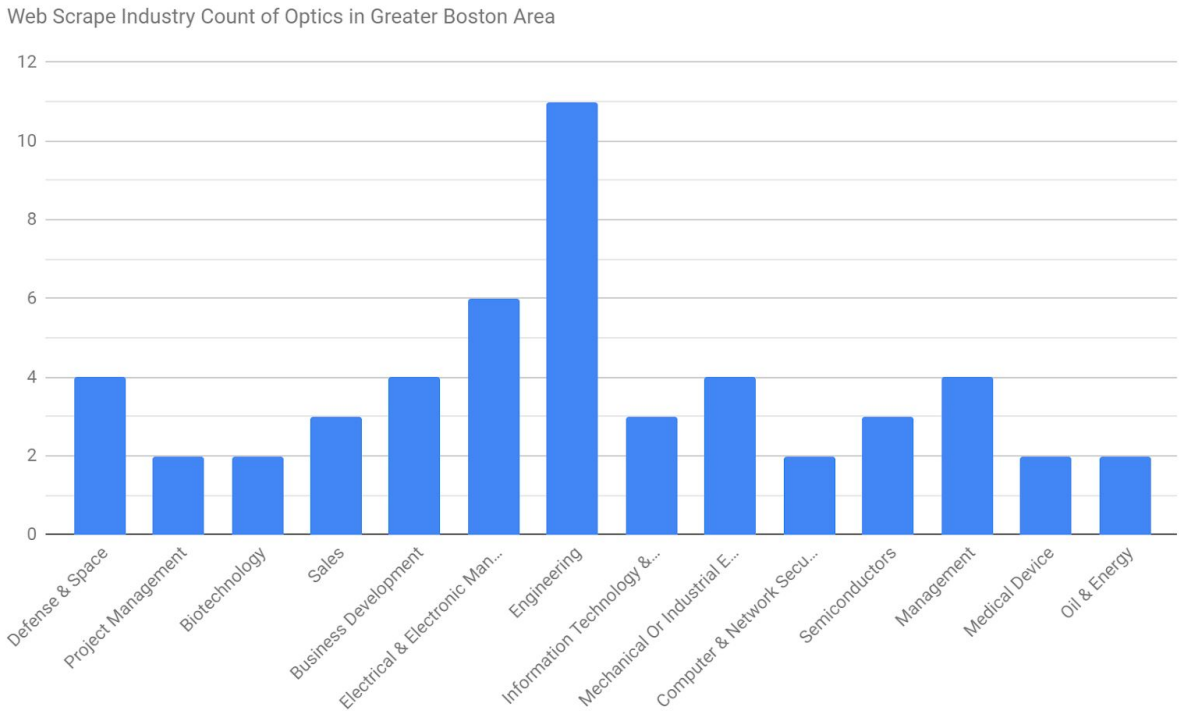

Figure 7. Web Scrape of Industries Hiring for Optics in Greater Boston Area

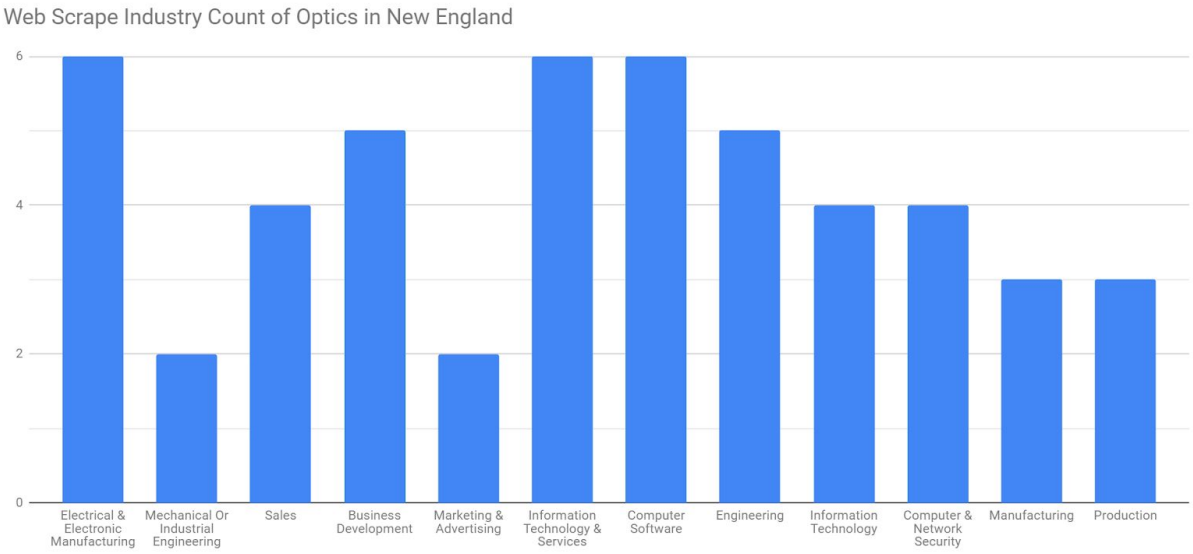

Figure 8. Web Scrape of Industries Hiring for Optics in New England

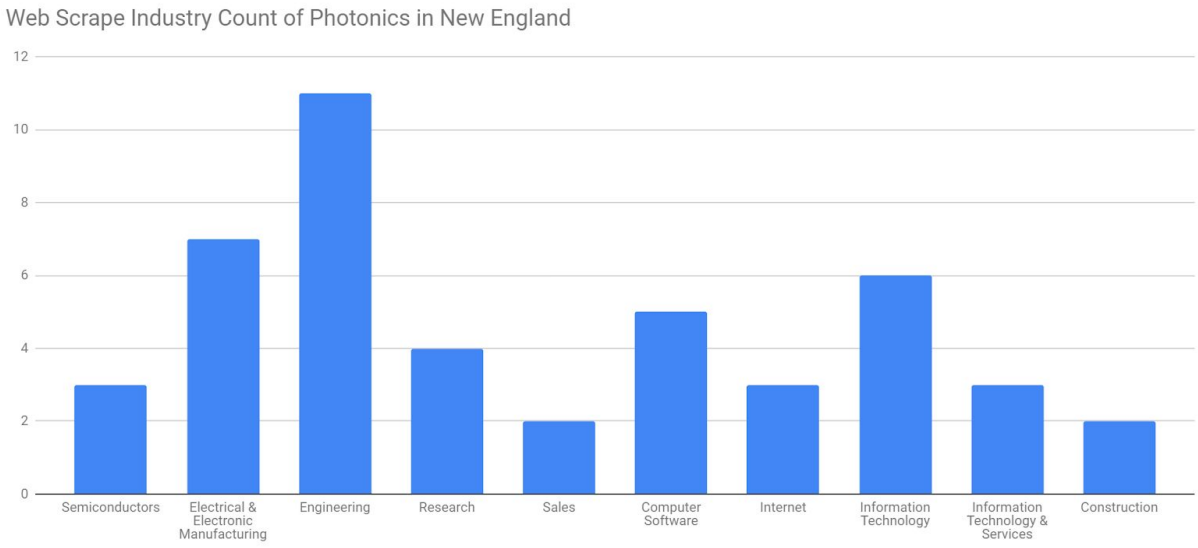

Figure 9. Web Scrape of Industries Hiring for Photonics in New England

After thoroughly analyzing our scraped data (see Appendix A for a full breakdown of industry data), we find that some of the most common industry types listed across job postings include: 'Electrical & Electronic Manufacturing', 'Information Technology & Services', 'Engineering', and 'Computer Software.' Our findings from our web-scraped data are consistent with our manual scrape. Some differences arise due to the web-scraping script pulling data from any job listing related to the search parameters, whereas for our manual scrape we were able to control for the exact information we were looking for.

One of the other issues we encounter when conducting our web-scraping is the acquisition of non-specific/ambiguous descriptors/keywords. For example, in Figure 3 we see that one of the most common industries is 'Engineering' while there is a second descriptor 'Mechanical or Industrial Engineering' which can fall under the engineering umbrella. Similarly, 'Engineering' may be broken up into more specific engineering types. Unfortunately, an automated web-scraping script simply cannot account for all the nuance that can be found in personal interviews. This is why it is important to consider both web-scraping and interviews as methodologies working in conjunction rather than individually, as both have their benefits and limitations.

