

Data Visualization Analytic for Understanding the Dynamics of Operating System Using Programming Language Paradigm

Adeagbo Moruf Adedeji ^{a,1,*}, Agbaje Halimah Adebimpe ^{b,2}, Kasali Abdulwakil Adekunle ^{c,3}

^a Department of Computer Science, First Technical University, Ibadan, Nigeria

^b Beijing School of Aeronautics and Astronautics, Beihang University, Beijing, China

^c Innovation Technology Center, The Federal Polytechnic Ede, Ede, Nigeria.

¹ adedegv@gmail.com, ² agbajehalimah@gmail.com, ³ 3gabdulwakil@gmail.com

* corresponding author

ARTICLE INFO

Article history

Received January 15, 2022

Revised March 7, 2022

Accepted June 10, 2022

Keywords

operating systems

programming languages

software

programmer

ABSTRACT

Evolution of Programming Languages lead to the emergence of new one while the old one slowly fade away. The choice of most developer on Operating System depends on the Programming Language to execute a project. This infer that there is a connection between Programming Languages and Operating System. Hence, this study proposes a data visualization analytic to observe the trend of programming language usages over the years and their effects on Operating Systems. Dataset containing the usage of Programming Languages and Operating System were retrieved from GitHub, Flourish and Stack overflow using text mining that employed regex techniques and Python was used to implement data analytic and visualization. Considering the first ten popular Programming Languages from 330,936 instances, our result similar to the result of other research output with Python leading the Programming Languages. The result of change over time of Operating System shows that Window OS is tending toward negative path while Linux OS and Mac OS are tending toward positive path. Also, Java and JavaScript are mostly used on Window OS, Objective C and Swift are mostly used on Mac OS and Python is mostly used on Linux. The overall results over the years showed that programmers are consistently shifting from Window OS to Linux OS and Mac OS.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

Trends have both promising and adverse impact on decision making [1]. As the evolutionary theory [2] cover every sphere of life, so also the Programing Languages [3] and Operating System (OS) [4]. In the programming world, Programming Languages and OS are two inseparable entities. OS is the underlying software that enables users interface with the computer to carry out basic task and allow software developer write and run code written in any Programming Languages [5]. OS can be found on both Personal Computer PC and Mobile or Handheld Device. Among the major OS that can be found on PC are Linux, Windows OS and Mac OS [4]. The developers make choice of OS that is specifically best suitable for programming and development of application [6]. In this context, Programming Languages are regarded as independent variables while OS(s) are regarded as dependent variable. As the OS and Programming Languages are inseparable entities, and evolution encompasses every sphere of life, one of the possible phenomena to explore is analyzing the trends of Programming Languages and its effect on the choice of Operating Systems among software developers. This enables software engineer, business enterprise, researchers and other stakeholders to understand the trend of OS in relation to Programming Language to aid their prediction and

decision making [7]. It also serves as basic factors to measure outstanding Programming Languages [1].

For more than 200 years, Programming Language has been in existence [3] and have a direct connection with the nature of project to be accomplished. Each with its set of rules, tools, applications and where it can be used. So, the developers pick Programming Languages for projects having known the project to be developed and understand the appropriate Programming Language for such software projects [8]. This serves a major reason a Programming Language is choosing over another or motivate the reason new Programming Languages are designed to execute a software project [1]. Therefore, based on evolutionary process, as the old Programming Languages slowly go into extinction so also the emergence of new Programming Languages [3]. Several approaches had been applied by researchers and expert to gather and analyze data in order to unravel the trend of Programming Languages. [9] provided answers to several research questions bothering interoperability, impact and popularity of Programming Languages by analyzing thousands of software project that are open source on GitHub repository. In this vein, [10] empirically carried out the analysis of Programming Language adoption through the identification of features embedded in Programming Languages that responsible for their adoption. It used thousands of datasets retrieved from SourceForge projects, Ohloh and various survey of software developers. [1] also unravel the most predominant Programming Language by analyzing project in open-source repository, that is, GitHub using Changelog Nightly. Changelog Nightly is a newsletter that dish out day-to-day information and current popular update on GitHub page which allows the development of scraper with a view to extracting relevant data. [11] [12] arrive at their result of most Popular Programming Languages through the analysis of result of web search by observing the aimed search phrase. [12][11] rates the Programming Languages based the counting of target word on search result of web pages on searching engine (e.g. Amazon, Bing, Baidu, Wikipedia, YouTube and Yahoo) while [11] is based on analyzing how frequent tutorial on Programming Languages are searched on the Google search engine. This is with the assumption that the more people are interested in a Programming Language, the more they search online for its tutorials.

