Q&A Website Study for Programming Language and Platform for Developers-A Stack Overflow Study

Golam Md. Muradul Bashir[1], Sajal Saha[2], Md. Raihan Talukder[3], Joy Karmaker[4], Md. Saiful Islam[5], Mohammad Jamal Hossain[6]

Faculty of Computer Science and Engineering
Patuakhali Science and Technology University
Patuakhali, Bangladesh

A B S T R A C T
Stack Overflow is a famous Q&A internet site which used by developers and programmers for plenty development and programming purposes. However amongst all the programming languages that are extra reliable and helpful for programmers is a burning question. We researched over Stack Overflow commits and reached in an end. Our main target become to research over Stack Overflow commits and to present an end result which can enhance the self belief degree of a programmer. In our research we attempted to expose an expansion over Stack Overflow. We laboured with previous dataset which turned into as much as 2015 and it contained 1139 Stack Overflow associated commits. Our new dataset which is 2016 to mid-2018 it contained 1800 Stack Overflow commits. After working with the new and former facts we look at that there are small variations among the 2 consequences. In our previous end result, we noticed that crowds are normally worried in "JavaScript" however in our further studies; we observed that "Java" replaces "JavaScript". If we don't forget responsiveness then "C" overtakes "Java" and possibility of having a commonplace solution "Java" suggests higher effects than "Python". If we examine the results in each situation we see that "net development" has huge involvement than different structures. Considering votes, the quantity of commits, reaction time we are able to see that "internet design" still remains in the equal function in both results. From our research builders and programmers can now apprehend which platform and language they have to pick out.

Keywords— Stack Overflow, platform, language, commit, crowd.

I. INTRODUCTION

Q&A websites are those kinds of websites where user can ask any types of questions and get proper answer of these questions. When user faces any problems, they get help from any Q&A websites. User should follow the websites rules and asked questions in the websites. There are also expert people crowd there. They try heart and soul to give the answer of following questions of the user. That's how people get help from the Q&A websites.

Every Q&A website user has to follow some rules and regulations for asking any question and answer any question. These rules are maintained to avoid unnecessary and irrelevant questions and assure the quality of questions. Every user need to create their personal account. There are also a rating and raking for every user. Users who gave correct answer its increase the rating of that particular user. In every Q&A website expert are divided into different sections.

Stack Overflow is such kind of website where programmers and developers asked their related fields questions and experts from this fields gave answer of their questions. For different level of programmers and developers Stack Overflow is the most popular among the other Q&A websites.

Large communities are always connected to this Q&A website. For this reason we select this website for our research purpose. We expect more accurate result over analyzing Stack Overflow and we get a clear knowledge about programming and development sectors. Every Q&A website are displayed in a structured manner. There is some important information stored in against every question. This information is the main element of our research study.

Before, we work with dataset which contains up to 2015 and now we work with the dataset which contains data of middle of 2018. We divided our whole work into several parts by answering some questions for better understanding and make our work simple to others.
1. Find out best programming language and development platforms.
-By analyzing the following question we can say the involvement of people in "Stack Overflow".

2. Trending technologies

-By answering the questions we can say that which technologies is now on trending and which programming language and platforms are related to this modern technologies.

3. Understand the evolution of Stack Overflow.

-In our research we tried to show an evolution over "Stack Overflow". Both research results show some differences. We work with the differences and visualize them through graphical representation.

Motivation

In some of the previous works

1. They tried to say that code is reuse from different sources. They also discussed when code reuses are mostly happen and which types of people are mostly involved in code reuse.

2. They also tried to show that, in which area "Stack Overflow" are most helpful and responsive.

3. In our previous research we worked with dataset up to 2015 and we tried to show the best programming language and platform according to our research result.

We can understand that a huge numbers of people are dependent on code reuse in development and programming sectors.

In computer science there are some popular topics and data mining is one of them. Modern days are the days of AI technology and data mining is one of the AI topics. Nowadays large organizations are mostly dependent on data mining. They decide which steps to take analyzing the huge number of their organizational data. We always tried to cope with modern technologies and that's why we get interested in data mining. So, we got knowledge about data mining and how to implement data mining in our research.

Case Study Setup

Our main purpose was to show evolution over "Stack Overflow" based on previous and new research results. For this reason we had to analyze over new data and we have to observe for which programming language and development platform people are mostly involved in "Stack Overflow".

For this reason we divided our work into several steps. We know that, data mining follows KDD process. KDD process also has some steps and after completing each of the following steps we finally get the knowledge over analyzing the data.