Nonetheless, we can make some conclusions and recommendations using the information we have found. WPI, being an engineering school, fosters an environment where many students pursue an academic and professional interest in the fields of these industries. Our recommendation for the LEAP@WPI/QCC facility to garner more membership from Photonics/Optics related companies and societies is to integrate and advertise the abundance of hiring potential in the area.

Students at WPI are most commonly hired at companies within the industries we find associated with Photonics/Optics[23]. LEAP@WPI/QCC may serve as a bridge between the Photonics/Optics industry and engineering students interested in the field. In the class of 2019 at WPI, 151 students graduated with a degree in 'Electrical & Computer Engineering'[24]; some of these students may be inclined to work within the Photonics/Optics sphere with more LEAP@WPI/QCC involvement on campus and general exposure to the field.

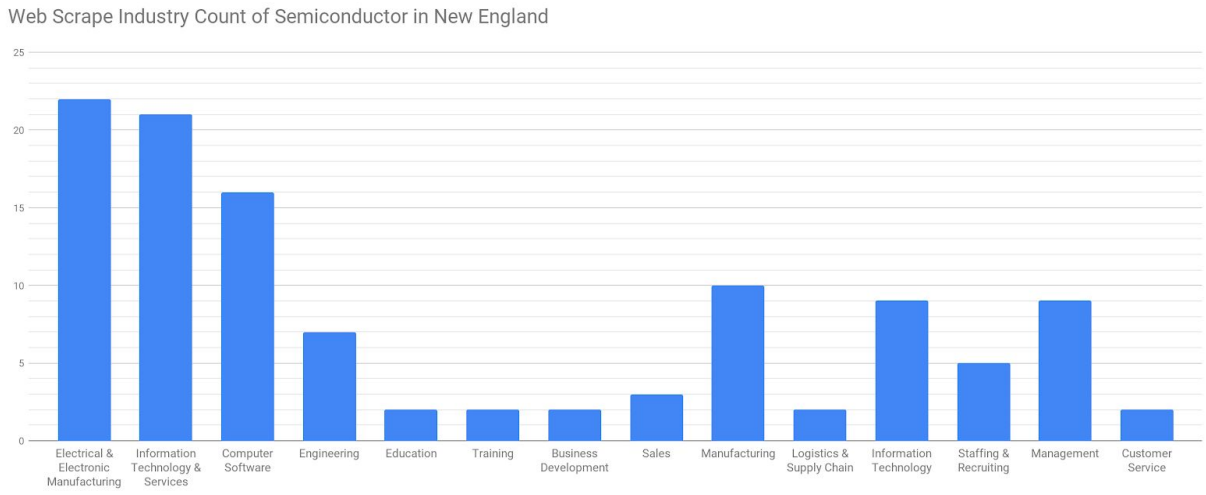

Figure 10. Web Scrape of Industries Hiring for Semiconductor in New England

With a quick glance at the graph for industries hiring for Semiconductors in the Greater Boston Area, we can see the similarities between the semiconductor industry and Photonics/Optics. Electrical & Electronic Manufacturing, Information Technology & Services,

23 https://www.wpi.edu/sites/default/files/inline-image/Offices/Career-Development-Center/CDC_StatReport_2018.pdf

24 https://public.tableau.com/profile/wpi.institutional.research#!/vizhome/DegreesAwarded_15722293635880/Story1

and other industries are prevalent in the hiring market for both. This is likely due to the similarities in the technologies and the integration of photonics tech in the semiconductor space.