2. Methods and Procedures

Trends of Programming Language can either be carried out by analyzing projects residing on open source repository or analyzing the result of web-search through the observation of key search phrase. Both method can also be extended to finding the prevalent OS among the programmers. While the first approach is commonly explored by the researcher and found in academic paper, the later is usually used to construct index [1]. For the purpose of this research, three different datasets were used. The first two sets of dataset were downloaded from Kaggle and the third dataset was downloaded from Stackoverflow. The first dataset contain data on usage of twenty-six Programming Language retrieved from Flourish between 1965 to 2019. The second dataset comprising data on usage of twenty-eight Programming Language retrieved from GitHub between 2004 and 2020. The third dataset contains information about Programming Languages used by programmers as well as the OS used to run those Languages on Stackoverflow between 2011 and 2020.

As shown in Figure 1, the first two dataset was benched mark over one another by finding the intersection of the first 10 most popular Programming Languages in both result after carrying out the ranking and changing overtime of Programming Languages. This is with the view to comparing the results of the two datasets and ensure that the prevailing Programming Languages are well represented. Cross mapping [13] help to reduce the number of Programming Languages from twenty eight to the bearest minimum before interfacing the result with the third dataset that comprise both Programming Languages and OS. Data was analyzed from CSV files using Python tools that embed Panda, NumPy and matplotlib in Jupyter Notebook environment. Also, the dataset in stack overflow was in text format which was transform into numeric value using data mining approach that embedded regex analysis to allow the carrying out of quantitative analysis. The Figure 2 shows the overall framework of the implementation process.