We can see that "Stack Overflow" is the most popular Q&A website and there were several research works done before related to "Stack Overflow". They tried to show different results after performing different kinds of computation. So, we were interested to show the evolution of "Stack Overflow" and we show this according to programming language and development platforms. In our previous research we worked with dataset up to 2015 and in our current work we perform the same operation on different dataset. Our current work contains data from 2016 to middle of 2018 and we tried to show differences. Our work mainly related to data mining. So, applying data mining process we tried to show the changes of user involvement according to programming languages and development platforms in "Stack Overflow". We studied "Stack Overflow" related research papers and we found out some difficulties. Though there were many researches over "Stack Overflow", no research was related to programming languages and development platforms. It is difficult for a beginner which programming language he should choose and which development platform he should prefer. They even don't know from where they get better learning materials if they want to choose their suitable language and platform. As "Stack Overflow" is a Q&A website and different level of experts are attached to it. The experts are take part in answering the following questions answer. So, beginner should have the idea which programming language and development platform are helpful for them according to "Stack Overflow".

II.   DATA COLLECTION

For the purpose of our study we need a dataset which is related to Stack Overflow commit.
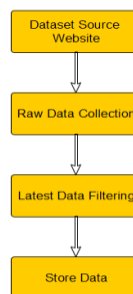
Fig. 1 – Data Collection Process.

According to this block diagram we can see how the dataset collected from the "Source website". Here we collected raw dataset from "GHTorrent". So, we can consider "GHTorrent" as our source website. The GHTorent project uses the Github API to collect raw data and extract, archive and share query able metadata. The created datasets are already being used for another task [2].

There are many data's on commits, pull requests, followers, issues, members, users, watchers etc. We need commits data from these dataset. We clean the data and collect commit data.

2.1 Data Selection

Then we did "string pattern matching" technique to find out stack overflow related commit. We made an algorithm for this task. Our algorithm is used for the purpose of the string matching. We find all commits that have "stack overflow" or its variants such as "stack overflow" or "Stack Overflow" etc. After applying that algorithm, we got 1149 stack overflow related commits. From these commits 50 "stack overflow" word was in the same commits. Thus we find out 1368 stack overflow commits. From these commits there were some commits which have no links and some commits were duplicated. We remove these non-linked and duplicated commits. Then we got 1139 stack overflow related commits. Thus from these huge dataset we select stack overflow related commits. We inspect every commit links for our research.

First of all, the imported data sets are in CSV format, where the file contains the commits details. The commit contains commit id, user id, commits links, commit submission time and commit submission date. In all data about commit, here we need the commit links only. But the commit links is described with some other garbage. So, we have to collect the required links. But the

CSV file of commits contains 1.5 million rows. In this case, to collect data links manually isn't easy from the huge large datasets. To solve the problem, we think about another solution. To find relevant data links, we use string pattern matching. The string pattern matching programed in C programming language.
Following algorithm shows us the steps of the program to separate "stack overflow" related commits.
The pattern matching program is made according to the algorithm:
1.   Open text input file.
2.   Open text output file.
3.   While (input is equal 1)
If (match ("keyword"))
        Print (input).
End of If.
        Increase count by 1.
4.   Return from program
Following flow chart is for the C program of string pattern matching technique. We can easily understand how the program match "Stack Overflow" related commits and then separately store them into another TXT file. The whole flow chart is according to the algorithm.
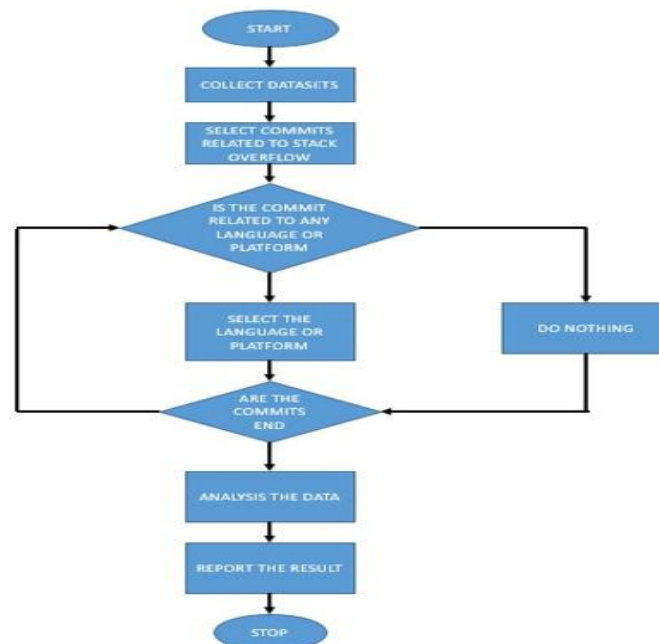


Fig. 2 – Data Selection Process.