Skills Findings and Recommendations

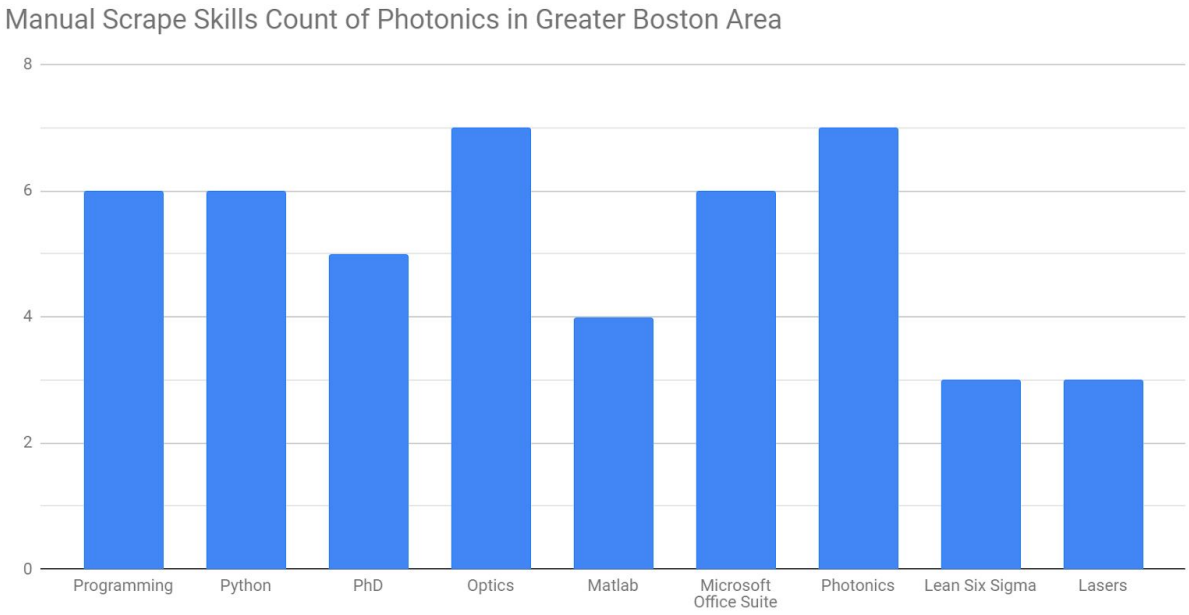

Figure 11. Manual Scrape of Skills for Photonics in Greater Boston Area

The next piece of data we will discuss pertains to the most common skills listed as requirements/qualifications on LinkedIn job postings in the Photonics/Optics industry. Figure 5 is a graphical representation of the skills we found when manually scraping data from LinkedIn with the parameters 'Photonics' and 'Greater Boston Area.' Unsurprisingly, two of the most commonly listed skills are 'Optics' and 'Photonics', though 'Programming', 'Python', and 'Microsoft Office Suite' come in second. Having a solid grasp on Microsoft Office tools is a necessary skill for most jobs, regardless of technicality or niche, which is why it is not imperative that we explore its importance in the field. As we will see in our next few graphs, 'Microsoft Office Suite' is not found to be a common skill in any of our web-scrapes, which points towards this skill being an outlier in our skill search.

As mentioned, scripting and programming skills with languages like ‘Python’ seem to be in demand. In a similar vein, we see ‘Matlab’, another programming language used extensively in engineering disciplines. We will see if these skills are corroborated by our web-scraped data. The next three graphs will present our web-scraped data, after which we will discuss our findings and recommendations.

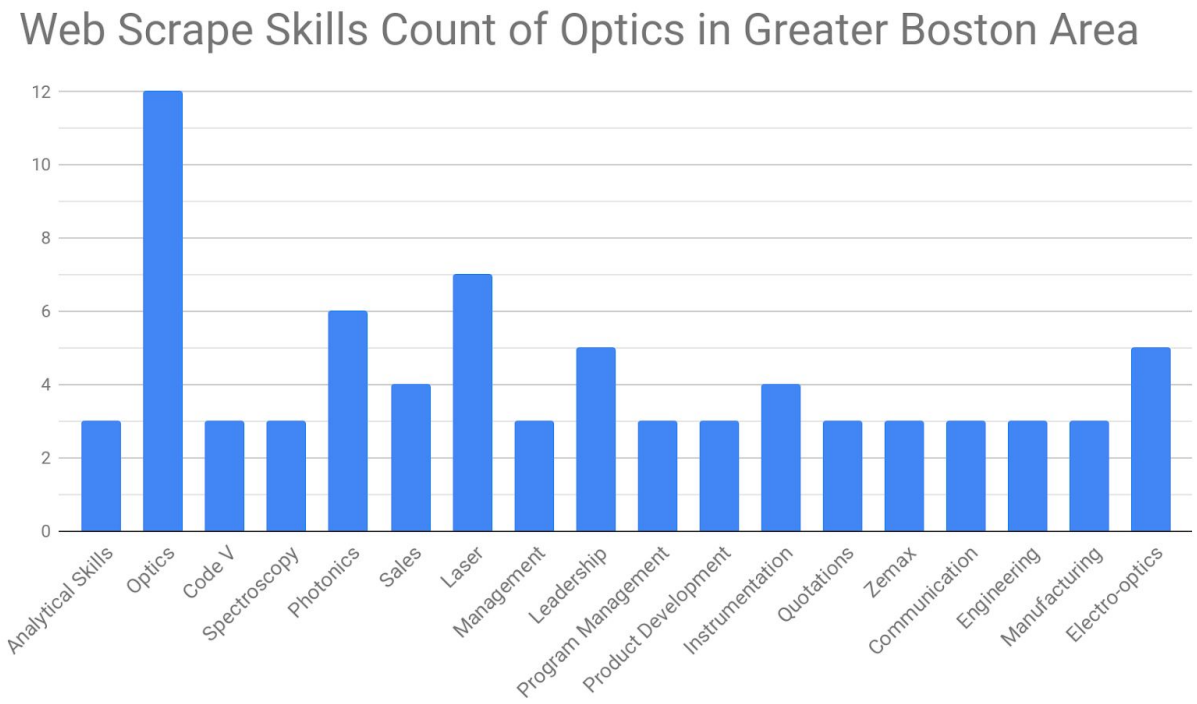

Figure 12. Web Scrape of Skills for Optics in Greater Boston Area

(next page)

