

Using Industry Demand to Inform Curriculum Decisions: A Preliminary Report

ANGELA BERARDINELLI, Mercyhurst University, USA
ELIZABETH BROSS, Mercyhurst University, USA

Abstract goes here.

Additional Key Words and Phrases: Curriculum design, text mining, job market

ACM Reference Format:

Angela Berardinelli and Elizabeth Bross. 2018. Using Industry Demand to Inform Curriculum Decisions: A Preliminary Report. 1, 1 (March 2018), 5 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

In the always-changing field of technology, it is important for educators to stay informed about the needs of industry partners in terms of skills, both technical and non-technical, so that we may provide the best preparation possible for our students' lives after graduation. Many programs do this by maintaining direct communication lines with industry partners, which helps give graduates an "in" at those companies since their staff had input into the curriculum. The drawback is that this gives a micro-view of what a few specific companies desire.

The researchers in this project were interested in analyzing larger trends in recent industry demand. By mining the job descriptions in a large dataset, we were able to rank specific technical and nontechnical skills according to their demand. Initial results indicate that job posting data, at least in the geographic areas and time periods studied, lines up with some available data about what developers report using on the job. For example, the top three languages in this study match up with the top three languages reported in the StackOverflow Developer's Survey over the same time periods.

The data source used in this study was provided by a global job listing site, LinkUp. This site indexes over 30,000 companies' postings daily and has hosted over 100 million job listings since 2007. We were able to obtain snapshots of their complete dataset as of June 2017 and January 2018 for analysis. With such a large underlying dataset, we've only just begun to learn what we can from this data. The first questions the researchers sought to tackle where:

- (1) What programming languages, frameworks, and other technologies are in demand for computing careers near Erie, PA right now?

Authors' addresses: Angela Berardinelli, Mercyhurst University, 501 East 38th Street, Erie, PA, 16546, USA, aberardinelli@mercyhurst.edu; Elizabeth Bross, Mercyhurst University, Erie, PA, USA, ebross@mercyhurst.edu.

© 2018 Copyright held by the owner/author(s).
This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in , <https://doi.org/10.1145/nnnnnnn.nnnnnnn>.

- (2) How has the demand for some languages, frameworks, and technologies changed over time? Which are increasing in employer demand? Which are decreasing in employer demand?
- (3) What "soft" or "non-technical" skills are in highest demand for computing careers near Erie, PA in recent years?

2 ABOUT THE DATA SOURCE

Need a paragraph or two here about what LinkUp is, what they do, what kind of data they collect.

The initial data feed received from LinkUp, dated June 1, 2017, included basic information (job title, company name, city, state, zip code, country) for over 110.87 million distinct jobs posted between 2007 and 2017. Of these, over 28.82 million included full job descriptions for the corresponding job posts. See Figure 1 for details.

Job metadata was provided in CSV format; job descriptions were provided in XML format. First, the job metadata was imported into an instance of MongoDB on a private, local server. Next, entries in the database were updated to add all available job descriptions.

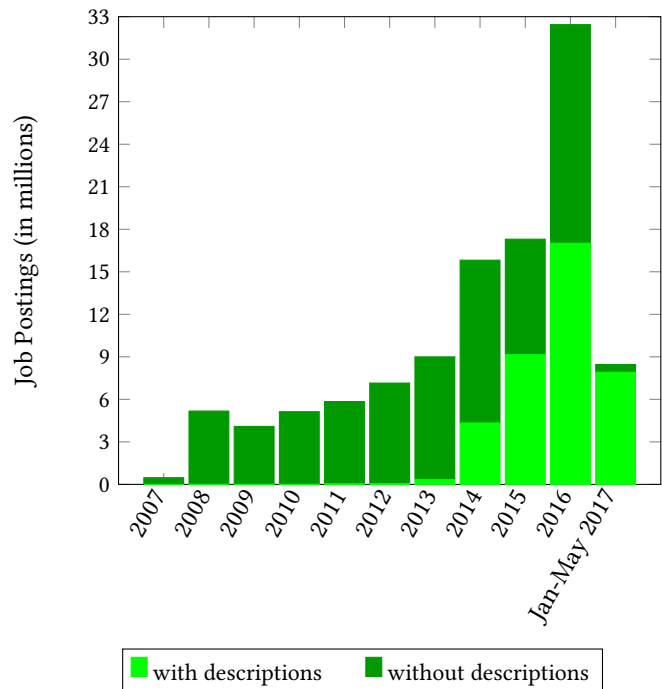


Fig. 1. All job postings included in the initial data upload dated June 1, 2017

Based on lists of computing job titles (sources for this?!) and preliminary exploration of the data set, the researchers compiled a

Terms included		Terms excluded
tech	technical	technician
technology	IT	technologist
UI	software	marketing
web	system	“data entry”
hardware	network	
data	database	

Table 1. Filter terms used for flagging jobs in computing.

list of key words and phrases that could be used to filter out computing job posts. (See Table 1.) This filtering led to approximately 10.96 million computing jobs posted between 2007 and May 2017, approximately 2.56 million of which had job descriptions available. See Figure 2 for details.

