



A Study on the Association between Computer Languages and Research Papers

Vatanak Lavan, Advisor: Ching-yu (Austin) Huang

School of Computer Science, Kean University, chuang@kean.edu

Abstract

The goal behind the undertaking of this project was to investigate the popularity of programming languages in the vast amounts of research published everyday. The ACM and IEEE online digital libraries are famous collections of all works published by the ACM and IEEE in their articles, magazines and conference proceedings. This study investigates the association between the type of computer language and the number of research papers published in the ACM and IEEE systems. Five programming languages - Java, Python, PHP, SQL, and JavaScript - were used as search keywords. In addition, data was also collected for combinations of each programming language with two different fields - Data Science and Cybersecurity. The search results were analysed and presented in charts and graphs. This study provides important insight into which computer languages are most prevalent in research all around the world.

Dataset

- ACM Digital Library
- IEEE Xplore Digital Library

	ACM	IEEE
	Date_Count	Date_Count
Search Keyword	4/3/2020	4/3/2020
Java	61,896	24,922
Python	22,486	3,288
PHP	7,758	1,116
SQL	22,941	4,741
JavaScript	14,740	1,664
Data Science	516,912	311,170
Security	103,374	224,622
Java AND Data Science	58,309	2,947
Java OR Data Science	521,227	333,145

Methods

The research during this project involved collecting the number of search results of keywords – Java, Python, PHP, SQL, and JavaScript – from ACM and IEEE as well as combinations with Data Science and Cybersecurity (e.g. “Data Science” AND Java) over a period of three weeks. Microsoft Excel was used to display the data in charts and graphs. Additionally, p -values were calculated using the Chi-Square test to find the correlation between programming language and number of search results which in turn, would determine the popularity of each language in the two systems (Fig. 9).