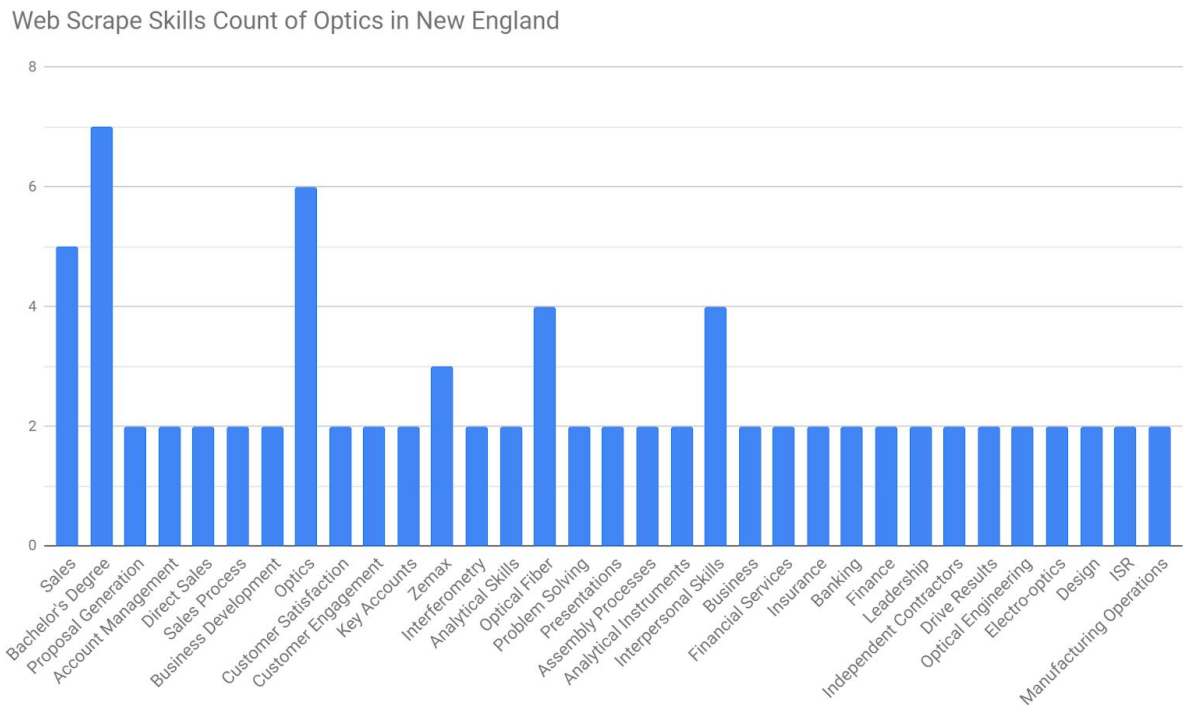

Figure 13. Web Scrape of Skills for Optics in New England

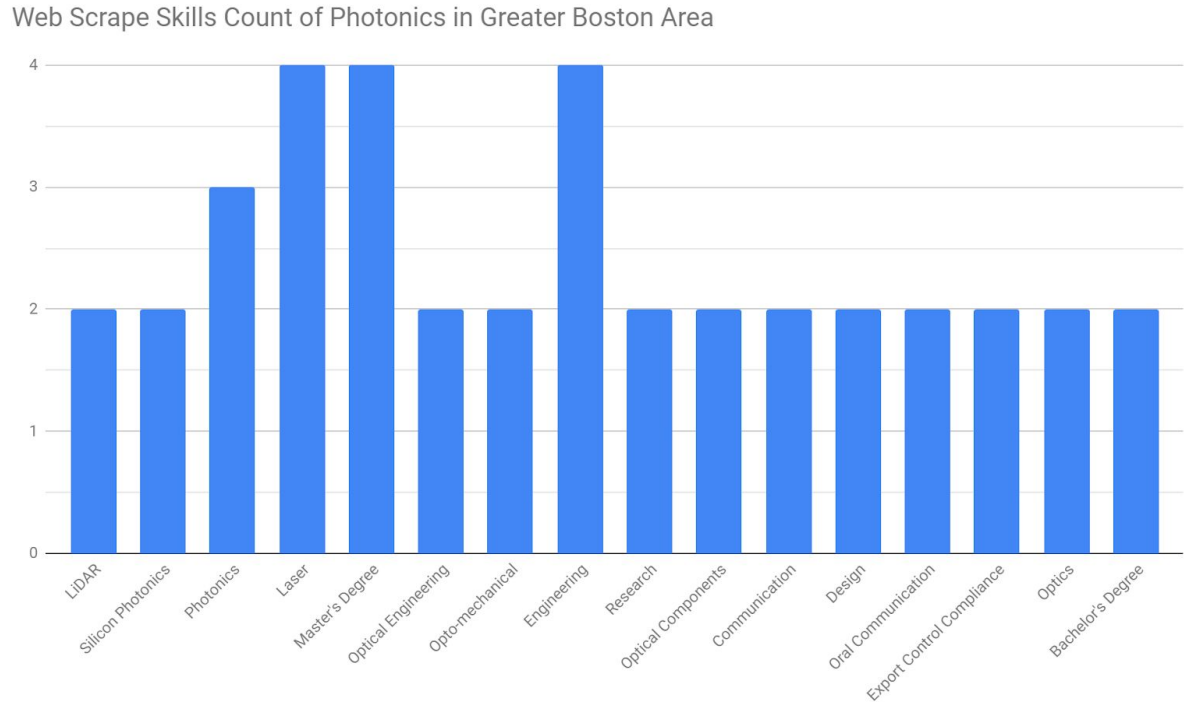

Figure 14. Web Scrape of Skills for Photonics in Greater Boston Area

Through each web-scrape of skills within the various parameters, we see that one of the 'skills' most sought out is a higher education degree. 'Bachelor's Degree', 'Master's Degree', and 'PhD' all show up in significant amounts across every scrape which is expected, considering that Photonics/Optics is a highly specialized field. Moreover, our scrape of skills across the different parameters has procured many more *specific* points of information that can be acted upon. 'Zemax', for example, is a commonly listed skill which likely refers to the software that the company Zemax produces which includes OpticStudio, OpticsBuilder, and OpticsViewer. All three of these are software tools centered around photonics and optics systems described as "leading optical system design software."[25]

LEAP@WPI/QCC can use this skills information to develop programs centered around highly sought out skills that may be beneficial to companies in the Photonics/Optics fields. One of the common complaints found during personal interviews (see Chapter 3 - Customer Discovery Results and Analysis) was that new hires are lacking in the skills and optics experience they were hoping for. This is where LEAP@WPI/QCC may be able to insert itself as a valuable asset, helping companies bridge that gap between new hires and their lack of necessary skills. Using our data and perhaps expanding upon the parameters for new web-scrapes can prove fruitful for developing these programs and the LEAP@WPI/QCC facility.

