# Github Commit Messages by Programming Language

Austin Stephens, Evan Childs, Sierra Johnson

## INTRODUCTION AND MOTIVATION

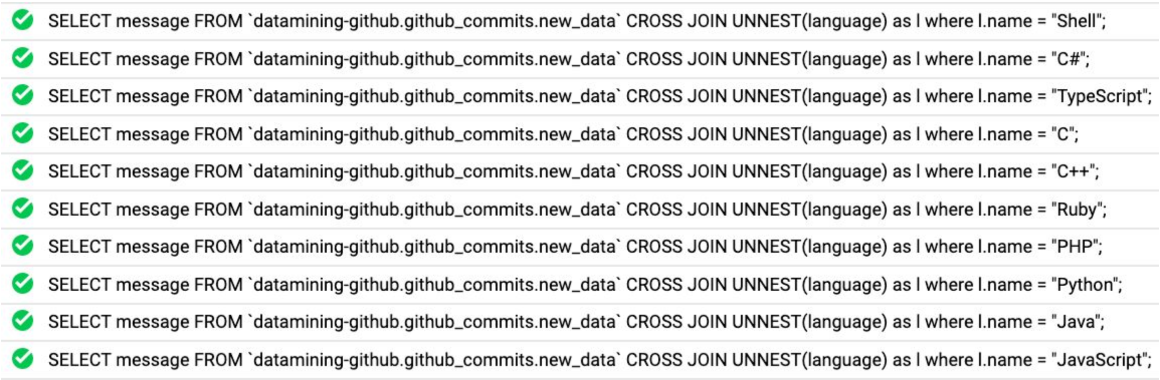
Github is a platform that millions of programmers trust not only with numerous source code files, but also with a reliable source of “version-control”. Anytime a programmer wishes to save their code using Github, they can “commit” the changes to their code, additionally providing a commit message (ideally describing more in depth what changes were made). Because of the versatility of this feature, it is commonly used around the world as well as across many different companies, languages, and programming languages.

Our interest was piqued when we realized that we couldn’t find research relating and comparing these commit messages, not only for commits within the same coding language but between languages as well. Our main goal of the project was to explore the similarities in commit messages within a programming language, as well as how the commit messages of a specific programming language relate to the messages of other programming languages. In the short amount of time we have been working with GitHub as students, we realized that commit messages we wrote would tend to be specific to whichever technology we were using. We theorized that we could find this underlying pattern in commits within the same language, as well as languages that are similar in syntax or content. For example, we theorized commit messages for a project written in C++ would be more likely to include words like “pointer”, “vector”, or “struct” and less likely to use words like “JSON”, “web”, or specific names of libraries available in other languages. We hoped to be able to cluster these languages into similar groupings based on the contents of the commit messages.

## DATA & COLLECTION

Collecting commit messages data was not clear-cut from the beginning as Github doesn’t have an interface on their website. Eventually, we decided that using Google’s BigQuery database for our research was going to be the best option. BigQuery is a powerful tool that let us leverage its traditional database structure to carefully select which data we wanted from the vast 3+ terabyte data collection. Since there are a large number of programming languages used on GitHub, we decided to limit our scope down to the top 10 most used languages on GitHub (in alphabetical order: C, C++, C#, Java, JavaScript, PHP, Python, Ruby, Shell, and TypeScript).

The data set provides a wide variety of attributes for each GitHub commit, many of which could provide further opportunity in the future for other interesting analysis. Categories such as Language, Commit Message, Date and Committer Information were all available, but with the vast amount of data and focus on the commit message contents, we decided to restrict our scope down to just the message contents. We were able to reduce down the 3+ terabyte data store into a more-manageable 32 gigabytes. The following are the specific queries we ran in order to collect our data:



```
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "Shell";
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "C#";
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "TypeScript";
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "C";
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "C++";
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "Ruby";
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "PHP";
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "Python";
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "Java";
SELECT message FROM `datamining-github.github_commits.new_data` CROSS JOIN UNNEST(language) as l where l.name = "JavaScript";
```

Once we had a copy of our data to work with, we realized that the format of the data would not be conducive to the tools we would be applying to it. The separation of the messages in the file were split across quotation marks and were full of symbols and punctuation. We decided that since we wanted to analyze the words contained in the messages, we should remove the symbols and punctuation allowing us to just look at words used. To accomplish this, we created a simple data cleaner program. The cleaner “sanitized” the data by removing punctuation and characters deemed irrelevant (such as commas, periods, etc.) throughout all of our data and changed all the characters to lower-case. For our benefit in creating other programs to analyze the data, the cleaner also put each commit message on its own line, allowing for us to easily tell where one message began and another ended without keeping track of quotation marks. After cleaning all of our data, our overall file size was reduced even further to 28 gigabytes. This was the last refinement performed on the data before we moved on to our actual analysis.