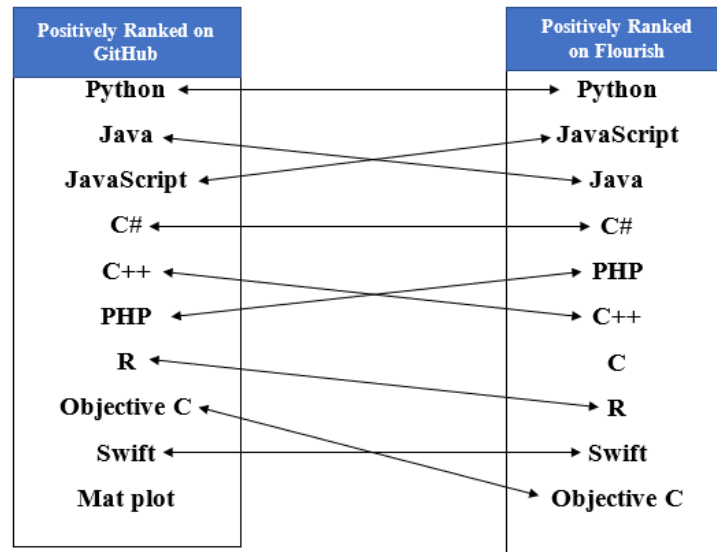


Figure 1. Cross Mapping of Ranked Result

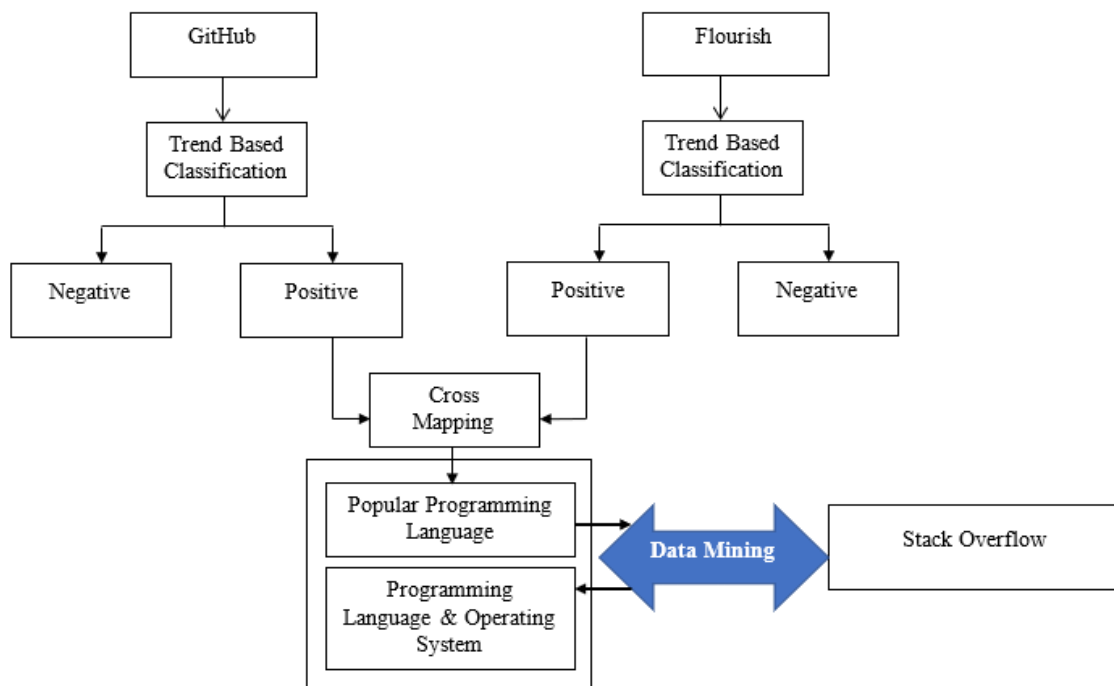


Figure 2. Process Framework

3. Results and Discussion

The work used dataset of Programming Languages usage over the years against the Operating System used to code in those Programming Languages. Data mining that embeds regex analysis was used to retrieved key phrase from CSV which was originally in text format. The extracted phrases were transformed from text form to numeric form to enable empirically analysis of the result. The transformation processes check for the presence of each Programming Language and Operating System based on the result in Figure 1 across rows with total instances of 330,936.

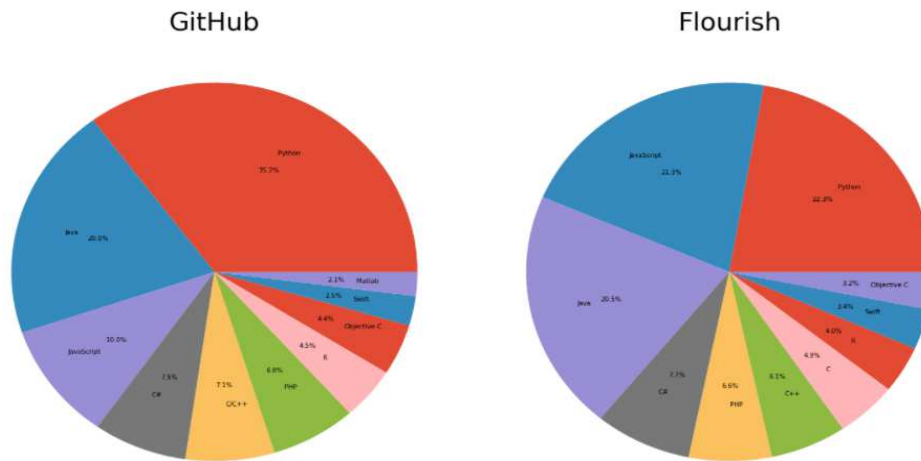


Figure 3. Trend of Programming Languages

As shown in Figure 3, the ranking of the prevailing Programming Languages on both GitHub and Flourish show Python to be the foremost Programming Language. Also, both results agreed nine Programming Languages to be among the first ten Programming Languages.

