Analysis of Current and Future Computer Science Needs via Advertised Faculty Searches

Craig E. Wills

Professor and Department Head
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609

WPI-CS-TR-14-06

November 2014

Abstract

This work studies where Computer Science departments are choosing to invest faculty positions by examining data obtained from advertised faculty searches for the current hiring season. While the number and areas for faculty searches does not necessarily translate into the same for faculty hires, we believe that they provide insight into current and future needs within the discipline.

In this work we analyzed ads from 223 institutions for hundreds of tenure-track faculty positions in Computer Science and found that the clusters of Big Data, Security and Systems/Networking are the areas of greatest investment. We also found that from a third to over a half of all hires for areas that are, or may be, interdisciplinary in nature.

Differences are also seen when analyzing results based on the type of institution. Positions related to Big Data are roughly twice as frequent for PhD institutions as for BS and MS institutions. Security is of most interest for top-100 PhD and BS institutions. The Systems/Networks cluster is in relatively high demand from all types of institutions. Software Engineering is much less in demand for top-100 PhD institutions relative to the other institutions in our study. Finally, the abundance of potentially interdisciplinary areas is even more pronounced for PhD institutions with at least a third and up to 60% of all positions devoted to these areas.

1 Introduction

The wealth of faculty searches in Computer Science during this hiring season for positions starting in the Fall of 2015 affords the opportunity to study areas of Computer Science where departments are choosing to invest in new faculty hires. This report details results from a study of faculty hiring ads in Computer Science.

The primary focus of this work is to study where departments specifically, and the discipline more generally, are choosing to invest precious faculty positions. It is an opportunity to understand where Computer Science departments think they are in terms of current needs as well as where they think they are going.

With this focus, there are a number of caveats to our study:

1. Our study is not exhaustive in that it does not necessarily take into account all searches currently underway for this hiring season. We describe the methodology used to discover ads, but ads may have been missed or may not have been placed in the timeframe of our study.

2. While we report on the number of searches we found, we do not attempt to make longitudinal comparisons on how the number of searches and number of positions being searched for compares with previous years.

3. While our study focuses on preferred areas for faculty applicants, not all ads identify such preferred areas. These searches are accounted for in the data, but are not considered when analyzing particular areas of interest.

4. Our study analyzes searches and not hires. The number and areas of actual faculty hires may not match what is being searched for.

2 Methodology

We used two primary sources for obtaining ads for Computer Science faculty positions: the Computer Research Association (CRA) Job postings[1] and the Association for Computing Machinery (ACM) list of jobs[2]. We considered ads posted on these venues between August, 2014 and November 14, 2014. In addition, we augmented these two sources with positions posted on the SIGCSE mailing list, which often includes ads for more teaching-focused institutions.

Only ads for tenured and tenure-track positions by departments containing Computer Science were considered. We did not consider non-tenure-track positions such as lecturers, instructors or researchers. We also did not consider searches for department head positions.

For each ad we coded the institution name and the number of positions being searched for. If the ad included specific areas of interest then these were coded as well.

[1] `http://cra.org/ads/`
[2] `http://jobs.acm.org/c/search_results.cfm?site_id=1603`

3 Results

Using this methodology our resulting dataset contains information for faculty searches from 223 institutions (212 are U.S. based). 182 (82%) of these institutions indicate a specific number of positions being searched for with the remaining searches using non-specific phrases such as "multiple positions," "several positions" or just "positions" to indicate the number.

Similarly not all ads listed specific or preferred areas of interest. 163 (73%) of the 223 institutions listed specific areas. In studying particular areas of interest we only considered ads with specific areas for our analysis.

In the initial step of our study, we determined the number of times that a specific area was mentioned in an ad. Thus an ad for a single faculty position with preferred interest in the areas of HCI, Security, Machine Learning and Robotics would count one "mention" for each of these four areas. Another institution looking to focus three positions in the area of Security would be one mention for Security. A total of 488 specific areas are mentioned in ads. Figure 1 shows the percentage of mentions for each area.