(next page)

[25] https://www.zemax.com/

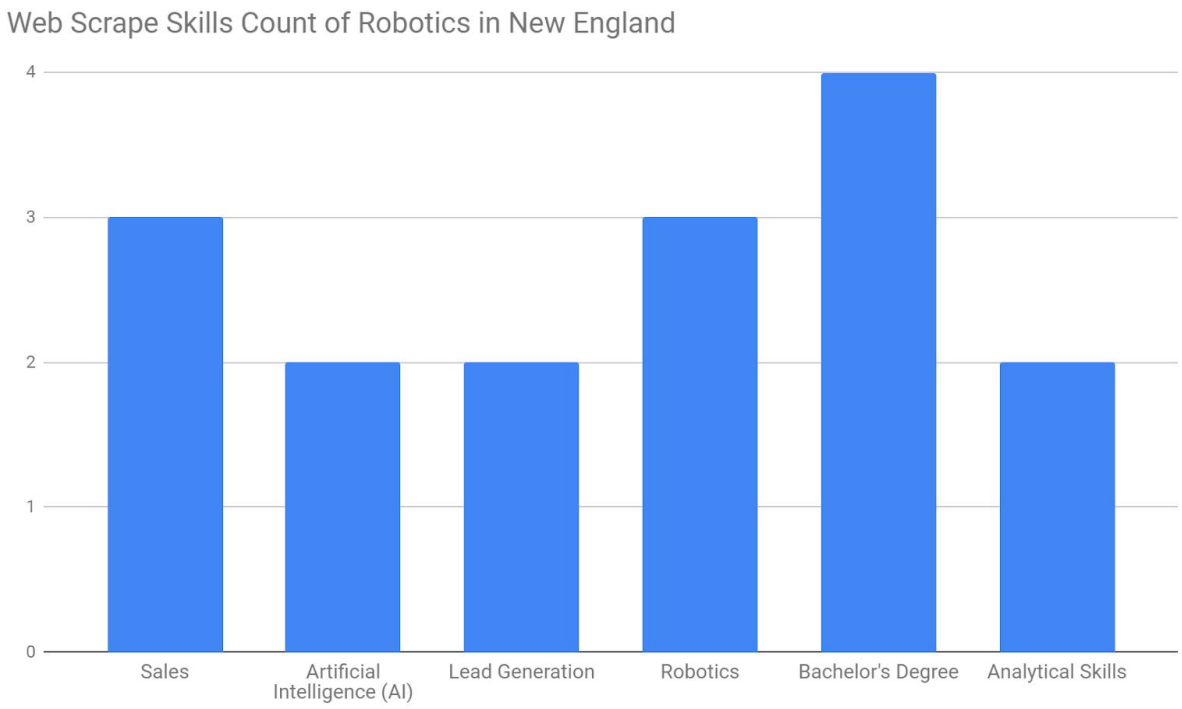

Figure 15. Web Scrape of Skills for Robotics in New England

Our web scrape of skills for Robotics in New England actually serves as an example for why web-scraping may sometimes fall through despite its ability to acquire large amounts of data. Due to the inconsistency of how job postings are presented on LinkedIn, the web-scraping script is not always able to find the fields it is looking for. In this case, out of 33 job postings scraped, the script was only able to pull the skills section from 7 of them. Although this graph's original purpose was to serve as a comparison between the commonly sought skills in Photonics/Optics and Robotics, there simply is not enough actionable data to make informed assumptions from. As previously mentioned, this is why it is important for us to use both web-scraping and interviews in tandem.

Applications to Companies in the Photonics/Optics Industry

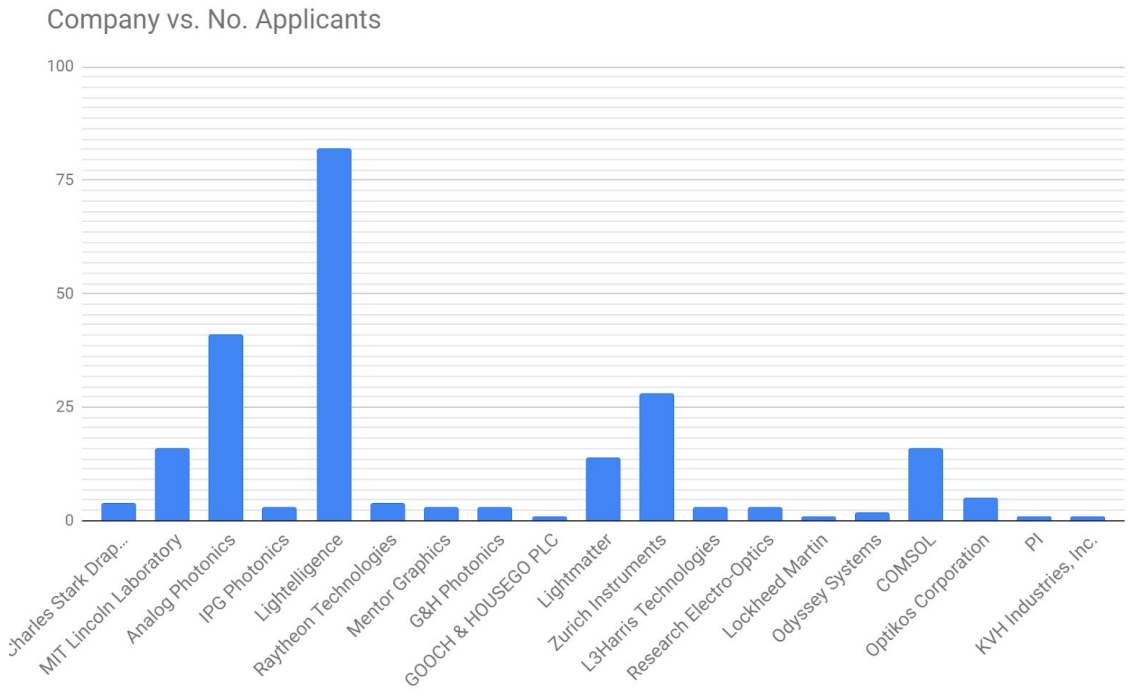

Figure 16. Company vs. No. Applicants from Manual Scrape

We will finally be looking at the total number of applicants to every company in our manual scrape. It is important to ascertain why certain companies like Lightelligence and Analog Photonics seem to get a majority of applicants. The three most popular companies to apply to for Photonics/Optics jobs in our manual scrape of Photonics in the Greater Boston Area show Lightelligence being the most popular, Analog Photonics second, and Zurich Instruments in third. In our manual scrape, there were a few companies with 0 applicants which you can find in the full breakdown of our data in Appendix B.

Examining the specific job postings for Lightelligence, Analog Photonics, and Zurich Instruments, we notice that every job from all three companies in our manual scrape is highly specialized. Each company asks for at least a Masters or PhD with specific knowledge in photonics, nanotechnology, and/or quantum computing. In comparison, job postings from companies like IPG Photonics in our manual scrape asked for an Associate or Bachelor's degree,

with less specialized skills. Our team member, Zachary Langlois, who has experience in the photonics industry working as an intern for IPG Photonics, speculates that low applicant numbers may be due to companies like IPG Photonics looking for applicants with technical and manufacturing backgrounds as opposed to academic/theoretical backgrounds.