2.2 Classify and Store

This is the most important task of this research. We observed all the links from the commits. Every link consists of a Stack Overflow page. We collected our essential data from that page for the evaluation. We collected response time, votes, acceptance answer from every links. To find out response time we find difference between first answer time and question time. Acceptance answer for the question is a green tick mark on the left side of the answer. We also collect the votes for the acceptance answer. These are the required data for our research.

We use Python program to do the work. We use two python library and some commands. This process is also known as web scarping. Performing this process we can retrieve data from required fields from any website. So, we perform this operation on Stack Overflow, get the required information and perform some programming and get the result and store the value in Microsoft excel sheet.

Algorithm of technical works:
1.   Open file in read mode.
2.   For (0 to EOF)
          Browse every link.
    Select ('div.fl0')
    Find all (class_="relativetime")
    Declare an array
          Store value in the array
3.   Sort the array value.
4.   Select ('div.owner')
    Find (class_="relativetime")
    Find ask_time
    response_time = array [0]-ask_time
    Write on the excel sheet
    Column increase by one.
5.   Find (class_="post-tag")
    Write on the excel sheet
    Column increase by one.
6.   Select ('div.js-voting-container')
          For (0 to Array length)
    If (select ('.js-accepted-answer-indicator'))
          Select ('.js-vote-count')
               Write on the excel sheet
    Initialize column value;
    Row increase by one.
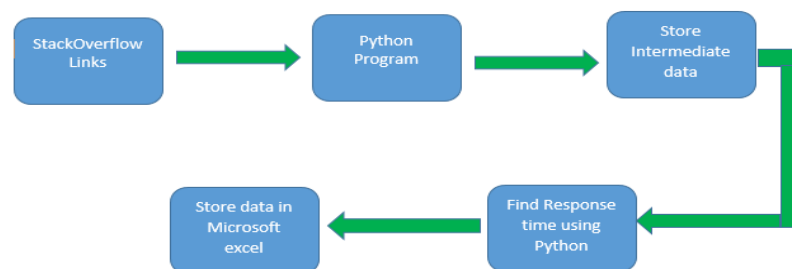
Fig. 3 – Data Store Process.

III. PRELIMINARY ANALYSIS

Our next target was to calculate Cohen's Kappa coefficient over our result of most committed language. Cohen kappa coefficient is a calculation that measures the agreement of certain limits of standard (divisive) elements. To calculate Cohen's Kappa coefficient we take agreements of two annotators. One is lecturer of Computer Science and Engineering and one is a BSc student of Computer Science and Engineering. Both are experts at programming language and has a very good idea on Stack Overflow. We use three scale range between -1 to +1. Where -1 means poorer than chance agreement, 0 means exactly chance agreement and +1 indicates better than agreement. Here are their decisions:

Table 1 – Agreement of annotators.

Language	Lecturer	BSc Student
Java	0	0
Java Script	1	1
Python	1	1
C++	0	0
Ruby	-1	-1
PHP	+1	+1
C	0	0
C#	0	1
Objective-C	0	0
Scala	-1	-1

Now we calculate the probability for three scales and the whole dataset to calculate the co-efficient. Following table shows similarity and dissimilarity for both annotators.

Table 2 – Cohen's Kappa calculation table.

		Lecturer			
		+1	**0**	**-1**	**Sum**
BSc Student	**+1**	2	0	0	2
	0	0	4	1	5
	-1	0	0	3	3
	Sum	2	4	4	10

Calculation:
The observed proportionate agreement, Po = (2 + 4 + 3)/10 = 0.9
Probability of -1, P-1 = (2/10) * (2/10) = 0.04
Probability of 0, P0 = (5/10) * (4/10) = 0.2
Probability of +1, P+1 = (3/10) * (4/10) = 0.12
Overall random agreement probability, Pe = 0.04 + 0.2 + 0.12 = 0.36
So, Cohen's Kappa Coefficient, k = (Po - Pe) / (1 - Pe) = (0.9 − 0.36) / (1 − 0.36) = 0.84
So, the law of agreement of two annotators is "0.84". This is an excellent agreement.

IV. CASE STUDY RESULT

4.1 Most Responsive Platform

To found out most responsive platform according to our new dataset, we calculated all commits delay time. Then we categorized these platforms and calculated the median of each platforms delay time. According to this calculation, which platform has the less delay time than all the other platforms, that is the most responsive platform among all platforms. We categorized platforms similarly as our previous datasets categorization. We considered commits less than 1 day to found out consistent commits. From our calculation we got following result,

Table 3 – Delay time of every platform.