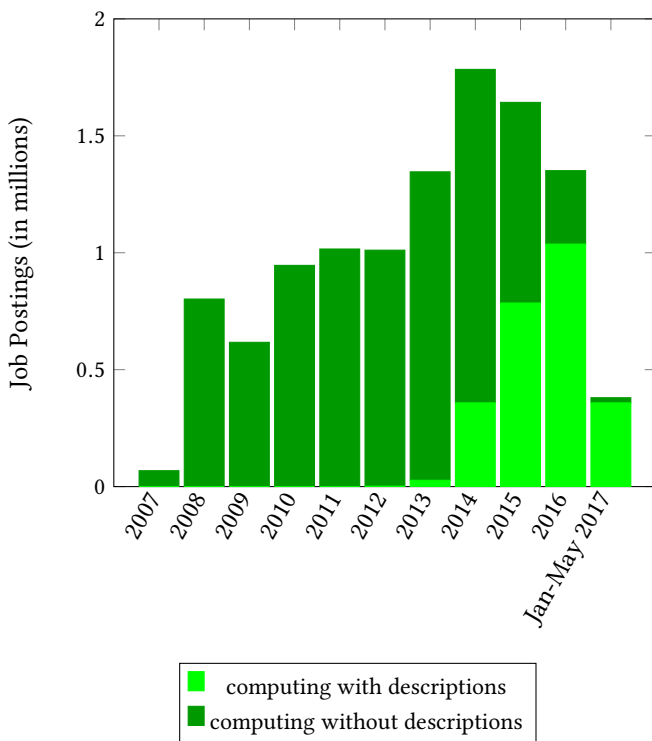


Fig. 2. All computing job postings included in the initial data upload dated June 1, 2017, using the filters in Table 1.

3 DATA PROCESSING AND ANALYSIS

Using the R statistical programming language:

- (1) We connected to the local MongoDB server using the `mongoLite` package and queried for all job postings with job descriptions in PA, NY, or OH.
- (2) We filtered on year, only keeping jobs posted in 2016 and 2017.
- (3) We performed basic text analysis preprocessing on the job description fields using the `tm` package: convert to corpus,

convert to lower case, remove numbers and stop words, remove symbols, remove white space.

Note: Not all standard stop words and symbols were removed. For example, we needed to be able to count instances of the programming languages C, C++, and C#. So the stop word `c` was not removed and the symbols `+` and `#` were not removed.

- (4) Using the `tidytext` package, we converted the corpus to a tidy data frame.
- (5) Using the tidy data frame and the `dplyr` package, we were able to generate four different sets of results:
 - the counts for every term that appeared in job titles and job descriptions
 - the counts for every programming language that appeared in job descriptions
 - the counts for every bigram (two-word phrase) that appeared in job titles and job descriptions
 - the counts for every trigram (three-word phrase) that appeared in job titles and job descriptions

4 RESULTS AND DISCUSSION

After filtering on computing jobs, states (PA, NY, or OH), and year (2016 and 2017), there were 110,469 jobs with descriptions for analysis.

Add some sort of analysis/context here for the results tables.

5 FUTURE WORK

Integrating Updated Source Data. We have received updated flat files from LinkUp dated January 1, 2018. We plan to import these into the database in the near future, before any additional analysis is completed.

Improving Computing vs. Non-Computing Accuracy. We would like to refine our term list for filtering computing jobs. During the initial analysis we confirmed that non-computing jobs were included in the sample and some computing jobs were likely left out.

Trends Over Time. We would like to look at trends over time rather than over a static 2.5 year period.

Enriching Geographic Trend Analysis. We would like to add geospatial information to the database to be able to convert zip code (which exists in the database) to latitude and longitude. This would allow us to use the built-in geospatial querying functionality of MongoDB to do analysis for localized regions and individual metropolitan regions, instead of only filtering on entire states. It may also be useful to compare an analysis of nationwide data to localized data.

Making the Results Accessible. Future plans may include a user-friendly web interface for sharing our results, and expanding beyond just computing to look at trends in other industries.

	Term	Relative Frequency
1	microsoft	1.000
2	sql	0.944
3	java	0.773
4	javascript	0.514
5	html	0.488
6	windows	0.479
7	linux	0.419
8	net	0.394
9	python	0.390
10	css	0.353
11	unix	0.306
12	c++	0.231
13	jquery	0.207
14	asp	0.181
15	hadoop	0.178
16	perl	0.144
17	objective	0.131
18	ruby	0.118
19	google	0.114
20	php	0.106
21	ios	0.098
22	spark	0.097
23	android	0.086
24	node	0.084
25	assembly	0.083
26	scala	0.073
27	mac	0.064
28	matlab	0.057
29	ada	0.053
30	apple	0.049
31	cobol	0.044
32	powershell	0.042
33	swift	0.036

Table 2. Relative Occurrences of Languages, Frameworks, and Systems in Computing Job Descriptions: 2016-2017 in PA, NY, OH (minimum frequency 500). Relative frequency computed with respect to top term.