## EXPERIMENT

Using our cleaned data, we applied a combination of many techniques we learned this semester. One of the challenges of working with this data was the shear size. At 28 gigabytes, we had to be careful not only with what tools we applied (as some of them scale to larger data better than others) but also with how we implemented them and with what tools. After some deliberation and testing, we chose to do all of our development work in C++. This offered a 20 fold speed increase over other options, such as python. Since we had so much data to process, these speed increases were crucial in getting the project completed in time.

We decided first that we wanted to examine the similarity of commit messages within a language. We hoped that we could cross reference this data with our later results in order to try and explain any outlying languages. For each language subset of our dataset, we applied FastMinHash using word 2-grams over all the commit messages in that language. We then compared the pairwise similarities of each message to every other message in order to compute the average similarity of the entire language. This turned out to be a huge problem, as comparing n choose 2 similarities for an n of size 10,000,000+ would be too computationally intensive for us to complete. To rectify this, we implemented reservoir sampling over the data set to get a random distribution of 25,000 data points per language, and used those to compute the average similarities. This process was repeated 5 times per language, and those results were averaged out to further ensure that we had a representative sample of our data set.

After we completed the average similarity of each language, we moved on to collecting the most commonly used words in each language. As we had such a large data set, we decided to use CountMinSketch to allow us find the most commonly used words without keeping track of all of the words (as we did not have enough computational resources, namely memory, to do a full count over the data). Using CountMinSketch, we collected the top 100 words from each language subset. In the initial runs using this approach, we found that our results contained a lot of english filler words, such as “a”, “the”, or “as”. We decided that since we were trying to find words that would set apart each language, these common words clouded our results and should be removed. After defining a list of common words that we left out, we ran the CountMinSketch program across all of our data set and collected the top 100 words per language.

The union of all of these words together gave us a common set of 186 words that could describe all of our data set. We used this union set to vectorize our data on a per language basis. We created another program that traversed our data sets and kept an accurate count of each of the words in the union set. These counts were used to create 186 dimensional vectors to represent each language. Finally, we used PCA dimensionality reduction to reduce these vectors into a 2 dimensional space, and clustered them using mean-link clustering.

## RESULTS

### Average Similarities

| Language | Average Similarity |
|---|---|
| C | 0.19% |
| C# | 0.12% |
| JavaScript | 0.10% |
| Python | 0.12% |
| Shell | 0.12% |
| C++ | 0.18% |
| Java | 0.15% |
| PHP | 0.13% |
| Ruby | 0.11% |
| TypeScript | 0.10% |

### Top 10 words per language (see appendix for full lists of 100)

C:

| Word | Count |
|---|---|
| off | 5540012 |
| signed | 5349419 |
| com | 5177128 |
| fix | 3992167 |
| add | 3798336 |
| org | 3743876 |

C#:

| Word | Count |
|---|---|
| added | 939968 |
| add | 818558 |
| fix | 807518 |
| merge | 763511 |
| update | 731888 |
| branch | 531739 |

| | |
|---|---|
| that | 2785870 |
| added | 2641871 |
| not | 2520351 |
| commit | 2433151 |

| | |
|---|---|
| com | 523774 |
| llvm | 493548 |
| fixed | 483640 |
| commit | 476378 |

C++:

| **Word** | **Count** |
|---|---|
| off | 5278135 |
| signed | 5095415 |
| com | 4957214 |
| fix | 3697383 |
| add | 3505808 |
| org | 3362927 |
| that | 2605247 |
| added | 2595762 |
| commit | 2541497 |
| not | 2340569 |

Java:

| **Word** | **Count** |
|---|---|
| merge | 2420757 |
| added | 2350049 |
| add | 2233418 |
| fix | 2043601 |
| update | 1935792 |
| com | 1366245 |
| test | 1274455 |
| commit | 1253104 |
| new | 1207386 |
| branch | 1160457 |

JavaScript:

| **Word** | **Count** |
|---|---|
| update | 8613319 |
| add | 6491657 |
| merge | 6191864 |
| fix | 5692279 |
| added | 5131286 |
| com | 4044482 |
| request | 3105385 |
| branch | 2988407 |
| pull | 2916442 |
| commit | 2768428 |