Platform	Delay(median)
Android	0.367
Database	0.1
iOS	0.17
Linux	0.3335
MAC	0.25
Web Design	0.1
Web Framework	0.27
Web Programming	0.2
Web Server	0.07
Windows	0.33

We made a graphical representation to understand the difference between all platforms response time. Following figure is the graphical representation of responsiveness of all platforms. Here, we can see that, "Web Server" is the most responsive platform. Also, "Web Design" and "Database" have very less number of response times.
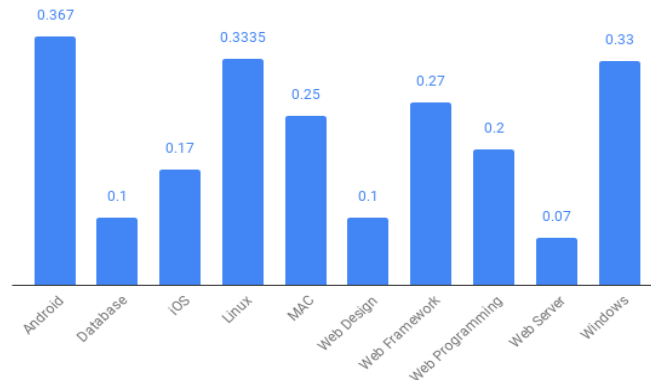
Fig. 4 – Graphical representation of delay time of platforms.
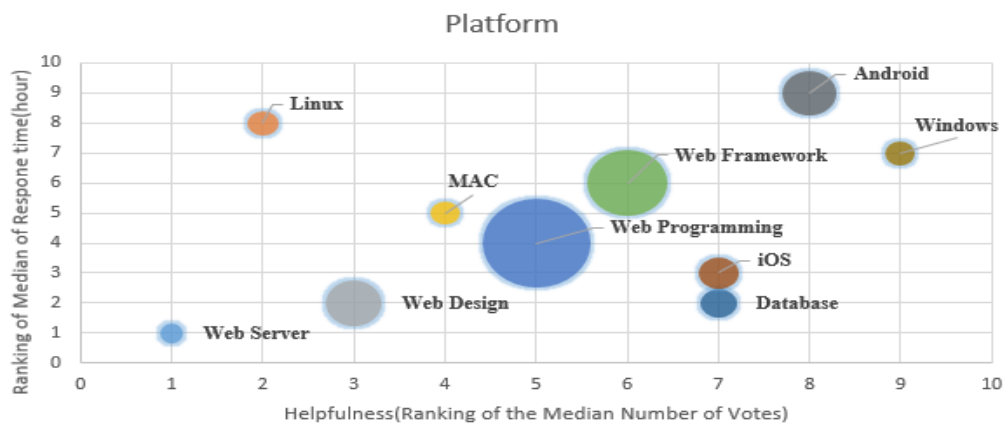
4.2 Best Platform

Now our target is to find out which platform is the best for developers. We consider some parameters such as "number of commits", "median of votes", "median of response time" to decide the best platform. We measure helpfulness ranking according to "median of votes" and responsiveness/delay ranking according to "median of response time".

Table 5 – Rank table for best platform calculation.

Platform	Commits	Answers	Median of Vote (accepted)	Rank (vote)	Median of Response time(hour)	Rank (delay)
Web Server	10	10	133.5	1	0.07	1
Linux	17	17	124	2	0.3335	8
Web Design	58	52	98.5	3	0.1	2
MAC	15	12	69.5	4	0.25	5
Web Programming	216	193	68.5	5	0.2	4
Web Framework	118	101	61	6	0.27	6
Database	23	21	53	7	0.1	2
iOS	28	21	53	7	0.17	3
Android	53	41	37	8	0.367	9
Windows	16	13	31	9	0.33	7

We used bubble plot to show the ranking of helpfulness and response time. Here, X-axis represents the ranking of helpfulness (median of votes), Y-axis represents the ranking of response time. The size of the bubbles represents the number of commits.

From Fig. 5, we can see that "Web Server" is the number 1 platform in both helpfulness and response time. But it has very poor in number of commits. According to this plotting "Web Design" has a good number of votes and response time ranking. It also has a good number of commits.

Fig. 5 – Bubble plot to graphically represent the best platform.

4.3   Most Commits

To find out, in which language most of the developers involved in Stack Overflow, so that we had to know which language had most number of commits. We consider all languages and make percentage like previous calculation.

Table 6 – Commit percentage of every language.