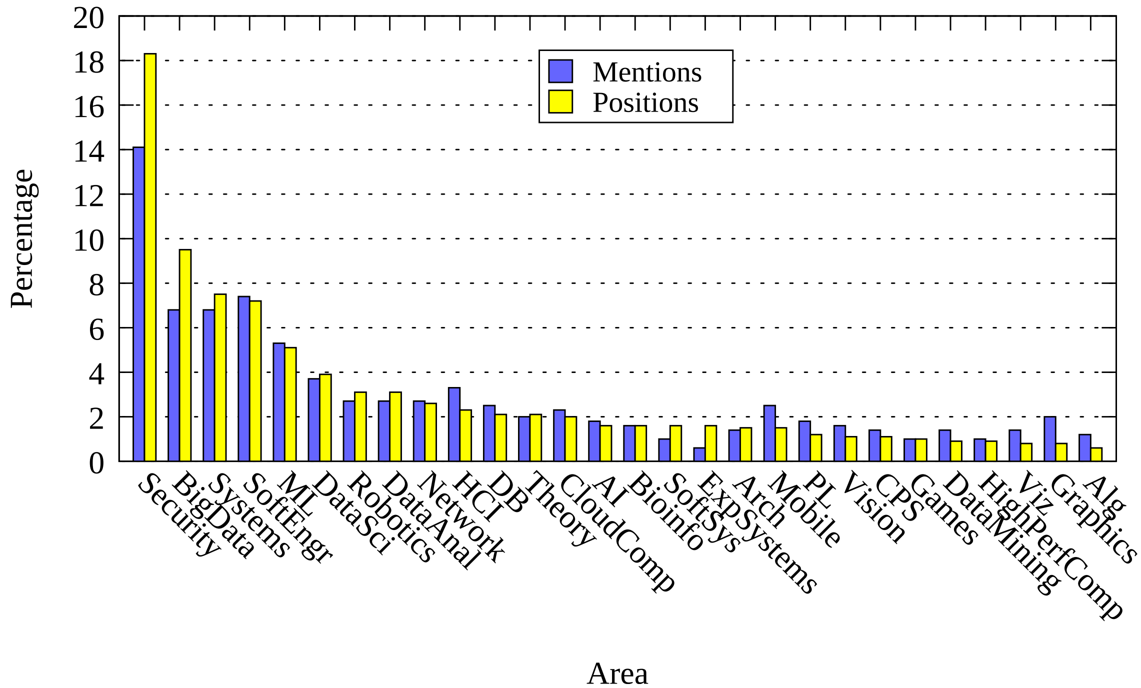

Figure 1: Area Percentage by Mentions and by Positions

While mentioned areas are one metric, another approach is to consider a faculty search as a "vote" for an area of current and future need. Using this approach a single position with four areas of interest would be investing 0.25 positions for each area, while three positions focused in a single area would invest 3.0 positions in that single area.

The problem with weighting areas based on the number of positions is that not all ads list a specific number of positions. For ads indicating multiple positions there are at least two, but otherwise the number is not known. To simplify our analysis we mapped each of these "multiple positions" searches to a fixed number of positions. We experimented with fixed values between two and four positions without significant differences in the overall results. All results shown in this report use a fixed value of three for multiple-position searches resulting in a total of 401 “positions” being searched for by the 223 institutions with 310 (77%) of the positions indicating preferences for specific areas. Figure 1 also shows the percentage of positions for areas with at least one percent for either mentions or positions. They are shown in rank order based on the number of positions.

The results show that the area of Security, which includes Privacy, accounts for the highest percentage of both mentions and positions, although it accounts for relatively more positions. Big Data is the area with the second most number of positions, although Software Engineering has the second most number of mentions. Apart from these two areas, the percentages for each metric are comparable for most of the remaining areas, although the percentage of mentions is a bit higher for areas such as HCI, Mobile Systems and Graphics.

Figure 1 does not show other mentioned areas that appear less frequently in ads. In order to account for these less-frequent areas and to combine similar areas, such as Big Data and Data Science, we grouped areas into clusters. These area clusters and the set of areas constituting the cluster are shown in Table 1. Areas with a small number of mentions and not clearly fitting into a cluster are included in two other clusters—one with areas in traditional Computer Science (OtherCS) and one with areas more interdisciplinary in nature (OtherInter).