1	experience	21	related	41	service
2	work	22	position	42	company
3	skills	23	environment	43	project
4	team	24	systems	44	employer
5	business	25	working	45	level
6	years	26	solutions	46	projects
7	management	27	communication	47	process
8	knowledge	28	qualifications	48	customers
9	ability	29	opportunity	49	computer
10	technical	30	provide	50	based
11	support	31	data	51	equal
12	requirements	32	strong	52	system
13	required	33	time	53	tools
14	including	34	design	54	ensure
15	degree	35	responsible	55	applications
16	development	36	preferred	56	develop
17	responsibilities	37	software	57	processes
18	information	38	high	58	status
19	technology	39	customer	59	written
20	services	40	quality	60	bachelor

Table 3. Top 60 terms overall in Computing Job Descriptions: 2016-2017 in PA, NY, OH

	Bi-gram	Relative Frequency		Bi-gram	Relative Frequency
1	communication skills	35.13%	27	full time	12.06%
2	bachelor degree	32.63%	28	duties responsibilities	11.77%
3	project management	27.61%	29	essential functions	11.56%
4	computer science	25.47%	30	products services	11.55%
5	problem solving	21.61%	31	world class	11.09%
6	information technology	21.51%	32	verbal written	10.72%
7	skills ability	21.23%	33	education experience	10.69%
8	team members	21.13%	34	information systems	10.49%
9	software development	20.34%	35	school diploma	10.46%
10	customer service	19.12%	36	business requirements	10.40%
11	experience working	18.11%	37	quality assurance	10.28%
12	qualified applicants	17.80%	38	design development	10.11%
13	work experience	15.44%	39	data management	9.93%
14	related field	15.35%	40	responsibilities include	9.78%
15	receive consideration	13.70%	41	work environment	9.78%
16	financial services	13.53%	42	work independently	9.71%
17	technical support	13.37%	43	skills strong	9.56%
18	working knowledge	13.27%	44	information security	9.54%
19	internal external	13.25%	45	written verbal	9.44%
20	applicants receive	12.86%	46	cross functional	9.36%
21	written communication	12.84%	47	experience required	9.27%
22	high school	12.73%	48	high quality	9.27%
23	fast paced	12.68%	49	employment opportunity	9.14%
24	duties assigned	12.21%	50	related experience	9.12%
25	policies procedures	12.16%	51	application development	9.12%
26	degree computer	12.13%	52	life cycle	9.10%

Table 4. Relative Occurrences of Bi-grams (two word phrases) in Computing Job Descriptions: 2016-2017 in PA, NY, OH (minimum frequency 10,000). Relative frequency computed with respect to total number of job postings in sample.

Tri-gram	Relative Frequency	Tri-gram	Relative Frequency
1 degree computer science	11.45%	30 school diploma equivalent	3.64%
2 high school diploma	10.42%	31 essential duties responsibilities	3.60%
3 written communication skills	10.31%	32 provide world class	3.55%
4 problem solving skills	8.59%	33 equivalent combination education	3.52%
5 minimum years experience	8.23%	34 excellent verbal written	3.50%
6 qualifications bachelor degree	7.62%	35 science related field	3.45%
7 verbal written communication	7.06%	36 world class technical	3.41%
8 bachelor degree computer	6.57%	37 support client customers	3.38%
9 ability work independently	6.18%	38 combination education experience	3.29%
10 fast paced environment	6.13%	39 project management skills	3.24%
11 computer science related	5.95%	40 requirements bachelor degree	3.12%
12 verbal communication skills	5.71%	41 experience bachelor degree	3.12%
13 communication skills ability	5.57%	42 mobile technology department	3.10%
14 written verbal communication	5.48%	43 experience minimum years	3.09%
15 years related experience	5.02%	44 valid driver license	3.06%
16 excellent communication skills	4.68%	45 global financial services	3.06%
17 school diploma ged	4.56%	46 computer science information	3.06%
18 development life cycle	4.55%	47 excellent written verbal	3.02%
19 oral written communication	4.44%	48 computer science engineering	3.02%
20 equivalent work experience	4.34%	49 financial services firm	2.96%
21 knowledge skills abilities	4.21%	50 sales associate responsible	2.96%
22 tech sales associate	4.16%	51 time management skills	2.91%
23 skills ability work	4.03%	52 project management experience	2.87%
24 experience years experience	3.94%	53 related field years	2.87%
25 perform essential functions	3.91%	54 cross functional teams	2.86%
26 bachelor degree equivalent	3.85%	55 leading global financial	2.80%
27 subject matter expert	3.70%	56 software development life	2.76%
28 concurrent technologies corporation	3.69%	57 degree years experience	2.75%
29 analytical problem solving	3.65%		

Table 5. Relative Occurrences of Tri-grams (three word phrases) in Computing Job Descriptions: 2016-2017 in PA, NY, OH (minimum frequency 3,000). Relative frequency computed with respect to total number of job postings in sample.