Also according to Zachary, IPG Photonics seemed to be very content with its workforce with low turnover and limited recruitment. Moreover, we hypothesize that companies like Lightelligence may be attracting many applicants due to their large funding on account of them being a relatively new start up with solid investment. These two factors could account for the low and high volume of applications, respectively. A form of promotion for LEAP@WPI/QCC could be positioning itself as a repository for potential hires considering one of the Annual Membership Benefits for LEAP is job posting.

Conclusion

Using Customer Discovery Data for Recommendations

What appears to be two of the main issues within the industry that LEAP@WPI/QCC could develop a model to address and eventually solve involve communication/networking and workforce development. In order to attract members to the facility and help the industry's growth and needs, LEAP@WPI/QCC needs to implement a model that can satisfy its potential members and contribute to industry growth.

Our recommendations for LEAP@WPI/QCC upon participating in this project would involve being consistent with communication and outreach to members and potential members, and LEAP@WPI/QCC evolving the workforce development programs for all levels of positions in the photonics industry. Communication from LEAP@WPI/QCC to its affiliates and potential members would be essential in order to allow the facility to ensure the relationship between a member or potential member is going in the best direction possible. Proper communication methods would be useful in order to guarantee that both parties in the relationship are clearly communicating about what is desired on both ends, as well as providing the most beneficial scenario for both parties. Consistent communication with members, companies and societies of interest, and personnel can be established through phone calls, personal email, or organizing meetings and events that allow the facility to maintain connections with its members effectively and keep the companies and societies interested in or involved with the facility engaged and satisfied with their membership. A personal form of contact also provides a better form of marketing for the LEAP@WPI/QCC facility. Personal contacting ensures that communication with interested members, or targeted members, can be established, and this provides an opportunity for various connections to be made effectively. This methodology for marketing can create many quality connections, which hopefully yields success in recruiting members.

By adjusting the workforce development programs to be a reflection of the photonics and integrated photonics industry's needs, LEAP@WPI/QCC can prepare the workforce of the industry in accordance with the skills and experience that are required of all levels of positions.

We believe that the workforce development programs could specifically focus on exposing students involved with projects based on integrated photonics circuits, equipment and tool sets used for integrated photonic circuit analysis, and perhaps course modules that relay specific theoretical background for the operation of different integrated photonic circuits and the equipment and tools being operated. As the industry evolves, the skills and experience with equipment and technology in the field of photonics is going to evolve for many jobs as well, but if LEAP@WPI/QCC can correspondingly adjust and improve workforce development programs for the different levels of jobs, then the industry's growth will be complemented by a well-prepared workforce.

In future projects associated with the LEAP facilities, we would recommend implementing our methodology of using cold calls to make contact with individuals within the photonics industry, in order to gather more data pertaining to industry needs and concerns. It is our recommendation that personal contact, not contact via email, is the most effective form of communication.

Using Our Web-Scraped Data and Recommendations for Future Projects

We believe that there is a lot of useful data to be found for LEAP@WPI/QCC in this section and Appendix B. However, analyzing and going through our scraped data showed us that web-scraping is not a perfect methodology and has its own shortcomings. On its own and without human intervention, it is not possible to make informed assumptions that would prove helpful to LEAP@WPI/QCC. Nevertheless, there is definitely even more data that can be found on a number of factors that our web-scraping code can be adapted towards. We recommend similar future projects to use web-scraping in some capacity, but to consider some of the obstacles we faced when analyzing and collecting our data. Some of these issues may be fixed with more time spent controlling for the variables on sites like LinkedIn, or finding a similarly popular professional platform that is more consistent in how it handles job postings. Moreover, developing a methodology to account for data in long description fields would alleviate many of the issues found in the current implementation. Our web-scraping and data analysis scripts are available in Appendix A for review.

The Importance of Hands-On Experience and Planting the "Seeds of Interest"

One of our IQP group members, Zachary Langlois, has had specific experience with electronics and microelectronics prior to being a student at WPI. Zachary attended Bay Path Regional Vocational Technical High School, a vocational high school in Charlton, Massachusetts, and took up electronics as the vocation he would study and practice during his four years there. This experience and practice with electronics provided Zachary with the opportunity to be prepared to take the next step and work towards a degree in the field of electrical engineering with a considerable amount of practical knowledge on analog and digital electronics. This knowledge was acquired through experience working on projects involving electronics circuits and components. Having the opportunity to work closely with electronics and microelectronics peaked Zachary's interest in the technology and drove him to pursue a career associated with electrical engineering. A similar scenario could become a reality for many children and young adults who could participate in K-12 outreach programs through LEAP@WPI/QCC. We feel as though the impact that these programs could potentially have on many young people is very important, especially considering that these programs could influence those who participate in becoming essential in moving the integrated photonics industry forward.

Zachary also had the opportunity to become a seasonal intern in the electro-mechanical assembly department of IPG Photonics in Oxford, Massachusetts over the course of three summers. This opportunity presented Zachary with the chance to work closely with the photonics industry, and this gave him the desire to work closely with photonics and integrated photonics upon graduating from WPI with a BS in electrical and computer engineering. However, in order to have the opportunity to work as an intern for IPG Photonics, Zachary was hired due to his background with electronics provided from his vocational high school. The skills and experience achieved while in high school had allowed Zachary to enter into this internship prepared to work on projects in the department. As time passed during his internship, Zachary also had the opportunity to sit in as a technician for a fellow employee, who was going on a vacation, which allowed him to further develop his experience and skills. After Zachary relayed his experience to our group, we understood how experience and training in high school and in the field had

presented him with the capability to develop skills he felt pushed him to pursue a career with photonics.

A Brief Example Related to the Importance of Practical Experience

The ability to have practical experience and training prior to entering into the industry will surely be invaluable. Hands-on training and development provides an individual with the ability to comfortably move into the industry and act on both theoretical knowledge and experience they now have. For instance, in the field of electrical engineering, there is a piece of equipment commonly used in circuit analysis called the oscilloscope. Today, students studying electrical engineering have the ability to simulate electric circuits in simulation software packages, and upon simulating an electric circuit they can also analyze the circuit using a virtual oscilloscope. As convenient as this is, this disconnects the student from being able to work hands-on with a piece of equipment that is essential to their discipline. Although the student can use the oscilloscope in simulations, they might have difficulty actually operating and using the oscilloscope effectively, and even understanding conceptually how the physical oscilloscope can be used. Unfortunately, this circumstance could prove troubling to a student in this position if they enter the industry and have trouble operating an oscilloscope. This example serves as a testament to how LEAP@WPI/QCC can present individuals with training for concepts, equipment, and skills associated with integrated photonics.