Table 1: Areas Grouped in Each Area Cluster

Area Cluster	Constituent Areas
AI	AI, Data Mining, Machine Learning
Arch	Architecture
BigData	Big Data, Data Science, Data Analytics, Data Computation, Information Analysis, Visualization, Visual Computing, Knowledge Representation
Bioinfo	Bioinformatics
Compiler/PL	Compilers, Programming Languages, Object-Oriented Languages
CompSci	Computational Biology, Computational Life Science, Computational Science
DB	Database, Data Management
Games	Games, Interactive Media, Digital Media
HCI	HCI, Human Computing, Interactive Computing, Virtual Reality
ImageSci	Graphics, Image Processing, Pattern Recognition, Vision
Mobile	Mobile Systems
Robotics/CPS	Cyber-Physical Systems, Robotics, Sensors, Autonomous Systems
Security	Forensics, Security&Privacy
SoftEngr	Software Engineering, Software Design, Software Systems
Sys/Net	Cloud Computing, Distributed Computing, High Performance Computing, Experimental Systems, Networking, Network Science, Parallel Computing, Systems
Theory/Alg	Algorithms, Theory, Formal Methods
OtherCS	CS Education, IT, Natural Language Processing, Software, System Design, System Verification, Web Technologies
OtherInter	Biomedical, Business Intelligence, Cognitive Model, Cognitive Systems, Communications, Energy Awareness, Health Computing, Health Information Systems, Interdisciplinary, Learning Science, Social Computing

Figure 2 shows the same results as Figure 1 using the clusters from Table 1 rather than the areas directly. It shows that the Big Data, Security and Systems/Networking area clusters clearly have the highest percentage of mentions and a relatively higher percentage of positions. The remaining area clusters tend to have more comparable percentages between the two metrics save for Image Science and Mobile Systems.

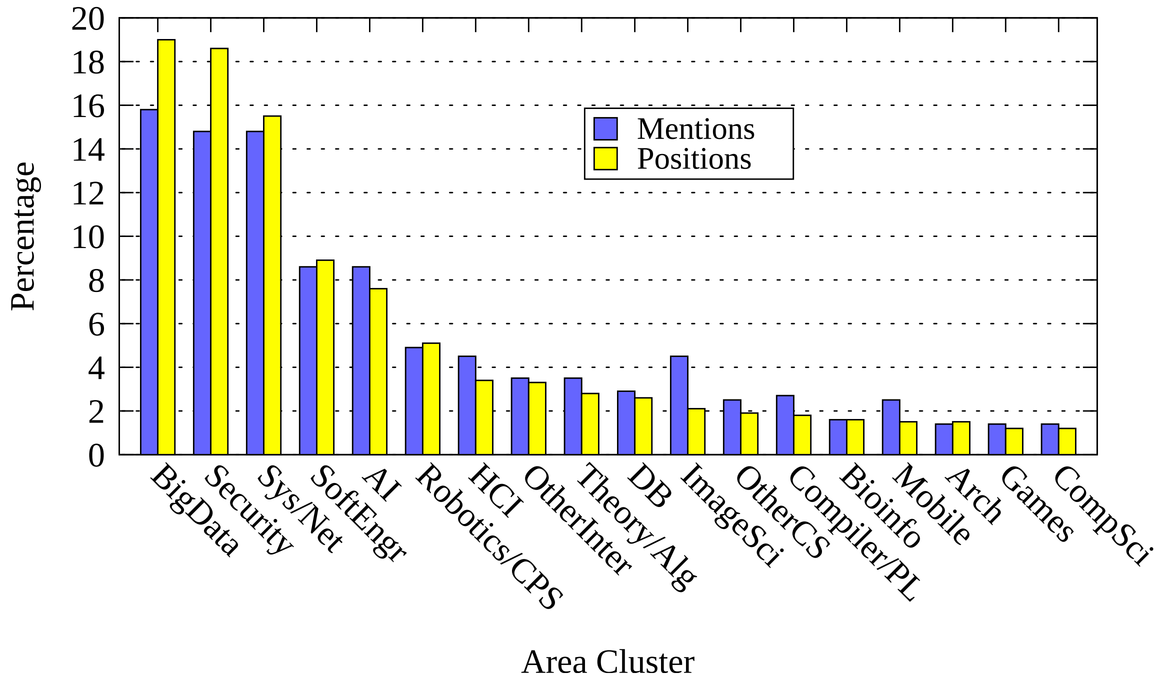

Figure 2: Clustered Area Percentage by Mentions and by Positions

Another question we examined is how the interdisciplinary nature of Computer Science is affecting hiring. Specific clusters in Table 1 that are more interdisciplinary include the Big Data, Robotics/CPS, Bioinformatics, Games, Computational Science, and Other Interdisciplinary clusters. Combining the results for these clusters from Figure 2 shows that 29% of the mentions and 31% of the positions are for more interdisciplinary areas. Moreover, other clusters such as AI and Security either support interdisciplinary work or may include work with other disciplines. Including these two clusters, which have some amount of interdisciplinary nature, results in up to 52% of the mentions and 57% of the positions being interdisciplinary in nature.