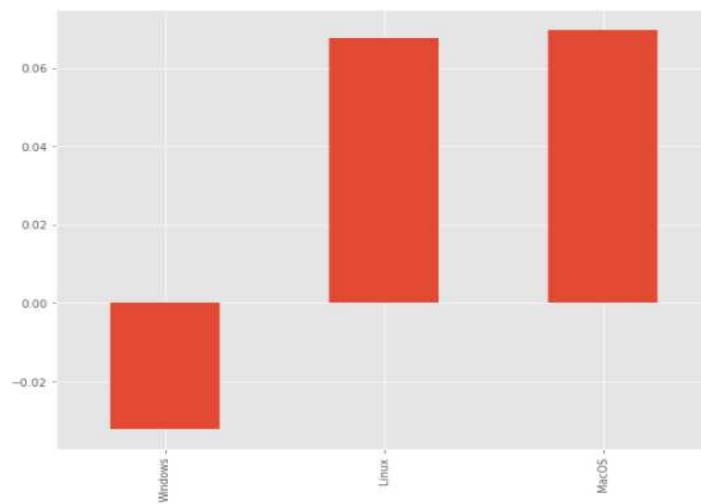


Figure 4. Change Overtime of OS

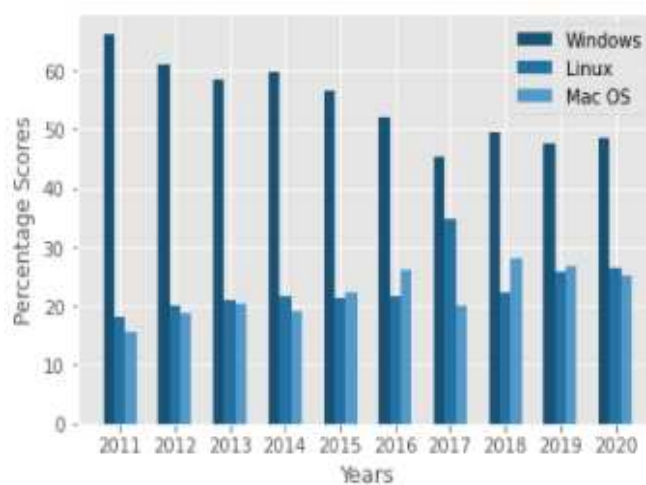


Figure 5. Trend of Operating System

As shown in Figure 4, the result of Operating System change overtime shows that Window OS is tending towards negative path while MacOS and Linux OS are tending towards positive path. Figure 5 shows that Window OS is consistently declining over the years while Linux OS and MacOS are considerably rising over the years. According to Figure 6, the result shows that Java, JavaScript, C# and C++ are mostly used on Window OS, Objective C and Swift are mostly used on MacOS and Python is mingling between Window OS and Linux OS. However, the trend in Figure 7 shows that C# and C++ are consistently declining over the years and Java and JavaScript have been able to consistently maintained their status of prevailing Programming Language over the years. This suggest that C# and C++ are likely responsible for the declining of Window OS while Java and JavaScript are likely responsible for Window OS still leading the trend. The result also suggests that the rising in MacOS is majorly due to increase in the usage of Objective C and Swift by Programmer. As depicted in the Figure 3, considering Python as the leading Programming Language could be responsible for rising in Linux OS as well as relative shifting of programmers in Java, R, PHP, and C++.