Although the oscilloscope example regarding practical experience was focused on the field of electrical engineering, practical experience is crucial for essentially any field of science. It is our group's belief that simulation software can be helpful in introducing concepts and knowledge, but these concepts should be reinforced using practical exercises and training in order to develop skillsets that coincide with knowledge and concepts established through simulations. AIM Photonics is currently developing a virtual laboratory platform that will enable community college, undergraduate, and graduate students, as well as industry professionals, the opportunity to examine integrated photonic circuits from a component level all the way up to a system level.[26] This virtual laboratory experience provides students and professionals the

[26] https://aimphotonics.academy/education/student-resources/virtual-lab

opportunity to become better acquainted with the relatively new technology of integrated photonics, but we hope that in educational or professional programs that this platform is utilized, there are also exercises and practices that can be used in order to solidify skills and knowledge introduced in AIM Photonics' virtual experience.

Appendix A

Data Scraping Scripts and Code

Data Scraping Script:

```
#Author: Hamayel Qureshi
#necessary imports for data scraping
import csv
from parsel import Selector
from time import sleep
from selenium import webdriver
from selenium.webdriver.common.keys import Keys

jobtitle = "Semiconductor" #job title scraping for
location = "New England" #location scraping for
numPages = 8 #number of google pages you would like to scrape, by default each google page has 10
results

writer = csv.writer(open('Semiconductor_New_England.csv', 'w+', encoding='utf-8-sig',
newline='')) #enter desired title for CSV
writer.writerow(['Job Name', 'Company', 'Job Type', 'Seniority', 'Industry', 'Skills',
'Location', 'URL']) #enter desired parameters for CSV

driver = webdriver.Chrome('') #enter path to chrome driver here
driver.get('https://www.linkedin.com/login?fromSignIn=true&trk=guest_homepage-basic_nav-header-si
gnin') #direct link to login page
sleep(1)
username = driver.find_element_by_name("session_key")
sleep(0.5)
username.click()
username.send_keys('') #enter LinkedIn login email here
sleep(0.5)

password = driver.find_element_by_name('session_password')
sleep(0.5)
username.click()
password.send_keys('') #enter LinkedIn login password here
password.send_keys(Keys.RETURN)
sleep(0.5)

#Using Google as a search engine to search for a number of links from LinkedIn relevant ot the
scrape
for pgNo in range(numPages):
    driver.get("https://www.google.com/search?q=site:linkedin.com/jobs+AND+%22" + jobtitle +
               "%22+AND+%22" + location +
"%22&rlz=1C1CHBD_enUS734US734&sxsrf=ALeKk03dKvBjVSyXOZ8F3cqxrBUsUPoIxg:1586741380828&ei=hMCTXp-NM
oWqytMPldiquA0&start=" + str(pgNo) +
               "0&sa=N&ved=2ahUKEwif3-qioOToAhUFlXIEHRWsCtcQ8tMDegQICxAt")
```

```
    sleep(0.5)
    urls = driver.find_elements_by_xpath('//*[@class = "r"]/a[@href]')
    urls = [url.get_attribute('href') for url in urls]
    sleep(0.5)

    for url in urls:
        driver.get(url)
        sleep(2)

        sel = Selector(text = driver.page_source)

        #looking for Job Title
        try:
            name = sel.xpath('//*[@class = "jobs-top-card__job-title
t-24"]/text()').extract_first().split() #job name done
            name = ' '.join(name)
        except:
            name = ''
            continue

        #looking for Company Name
        try:
            company = sel.xpath('//*[@class = "jobs-top-card__company-url
ember-view"]/text()').extract_first().split() #company name done
            company = ' '.join(company)
        except:
            company = ''
            pass

        #Looking for Job Location
        try:
            location = ' '.join(sel.xpath('//*[@class =
"jobs-top-card__bullet"]/text()').extract_first().split()) #location done
        except:
            try:
                location = ' '.join(sel.xpath('//*[@class = "jobs-top-card__exact-location
t-black--light link-without-visited-state"]/text()').extract_first().split())
            except:
                location = ''
                pass

        #Looking for Job Type (Full Time/Part Time/etc.)
        try:
            jobType = sel.xpath('//*[@class = "jobs-box__body
js-formatted-employment-status-body"]/text()').extract_first().split()
            jobType = ' '.join(jobType)
        except:
            jobType = ''
            pass
```

```
        #Looking for Job Seniority level
        try:
            seniority = sel.xpath('//*[@class = "jobs-box__body
js-formatted-exp-body"]/text()').extract_first().split()
            seniority = ' '.join(seniority)
        except:
            seniority = ''
            pass

        #Looking for Industries the job/company is involved in
        try:
            industry = ''
            ilist = []
            industries = driver.find_elements_by_xpath('//*[@class = "jobs-box__list-item
jobs-description-details__list-item"]')
            try:
                for el in industries:
                    temp = el.get_attribute("innerHTML")
                    temp = temp.strip('\n').strip()
                    ilist.append(temp)
                    industry = ', '.join(str(x) for x in ilist)
            except:
                industry = ''
        except:
            industry = ''
            pass

        #Looking for skills requested for job
        try:
            skills = ''
            slist = []
            skillEls = driver.find_elements_by_xpath('//*[@class = "jobs-ppc-criteria__value t-14
t-black t-normal ml2 block"]')
            try:
                for el in skillEls:
                    temp = el.get_attribute("innerHTML")
                    temp = temp.strip('\n').strip()
                    slist.append(temp)
                    skills = ', '.join(str(x) for x in slist)
            except:
                skills = ''
                pass
        except:
            skills = ''
            pass

        url = driver.current_url

        #Printing found data to terminal for confirmation
        print('\n')
```

```
        print('Job Name: ', name)
        print('Company: ', company)
        print('Job Type: ', jobType)
        print('Seniority: ', seniority)
        print('Industry/ies: ', industry)
        print('Skills: ', skills)
        print('Location: ', location)
        print('URL: ', url)
        print('\n')

        #Writing to CSV
        writer.writerow([name,
                    company,
                    jobType,
                    seniority,
                    industry,
                    skills,
                    location,
                    url])

driver.quit()
```