4 Results By Type of Institution

We repeated our analysis based on the type of the program at each institution. Undergraduate-focused programs may not have the same needs as PhD programs. For this portion of the study we augmented our dataset to include the highest degree offered by each program—BS, MS or PhD. Our dataset includes 122 PhD institutions. In order to study faculty investments at the most prominent U.S. programs, we further subdivided this group by using the U.S. News Rankings of the 100 Best Graduate schools[3]. This "PhD100" list account for 70 institutions in our dataset. The

[3] `http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-science-schools/computer-science-rankings`

remaining PhD programs, including the 11 non-U.S. based, are denoted as "PhDOther". Table 2 shows summary results based on the four institution types.

Table 2: Summary of Position Searches by Institution Type

Institution Type	Number of Institutions	Advertised Number of Positions 1	2	3+	Multiple	Total Positions	%Specific Area
PhD100	70	20 (29%)	13 (19%)	7 (10%)	30 (43%)	167	78%
PhDOther	52	20 (38%)	17 (33%)	8 (15%)	7 (13%)	106	87%
MS	48	31 (65%)	13 (27%)	0 (0%)	4 (8%)	69	80%
BS	53	48 (91%)	4 (8%)	1 (8%)	0 (2%)	59	56%
All	223	119 (53%)	47 (21%)	16 (7%)	41 (18%)	401	77%

Table 2 reveals differences between the different types of institutions. Ads for 91% of the BS institutions are for a single position while 43% of the ads for PhD100 institutions are for multiple positions. Only 56% of positions from BS institutions identify specific areas of interest.

In order to understand differences on areas of interest between different types of institutions, we use the position metric results and area clusters shown for all institutions in Figure 2. Figure 3 shows these same data (in the same rank order as Figure 2) grouped by the four types of institutions.

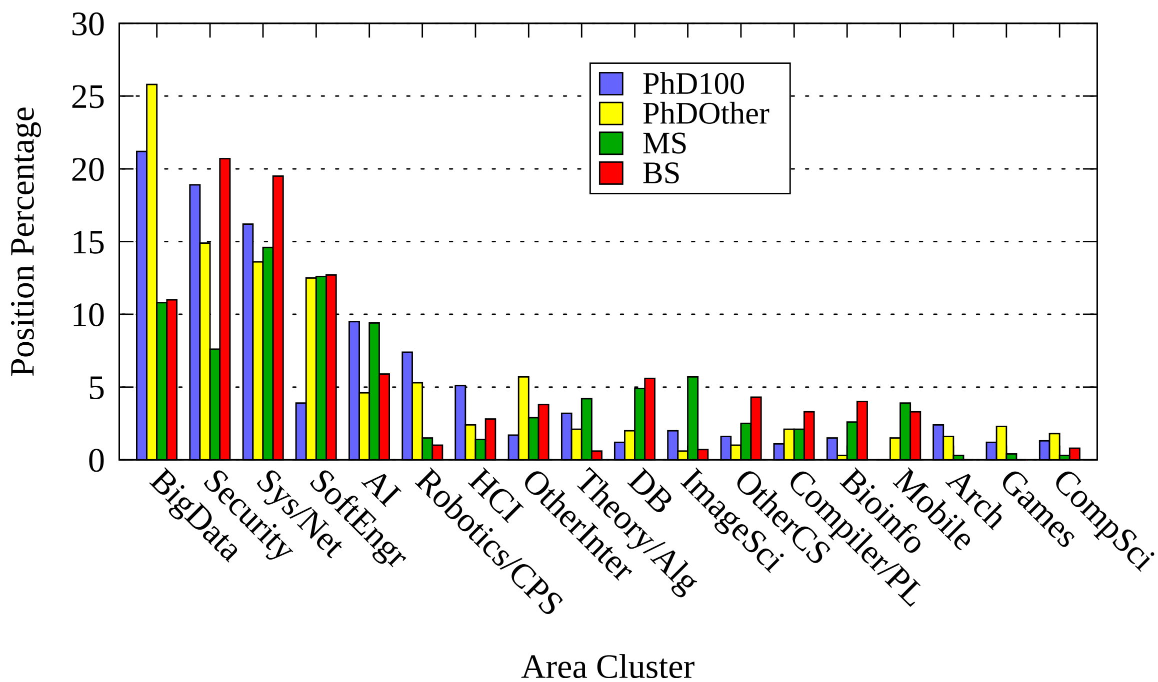

Figure 3: Area Cluster Percentage by Institution Type

The figure shows a number of interesting results. Positions related to Big Data are roughly twice as frequent for PhD institutions as for BS and MS institutions. Security is of most interest for PhD100 and BS institutions. The Systems/Networks cluster is in relatively high demand from all types of institutions. Software Engineering is much less in demand for PhD100 institutions relative to the other three types. The Robotics/CPS cluster is relatively much more in demand by the PhD institutions. The Database cluster is relatively much more in demand by the BS and MS institutions.