Results

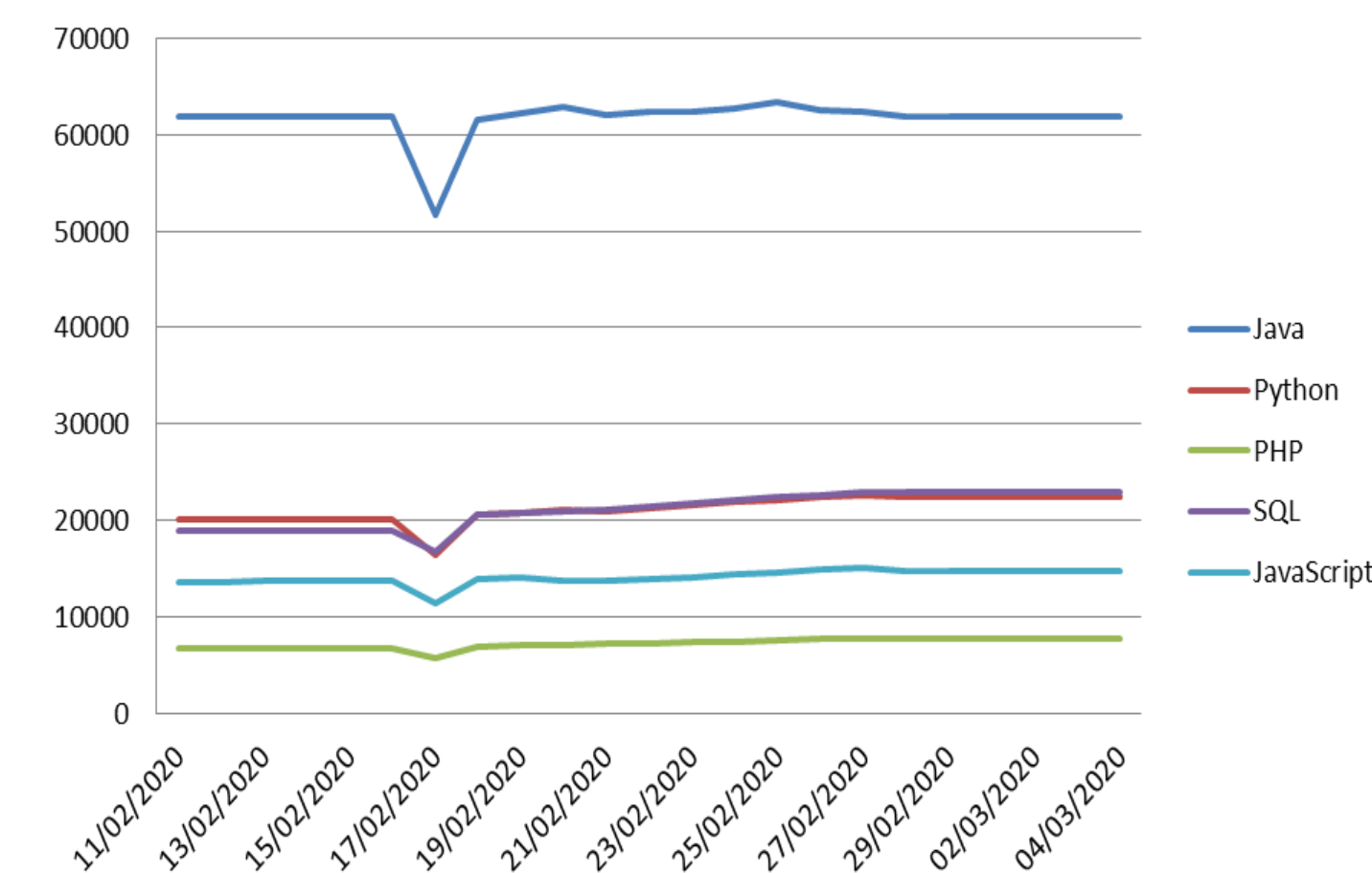


Figure 1. Number of search results for languages in ACM

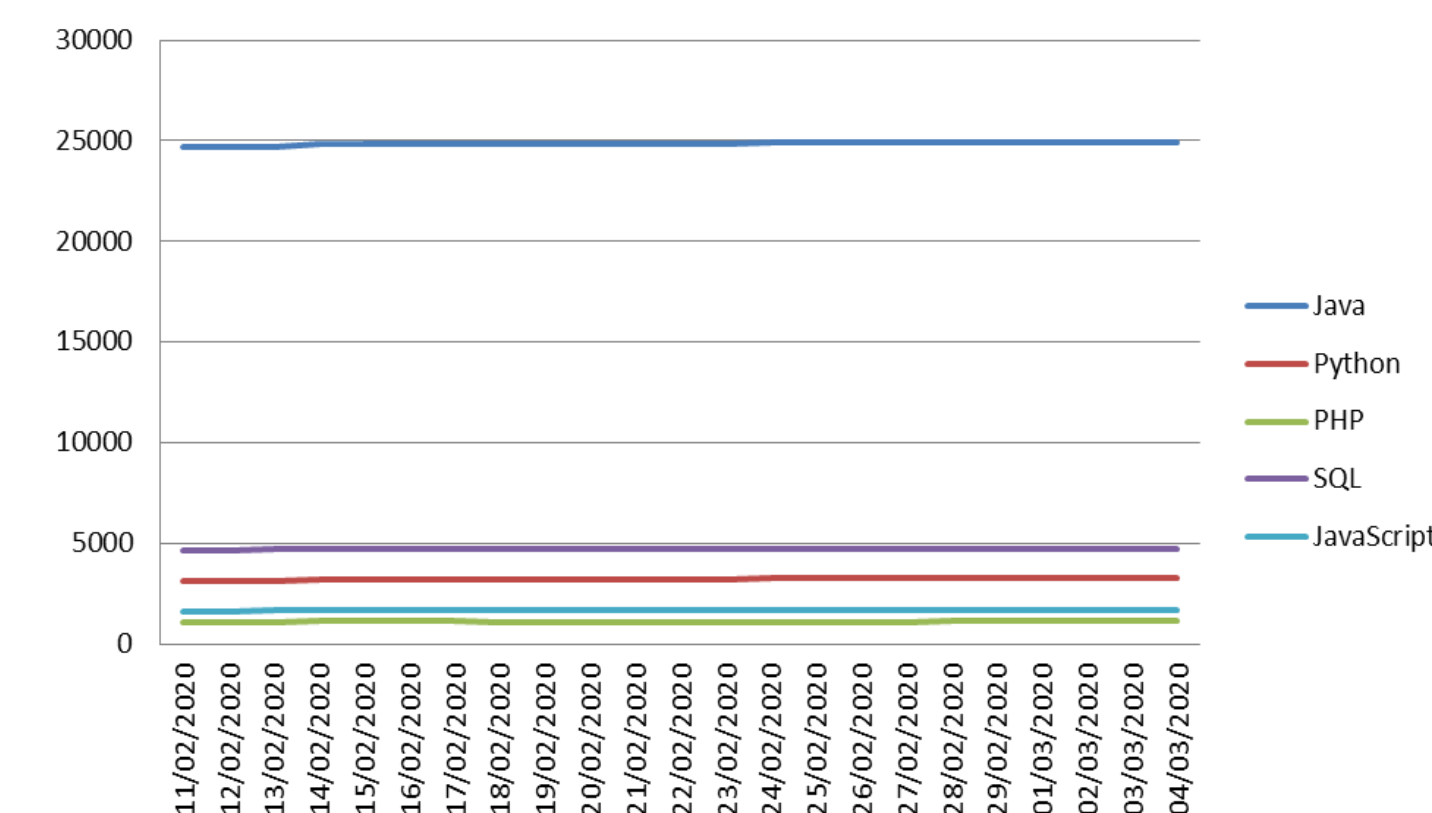


Figure 2. Number of search results for languages in IEEE

Materials

The materials used during the process of this project include:

- Microsoft Excel
- Social Science Statistics Chi-Square Calculator

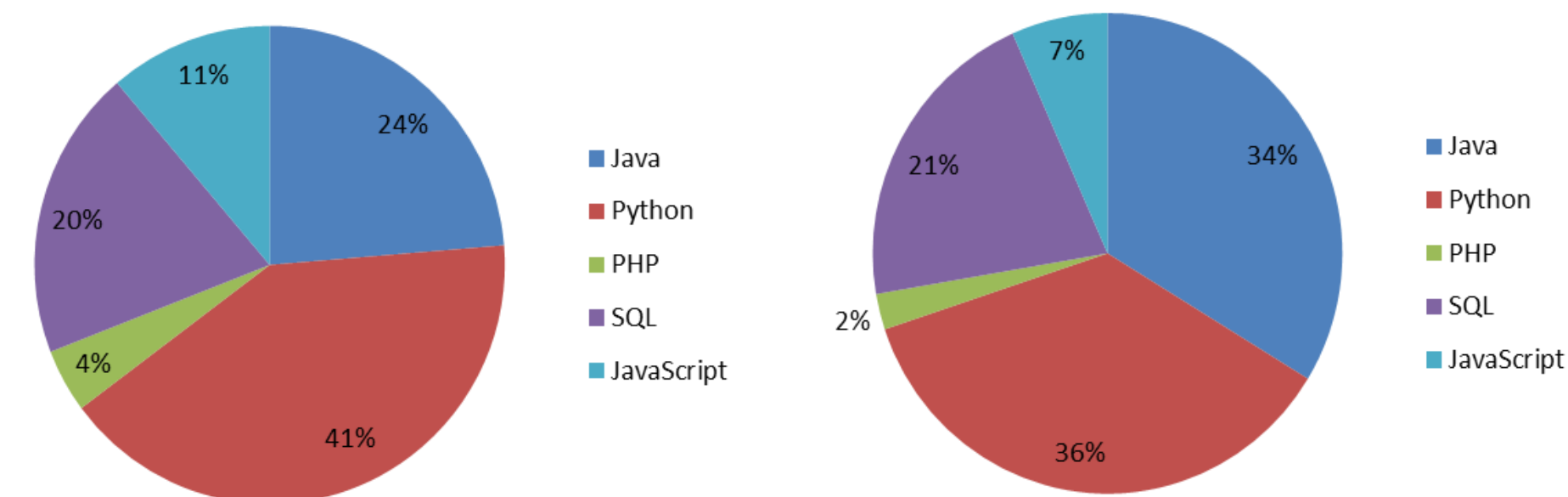


Figure 5 & 6. Percentage of search results for languages with Data Science in ACM vs. IEEE