Graphical Data Analysis Script:

```
#this script pulls data from CSVs and plots it out
#for the final report, the results and analysis section contains graphs created
#using Google Sheets instead, however this is a useful preliminary tool
import pandas as pd
import matplotlib.pyplot as plt
from collections import Counter
import numpy as np
import csv

def isNaN(x):
    return (x is np.nan or x != x)

filename = 'Robotics_New_England.csv'
indparam = 'Semiconductor'
locparam = 'New England'

#INDUSTRY SCRAPE

industry = pd.read_csv(filename)
industryNew = industry['Industry']
print(industryNew.head())

industryArray = []

for row in industryNew:
    if row and not isinstance(row, float):
        if('  ' in row):
            newRow = row.replace('\n', '').split('  ')
        elif('&amp;' in row):
            newRow = row.replace('&amp;', '&').split(', ')
        else:
            newRow = row.replace('\n', '').split(', ')
        for el in newRow:
            industryArray.append(el)

industryCount = dict(Counter(industryArray))
print(industryCount)

industryCountPlt = {key:val for key, val in industryCount.items() if val > 2}

print(industryCount)
indFig = plt.figure()
indFig.suptitle('Web Scrape of Industries for ' + indparam + ' in ' + locparam)
plt.locator_params(integer=True, tight=False)
plt.bar(industryCountPlt.keys(), industryCountPlt.values())
plt.show()

# #SKILLS SCRAPE
```

```
skills = pd.read_csv(filename)
skillsNew = skills['Skills']

skillsArray = []

for row in skillsNew:
    if row and not isNaN(row):
        newRow = str(row).split(', ')
        for el in newRow:
            if('&amp;' in el):
                el = el.replace('&amp;', '&')
                skillsArray.append(el)
            else:
                skillsArray.append(el)

skillsCount = dict(Counter(skillsArray))

skillsCountPlt = {key:val for key, val in skillsCount.items() if val > 2}

print(skillsCount)
indFig = plt.figure()
indFig.suptitle('Web Scrape of Skills for ' + indparam + ' in ' + locparam)
plt.locator_params(integer=True, tight=False)
plt.bar(skillsCountPlt.keys(), skillsCountPlt.values())
plt.show()

#LOCATION SCRAPE

location = pd.read_csv(filename)
locationNew = location['Location']

locationArray = []

for row in locationNew:
    if not isNaN(row):
        locationArray.append(row)

locationCount = dict(Counter(locationArray))

locationCountPlt = {key:val for key, val in locationCount.items() if val >= 2}

print(locationCount)
indFig = plt.figure()
indFig.suptitle('Web Scrape of Location for ' + indparam + ' in ' + locparam)
plt.locator_params(integer=True, tight=False)
plt.bar(locationCountPlt.keys(), locationCountPlt.values())
plt.show()

#COMPANY SCRAPE
```

```
company = pd.read_csv(filename)
companyNew = company['Company']

companyArray = []

for row in companyNew:
    if not isNaN(row):
        companyArray.append(row)

companyCount = dict(Counter(companyArray))

companyCountPlt = {key:val for key, val in companyCount.items() if val > 1}

print(companyCount)
indFig = plt.figure()
indFig.suptitle('Web Scrape of Companies for ' + indparam + ' in ' + locparam)
plt.locator_params(integer=True, tight=False)
plt.bar(companyCountPlt.keys(), companyCountPlt.values())
plt.show()

#SENIORITY SCRAPE

seniority = pd.read_csv(filename)
seniorityNew = seniority['Seniority']

seniorityArray = []

for row in seniorityNew:
    if not isNaN(row):
        seniorityArray.append(row)

seniorityCount = dict(Counter(seniorityArray))

seniorityCountPlt = {key:val for key, val in seniorityCount.items() if val > 0}

print(seniorityCount)
indFig = plt.figure()
indFig.suptitle('Web Scrape of Seniority for ' + indparam + ' in ' + locparam)
plt.locator_params(integer=True, tight=False)
plt.bar(seniorityCountPlt.keys(), seniorityCountPlt.values())
plt.show()

COMPANY VS NO. APPLICANTS

compapp = pd.read_csv('manual_scrape_photonics_gba.csv')
compappNew = compapp[['Company', 'No. Applicants']]
# print(compappNew.head())

df = pd.DataFrame(data = compappNew)
```

```
newDict = {}

for index, row in df.iterrows():
    if row['Company'] not in newDict.keys():
        newDict[row['Company']] = row['No. Applicants']
    elif row['Company'] in newDict.keys():
        newDict[row['Company']] = newDict.get(row['Company']) + row['No. Applicants']
    else:
        continue

compappPlt = {key:val for key, val in newDict.items() if val > 2}

# print(compappPlt)
compappFig = plt.figure()
compappFig.suptitle('Manual Scrape of Companies vs. No. of Applicants for Photonics in Greater
Boston Area')
plt.locator_params(integer=True, tight=False)
plt.bar(compappPlt.keys(), compappPlt.values())
plt.show()

print(newDict)
```

Python Dictionary to CSV Converter:

```
#this script takes python dictionaries output from the graphical data analysis script
#and converts them into separate CSVs for easier graphing using google sheets
import csv

companydict = {} #fill with your desired python dictionary

with open('companycount.csv', 'w') as f:
    w = csv.writer(f)
    w.writerow(companydict.keys())
    w.writerow(companydict.values())
```

Appendix B

Data Scraping CSVs and Graphs

The complete CSVs and graphs created using our scripts can be found and downloaded at the following link:

https://www.dropbox.com/sh/5mbghkdrbcj7rhe/AAAOZ0gB7cyukgEaSGzyETyya?dl=0

Appendix C

Customer Discovery: List of Interviewees

1) Matthew Bracy, iMAPS (Vice President)
2) Amaresh Mahapatra, Linden Photonics, iMAPS (Education Chair)
3) Dipak Sengupta, iMAPS (Education Program Committee Chair)
4) Harvey Smith, iMAPS (Publicity Chair)
5) Edward Ackerman, Photonics Systems Inc. (Engineer)
6) Joshua Brown, MIT Lincoln Labs (Engineer)
7) Richard Dumont, Northrop Grumman (Engineering Manager)
8) Daniel Farley, AFL Global (Engineering Director)
9) Michael Foley, Incom (Engineer)
10) Edward Freniere, Lambda Research Corporation (President)
11) George Hanlan, SCHOTT (Engineer)
12) Franklin Leard, Seurat Tech (Engineer)
13) David Lees, Collins Aerospace, NES/OSA (Engineer, Program Committee Chair)
14) Jon Medernach, MRSI Systems (Key Account Sales Engineer)
15) Douglas Naylor, Headwall Photonics (Engineer)
16) Eugene Sokolov, VPI Photonics (Software Developer)
17) Niels Wijnaendts van Resandt, Heidelberg Instruments (Sales Director)