Finally, Figure 4 shows the percentage of positions that are devoted to areas more and some amount of interdisciplinary in nature. The results show that the impact of interdisciplinary areas is even more pronounced for PhD institutions with at least a third and up to 60% of all positions devoted to these areas.

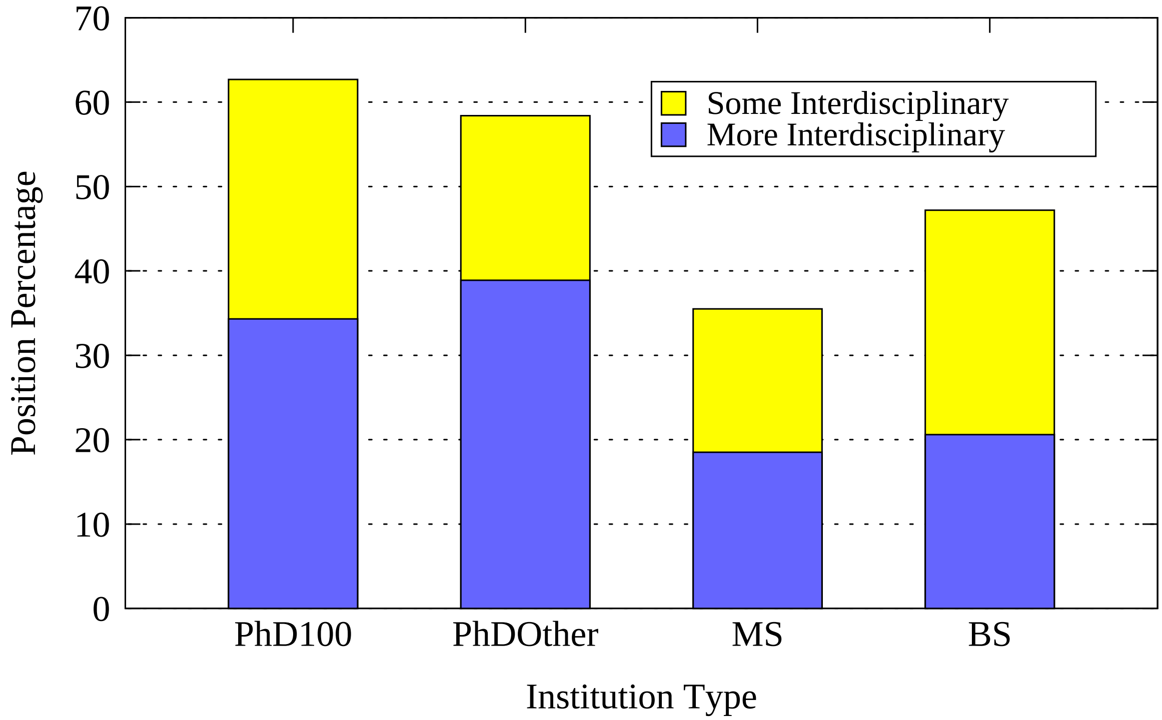

Figure 4: Investment in Interdisciplinary Area Clusters by Institution Type

5 Summary

This work studies where Computer Science departments are choosing to invest faculty positions by examining data obtained from advertised faculty searches for the current hiring season. While the number and areas for faculty searches does not necessarily translate into the same for faculty hires, we believe that they provide insight into current and future needs within the discipline.

In this work we analyzed ads from 223 institutions for hundreds of tenure-track faculty positions in Computer Science and found that the clusters of Big Data, Security and Systems/Networking are the areas of greatest investment. We also found that from a third to over a half of all hires for areas that are, or may be, interdisciplinary in nature.

Differences are also seen when analyzing results based on the type of institution. Positions related to Big Data are roughly twice as frequent for PhD institutions as for BS and MS institutions.

Security is of most interest for top-100 PhD and BS institutions. The Systems/Networks cluster is in relatively high demand from all types of institutions. Software Engineering is much less in demand for top-100 PhD institutions relative to the other institutions in our study. Finally, the abundance of potentially interdisciplinary areas is even more pronounced for PhD institutions with at least a third and up to 60% of all positions devoted to these areas.