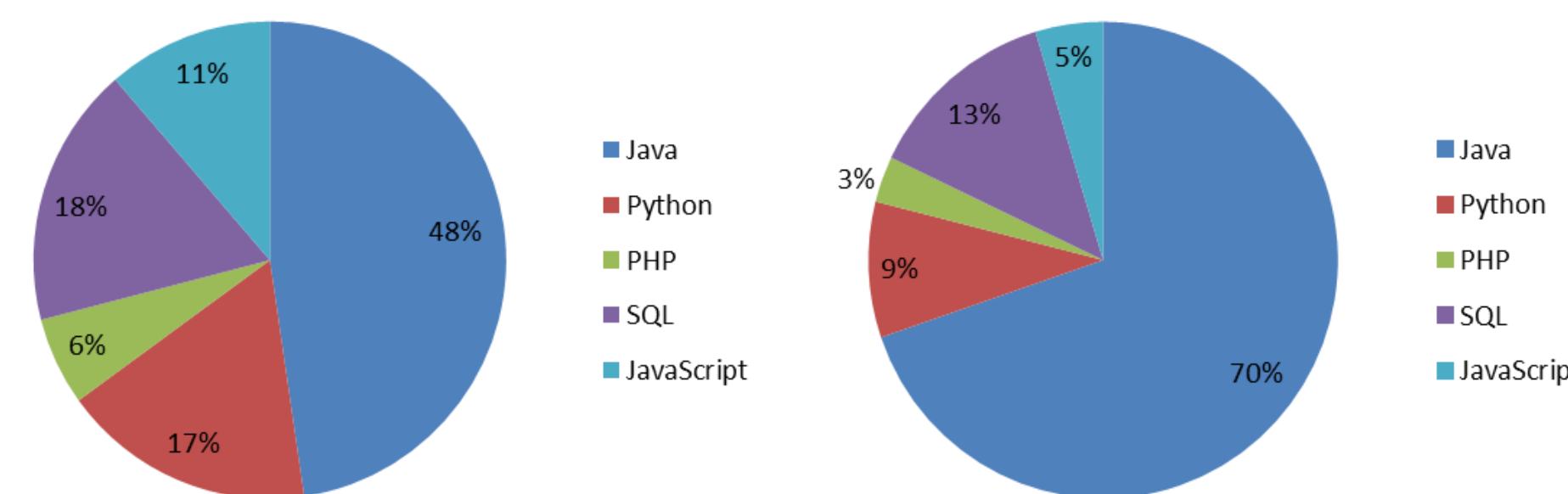


Figure 3 & 4. Percentage of search results for languages in ACM vs. IEEE

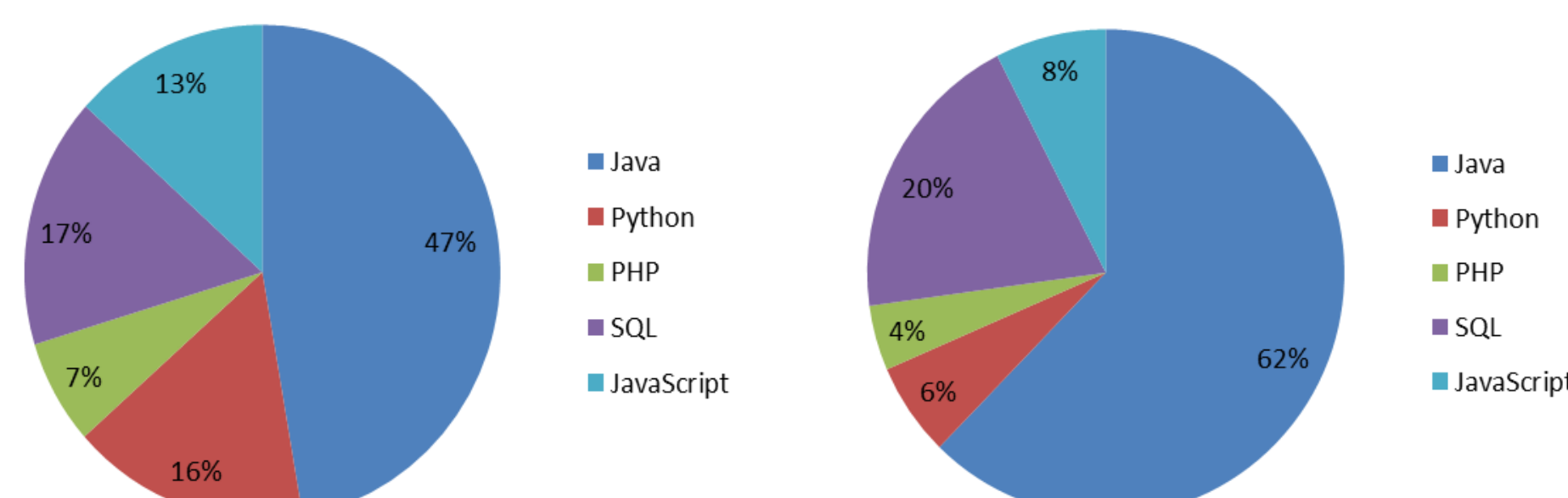


Figure 7 & 8. Percentage of search results for languages with Security in ACM vs. IEEE

Language	ACM	IEEE
Java	$p < 0.00001$ (Significant)	$p = 0.000048$ (Significant)
Python	$p < 0.00001$ (Significant)	$p < 0.00001$ (Significant)
PHP	$p = 0.000014$ (Significant)	$p = 0.305166$ (Not significant)
SQL	$p = 0.000464$ (Significant)	$p = 0.401314$ (Not significant)
JavaScript	$p = 0.008637$ (Significant)	$p = 0.866648$ (Not significant)

Figure 9. p -values of languages in ACM vs. IEEE at $p < 0.05$

Summary

Considering its long history, Java accounts for the most number of research papers published to this day (Fig. 1-4). As Python is gaining popularity, it places below Java closely with SQL (Fig. 1-4). In Security, Java again takes up a majority of the search results with SQL, Python, and JavaScript as notable contenders (Fig. 7&8). In Data Science, Python is the most popular, but Java is still considerable in number (Fig. 5&6). According to the p -values, it seems Java and Python are overall the most significant programming languages in ACM and IEEE (Fig. 9).

References

1. <https://dl.acm.org/>
2. <https://ieeexplore.ieee.org/Xplore/home.jsp>
3. <https://hackr.io/blog/best-programming-languages-to-learn-2020-jobs-future>
4. <https://www.youtube.com/watch?v=kXdufyvzGtE>
5. <https://www.youtube.com/watch?v=mUxS-35qO44&>
6. <https://www.youtube.com/watch?v=Og847HVwRSI>

Acknowledgement

We would like to express special thanks of gratitude to the Freshman Research Initiative who gave the opportunity.