Language	Commits	Percentage(New)
Java	130	20.47
JavaScript	118	18.58
Python	114	17.95
C++	88	13.86
Ruby	55	8.66
PHP	39	6.14
C	31	4.88
C#	28	4.41
Objective-C	18	2.83
Scala	10	1.57
R	3	0.47
Go	1	0.16

We can see that, "Java" has the most commits. So, we can say, developers are mostly involved in "Java". We can also see that percentages of both "JavaScript" and "Python" language are very close to "Java" from this we can say that a huge number of programmers are involved in "JavaScript" and "Python". From this table we can also see that "Scala", "R" and "Go" language programmers are rarely used.
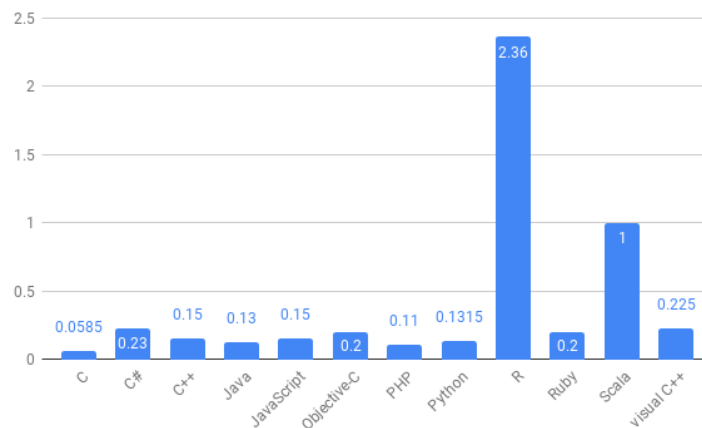
4.4   Most Responsive Language

In this section, we calculated programming languages and their corresponding response time. We calculate median of response time of all languages.

Then, we made a graphical representation of the response time to find out most responsive language. From graphical representation we can easily understand the responsiveness of all programming language.

Table 7 – Delay time of every language.

Language	Delay(median)
C	0.0585
C#	0.23
C++	0.15
Java	0.13
JavaScript	0.15
Objective-C	0.2
PHP	0.11
Python	0.1315
R	2.36
Ruby	0.2
Scala	1
visual C++	0.225

Fig. 6 – Graphical representation of delay of languages.

From this Fig. 6, we can say "C" is the most responsive among all the programming language. We also see that "PHP" and "Java" also require less response time. From this we can say that most of the programmers are involved in "C", "PHP" and "Java". From this graph we can see that "R", "Scala" and "Visual C++" has huge response time to get an answer.

4.5   *Most Accepted Language*

Now, our target is to find out, which language has the most probability to get an accepted answer. To find out that, we calculate the probability of getting accepted answer of all the programming language. But we consider only the popular languages.

Table 8 – Acceptance probability of popular languages.

Language	Total Question	Accepted Answer	Acceptance probability
C	31	28	0.9032258065
C++	88	77	0.875
Java	130	113	0.8692307692
JavaScript	118	108	0.9152542373
PHP	39	35	0.8974358974
Python	114	97	0.850877193

Then we made a pie chart to graphically represent this probability.

From Fig. 7 we can see that, Java has the most probability to get accepted answer. From this graphical representation we can also see that "JavaScript" and "Python" has almost same probability of get a correct and against every commits. If we see carefully we can see that "C", "PHP" and "C++" has less probability to get a correct answer from the Stack Overflow user.
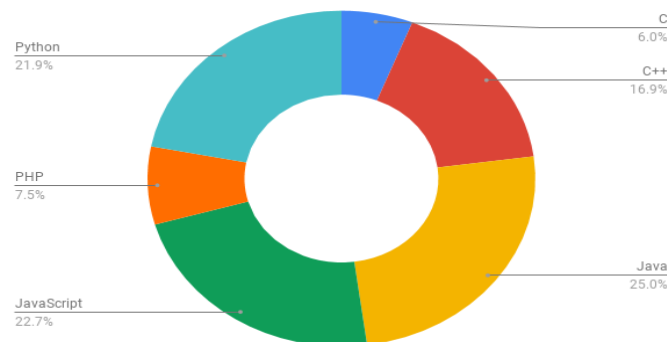
Fig. 7 – Pie chart to graphically represent the acceptance probability of popular languages.

V.   COMPARISON (NEW VS. OLD)

Up to 2015 and after 2015 there are a huge number of changes in dataset. We consider before 2015 dataset as old dataset and after 2015 dataset as new dataset. Now we show the computational differences of the two datasets.

5.1 *Most Responsive Platform*

We considered the same platform categorization for both databases. We got some findings in both databases. Now it's time to see the evolution of response of each platform from 2015 to 2018. To see the evolution of each platform we made a comparison between these two databases. In both cases, we considered commits which have response time less than 24 hours. We calculated the median of all commits response time, so number of commits did not affect the matter for this comparison.

Table 9 – Responsiveness comparison of platforms.