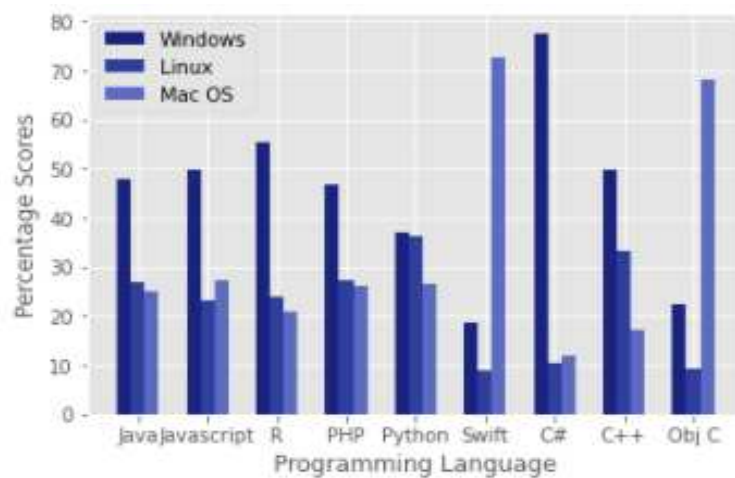


Figure 6. OS and Program Language

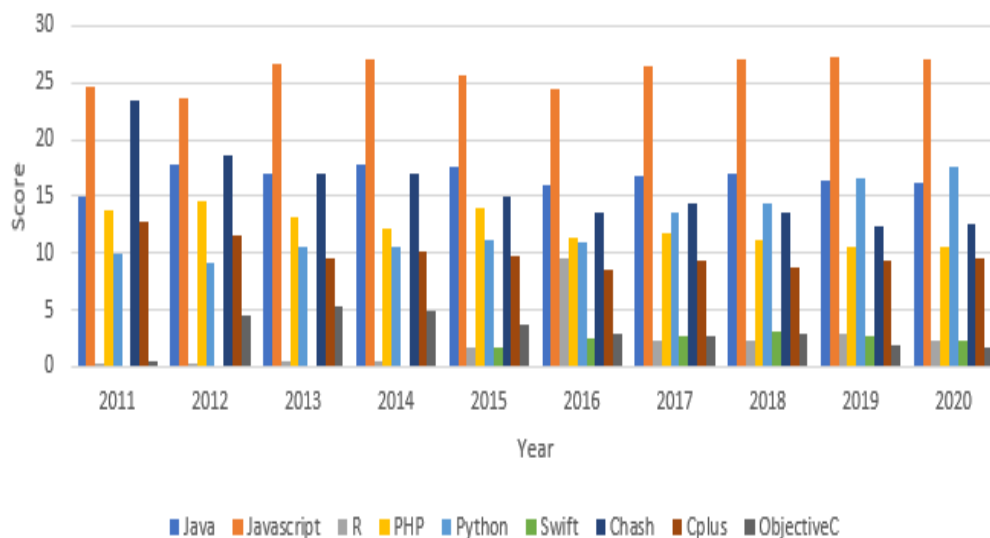


Figure 7. Language Usage over the Years

4. Conclusion

The research proposed the choice of Operating System among programmer using Programming Language paradigm. The work used dataset of Programming Languages usage over the years against the Operating System used to code in those Programming Languages. The study analyzed the trend of Programming Languages and correlate it against the trend of Operating System. Considering the first ten popular Programming Languages, our result is fairly similar with the result of other research output [1] and real time index [12][11] on Programming Languages that are prevailing. Regarding the change over time of Operating System, the result shows that Window OS is tending toward negative path while Linux OS and Mac OS are tending toward positive path. In the same vein, Java and JavaScript mostly on Window OS, Objective C and Swift are mostly used on Mac OS while Python is mostly used on Linux. Also, all the represented Programming Languages are relatively shifting toward the use of Linux OS and Mac OS. The overall results over the years showed that programmers are consistently shifting from Window OS to Linux OS and Mac OS.

Programmer make choice of the Operating System that suitably fit Programming Language to execute a project [6]. This implies that there are inherent features embedded in Operating System that make programmer choose one over another. Therefore, further research is suggested to find those inherent features in each Operating System that make it suitable for one Programming Languages over the other. The dataset used in this research are subset of representation of the entire programmers and can be further expanded to accommodate data from other sources. Also, machine learning that predict the suitable Operating System based on Programming Language is suggested as future work.

References

- [1] K. Gunnarsson and O. Herber, "The Most Popular Programming Languages of GitHub 's Trending Repositories," KTH Vetenskap OCH Konst, 2020.
- [2] C. Darwin, "On the Origin of Species." John Muray, p. 502, 1859, [Online]. Available: http://www.vliz.be/docs/Zeecijfers/Origin_of_Species.pdf.
- [3] K. R. Chowdhary, "On the Evolution of Programming Languages," no. June, 2020, [Online]. Available: <http://arxiv.org/abs/2007.02699>.
- [4] S. Kabiraj, S. K. Chandra, and A. Gupta, "Operating System a Case Study," *Int. J. Trend Sci. Res. Dev.*, vol. Volume-2, no. Issue-3, pp. 166–175, 2018, doi: 10.31142/ijtsrd10780.
- [5] K. Tsvetkov, "Operating Systems. The Past, Present and Future," 2020. doi: DOI: 10.13140/RG.2.2.36259.07202.
- [6] g2.com, "Best Operating Systems," 2021. <https://www.g2.com/categories/operating-system> (accessed May 29, 2021).
- [7] Y. Chen, R. Dios, A. Mili, L. Wu, and K. Wang, "An empirical study of programming language trends," *IEEE Softw.*, vol. 22, no. 3, pp. 72–79, 2005, doi: 10.1109/MS.2005.55.
- [8] P. Kumar, "Best Programming Languages to Learn in 2020," 2019. <https://studywarehouse.com/best-programming-languages-learn-2020/> (accessed May 28, 2021).
- [9] T. F. Bissyande, F. Thung, D. Lo, L. Jiang, and L. Reveillere, "Popularity, interoperability, and impact of programming languages in 100,000 open source projects," in *International Computer Software and Applications Conference*, 2013, pp. 303–312, doi: 10.1109/COMPSAC.2013.55.
- [10] L. A. Meyerovich and A. S. Rabkin, "Empirical analysis of programming language adoption," in *Proceedings of the 2013 ACM SIGPLAN International Conference on Object Oriented Programming Systems Languages & Applications*, 2013, vol. 48, no. 10, pp. 1–18, doi: 10.1145/2544173.2509515.
- [11] P. Carbonnelle, "PYPL PopularitY of Programming Language," *PYPL index*, 2021. <http://pypl.github.io/PYPL.html> (accessed May 28, 2021).
- [12] TIOBE Software, "TIOBE Index for May 2021," *TIOBE*, 2021. <https://www.tiobe.com/tiobe-index/> (accessed May 28, 2021).
- [13] S. Paul, L. Jolla, and S. Barbara, "from short time series R eports R eports," *Ecology*, vol. 96, no. 5, pp. 1174–1181, 2015, [Online]. Available: <https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/14-1479.1>.