Platform	New Delay(median)	Previous Delay(median)
Android	0.367	0.85
Database	0.1	0.185
iOS	0.17	1.085
Linux	0.3335	0.2
MAC	0.25	0.35

|		
Web Design	0.1	0.125
Web Framework	0.27	0.87
Web Programming	0.2	0.28
Web Server	0.07	2.5
Windows	0.33	0.5

Here we can see all platforms median of previous response time and median of new response time. There is a lot of difference between new and previous response time. Now, we make a graphical representation on these differences to make a proper analysis on them and find out some results from this analysis.

We consider bar chart to represent this comparison graphically. Here, blue bars indicate the new delay time and red bars as well as the previous delay time. We can see some findings from this graphical representation.

Findings:
1.   Almost all platform except "Linux", get some improvement in responsiveness.
2.   "Web Server" has got a huge improvement in responsiveness. It was the most delayed language according to previous dataset and now it is the most responsive according to new dataset.
3.   Though "Web Design" has less response time according to new dataset, it become 2nd for huge improvement of "Web Server".
4.   We can say, developers and crowds become more active in Stack Overflow because almost all platforms improved in responsiveness.
These are the findings from the comparison of responsiveness between new and old dataset.
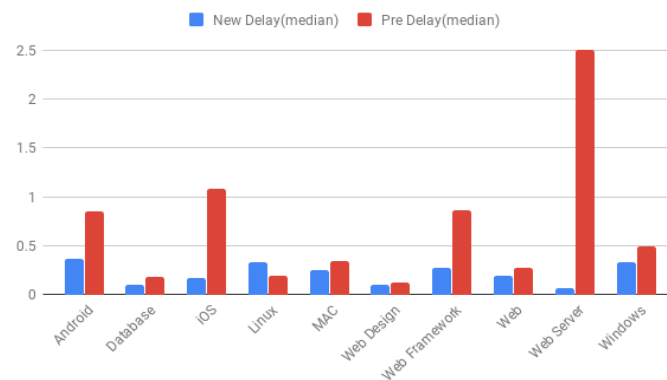
Fig. 8 – Graphical representation of responsiveness comparison of platforms.

5.2   Best Platform
From following table we can easily understand the best platform according to both dataset.

Table 10 – Best platform comparison.

Platform	Median of Vote (New Dataset)	New Rank (vote)	Median of Response time (New Dataset)	New Rank (delay)	Median of Vote (Previous Dataset)	Previous Rank (vote)	Median of Response time (Previous Dataset)	Previous Rank (delay)
Web Server	133.5	1	0.07	1	5	10	2.5	10
Linux	124	2	0.3335	8	22	5	0.2	3
Web Design	98.5	3	0.1	2	51.5	1	0.125	1
MAC	69.5	4	0.25	5	30	3	0.35	5
Web Programming	68.5	5	0.2	4	36.5	2	0.28	4
Web Framework	61	6	0.27	6	15	7	0.87	8
Database	53	7	0.1	2	29	4	0.185	2
iOS	53	7	0.17	3	18	6	1.085	9
Android	37	8	0.367	9	14	8	0.85	7
Windows	31	9	0.33	7	10	9	0.5	6

Findings:
1.   "Web Design" has a good ranking in both dataset.
2.   "Web Server" improves a lot in rank.

5.3 Most Commits

To find out comparison between the most trending programming language of new and old dataset, we consider commit percentage of all programming languages. Following table shows the percentages of commits of all programming languages occurred by developers.

Table 11 – Commit percentage comparison of languages.

Language	Percentage(New)	Percentage(Old)
Java	20.47	21.06
JavaScript	18.58	22.12
Python	17.95	15.91
C++	13.86	11.36
Ruby	8.66	5.15
PHP	6.14	7.73
C	4.88	6.36
C#	4.41	2.88
Objective-C	2.83	4.85
Scala	1.57	1.36
R	0.47	0.61
Go	0.16	0.45

Then we graphically represent to understand this comparison easily.
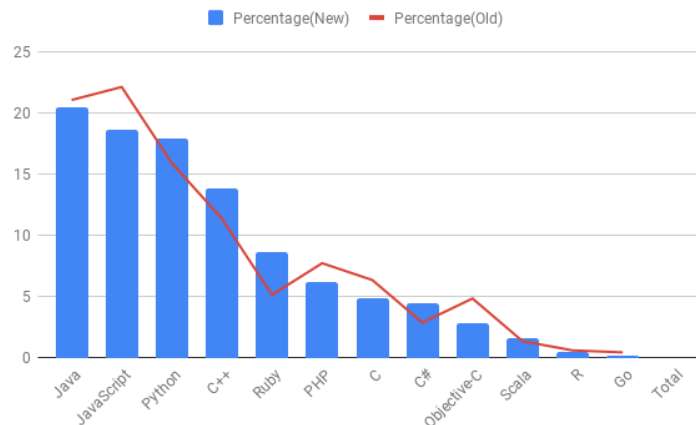
Fig. 9 – Graphical representation of language commit percentatge.

We consider blue bars as new dataset commit percentages and red line as previous dataset commit percentages. Here are some findings from this graphical representation.

Findings:

3. Java has become most trending according to new dataset. Though its' commits percentage is less than previous dataset, it become more trending than other programming languages.

4. JavaScript got less percentage of commits than previous dataset. It has become 2nd among all languages.

5. Developers and programmers involved in Python, C++, Ruby etc. languages much more than before 2015.
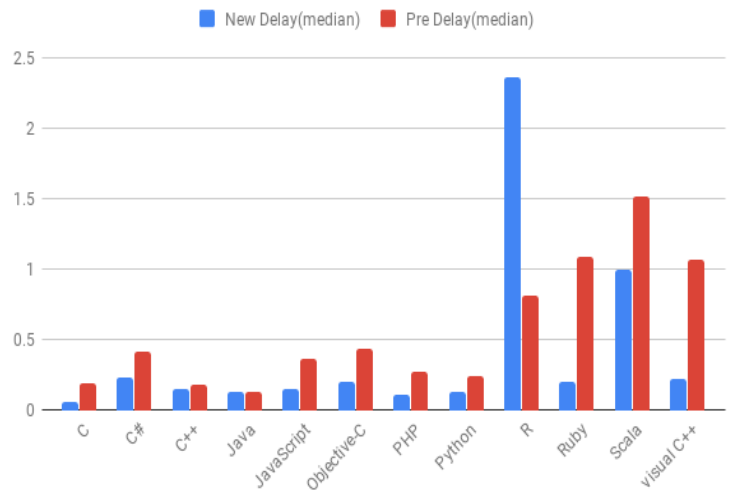
5.4 Most Responsive Language

We calculate responsiveness of all language for both new and previous dataset. To compare responsiveness of all languages we consider median of response time in hour as unit. Here are the median of response time of all language in both dataset.

Table 12 – Responsiveness comparison of languages.

Language	New Delay(median)	Pre Delay(median)
C	0.0585	0.195
C#	0.23	0.42
C++	0.15	0.177
Java	0.13	0.13

JavaScript	0.15	0.367
Objective-C	0.2	0.432
PHP	0.11	0.27
Python	0.1315	0.24
R	2.36	0.81
Ruby	0.2	1.09
Scala	1	1.516
visual C++	0.225	1.07

Now we make a graphical representation of this comparison to understand easily.

Fig. 10 – Graphical representation of responsiveness comparison of languages.

Here, blue bar is for new dataset response time (median) and red bar is for old dataset response time (median).

Findings:
1.   Except "R" language, all languages got improvement in responsiveness.
2.   "C" becomes the most responsive language.
3.   Response time of "Java" remains same.

5.5   *Most Accepted Language*
        We calculate the probability of getting answers of all programming language for both new and previous dataset. But we consider only the common popular languages in both dataset to compare responsiveness of all languages properly. We first calculate the probability according to the number of accepted answer commits and then find out probability for each dataset. Then we make following table where both dataset probability is showed.
Then we make graphical representation to understand the difference properly.

Table 13 – Acceptance probability comparison.

Language	Probability Percentage (new)	Probability Percentage (old)
C	6	15.8
C++	16.9	17.2
Java	25	16.2
JavaScript	22.7	16.7
PHP	7.5	16.4
Python	21.9	17.7
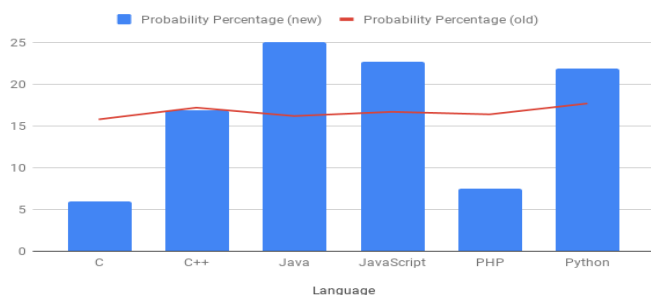
Fig. 11 – Graphical representation of acceptance probability comparison.

Findings:
1.   Acceptance probability of languages changed in a huge number.
2.   "Java" has the most acceptance probability.
3.   "JavaScript" and "Python" also improves a lot.

VI.   OBSERVATIONS AND ANALYSIS

We divided our observation and analysis part into w division. One is for the platforms and one for the languages.

6.1 Observation and Analysis (Platforms)

From figure 6.1 we can see that response time of "Web design" remains very low for both datasets. That means questions related to "Web design" user get fastest answer than other platforms by the crowd. If we consider responsiveness and helpfulness, we can see that the ranking of "web design" is well enough in both condition. Now we understand that Stack Overflow is the most helpful for "Web design". If any programmer and developer face any problem related to "Web design" there is a huge probability to get an accurate answer in a short period of time.

6.2   Observation and Analysis (Languages)

Figure 6.3 indicates that recently most of the commits occur in "java" language and analyzing previous dataset we knew that involvement in "Java" is quite well. So, from this analysis we can say that programmers are mostly involved in "Java" language. If we think about this, we can say that most of programmers and developers are involved in "Java". This diagram also shows that "JavaScript" and "Python" has a huge number of commits. If we consider responsiveness, we can say that "C" has a high responsiveness but for both dataset "Java" shows quite well. From fig 6.4 we can say that probability of getting acceptance answer for "Java" language is excellent. So, without any hesitation we can say that "Java" is the most helpful in "Stack Overflow". Most of the questions are related to "Java" and get the fastest answer. The probability of getting answer in "Java" is very high.

Evolution:

After completing our research work on both datasets, we can see that there is a quite difference between the both results. In our previous result, we saw that crowds are mostly involved in "JavaScript" but in our further research, we found that "Java" replaces "JavaScript". If we consider responsiveness, then "C" overtakes "Java" and probability of getting an accepted answer "Java" shows better outcomes than "Python". If we observe the results in both conditions, we see that "Web development" has large involvement than other platforms. Considering votes, the number of commits, response time we can see that "Web Design" still remains in the same position in both results. As we see the use of "Java" increases that means large number of developers are now involving in Android and desktop applications. Before 2015 few numbers of user use Android and after 2015 the number of Android user increase in a multiplicative order. Nowadays most of the smart phone company use Android is their primary operating system. For this reason, we can say that the development of Android application rises in a simultaneous order. Most of the Android developers mainly prefer "Java" as their main language for Android development. Now we can say that the uses of "Java" increase highly, and crowd get fast response of their "Java" related questions.

VII.   CONCLUSION

In this research, we investigated the most helpful language for a programmer and the most helpful platform for a developer. Nowadays the importance of the Stack overflow is increasing more and more. Most of the cases, our research seems very helpful for those users who are attached to technical tasks like programming and developing. Our research result will lead to the new Stack Overflow users to determine, what they should learn. It provides a better way to take the decision to build an efficient and bright career opportunity.

REFERENCES
[1].   R. Abdalkareem, E. Shihab, J. Rilling, "What Do Developers Use the Crowd For? A Study Using Stack Overflow", IEEE Software, Volume 34 Issue 2, Page 53-60, March 2017.
[2].   Georgios Gousios, "The GHTorent Dataset and Tool Suite", Mining Software Repositories(MSR), Page 233-236, 2013.

[3].   Rabe Abdalkareem, Emad Shihab, Juergen Rilling, "On code reuse from Stack Overflow: An exploratory study on Android apps", Information and Software Technology, Volume 88, Pages 148-158, August 2017.

[4].   A. Barua, S. W. Thomas, and A. E. Hassan, "What are developers talking about? an analysis of topics and trends in stack overflow," Empirical Software Engineering (EMSE), vol. 19, no. 3, pp. 619–654, 2014.

[5].   B. Vasilescu, Y. Yu, H. Wang, P. Devanbu, and V. Filkov, "Quality and productivity outcomes relating to continuous integration in github," in Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering, ser. ESEC/FSE'15. ACM, 2015, pp. 805–816.

[6].   A. Zagalsky, C. G. Teshima, D. M. German, M.-A. Storey, and G. Poo-Caamaño, "How the r community creates and curates knowledge: a comparative study of stack overflow and mailing lists," in Proceedings of the 13th International Workshop on Mining Software Repositories, ser. MSR'16. ACM, 2016, pp. 441–451.

[7].   B. Vasilescu, V. Filkov, and A. Serebrenik, "Stackoverflow and github: Associations between software development and crowdsourced knowledge," in Proceedings of International Conference on Social Computing, ser. SocialCom'13. IEEE, 2013, pp. 188–195.

[8].   C. Rosen and E. Shihab, "What are mobile developers asking about? A large scale study using stack overflow," Empirical Software Engineering (EMSE), vol. 21, no. 3, pp. 1192–1223, 2016 R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[9].   J. Fleiss, "The measurement of interrater agreement," Statistics methods for rates and proportions, pp. 212–236, 1981