PHP:

| **Word** | **Count** |
|---|---|
| update | 2165712 |
| fix | 1977942 |
| added | 1929198 |
| add | 1928928 |
| merge | 1825727 |
| com | 1456265 |
| branch | 1160659 |
| org | 1085027 |
| php | 1064526 |
| commit | 1035091 |

Python:

| Word | Count |
|---|---|
| com | 5955902 |
| off | 5558730 |
| add | 5516950 |
| signed | 5372963 |
| fix | 5223149 |
| update | 4681376 |
| merge | 4397191 |
| added | 3812603 |
| org | 3611774 |
| commit | 3455537 |

Ruby:

| Word | Count |
|---|---|
| add | 2515291 |
| update | 2375670 |
| merge | 2118037 |
| fix | 1930579 |
| added | 1700363 |
| com | 1500616 |
| request | 1167814 |
| pull | 1085429 |
| branch | 982834 |
| new | 962079 |

Shell:

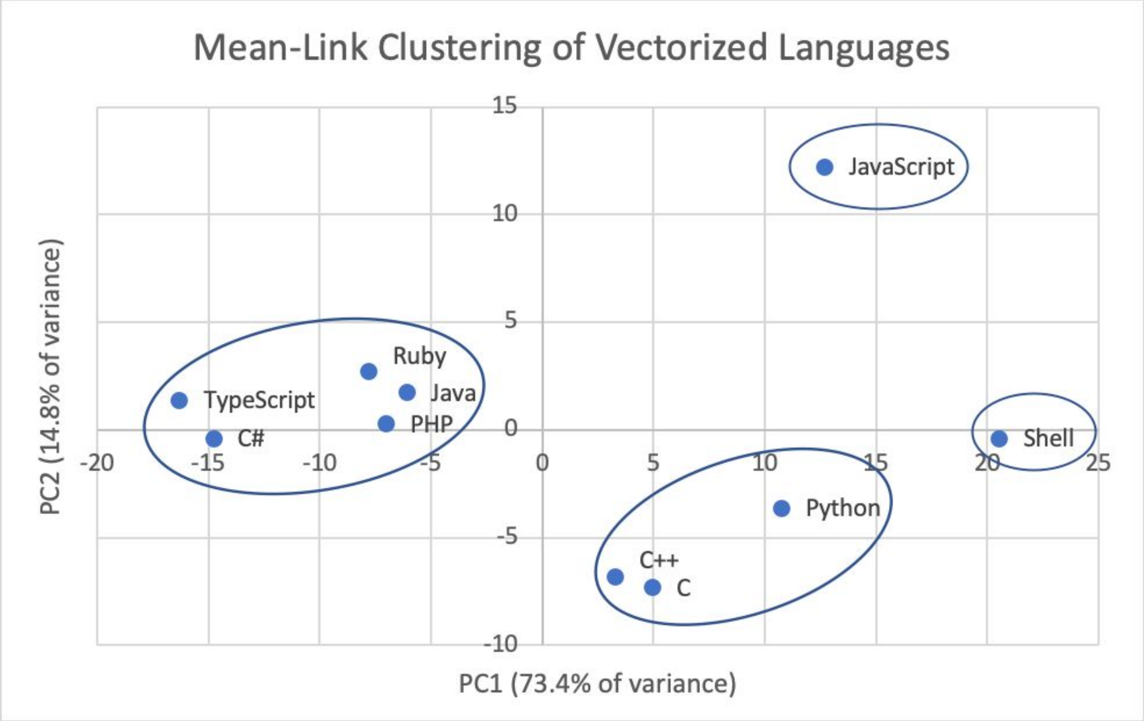
| Word | Count |
|---|---|
| add | 7424005 |
| fix | 7242102 |
| com | 7042376 |
| update | 6404732 |
| off | 6188351 |
| merge | 6146365 |
| signed | 5972214 |
| added | 5066512 |
| commit | 4013272 |
| org | 3964529 |

TypeScript:

| Word | Count |
|---|---|
| fix | 621144 |
| update | 605130 |
| add | 577730 |
| merge | 519678 |
| com | 323513 |
| added | 303575 |
| request | 300410 |
| pull | 285595 |
| https | 238185 |
| github | 228979 |

**Principal Component Variances**

| | |
|---|---|
| PC1 | 73.4% |
| PC2 | 14.8% |
| PC3 | 4.1% |
| PC4 | 2.8% |
| PC5 | 1.9% |



## CONCLUSION

We found it interesting that C and C++ have the highest average similarities of 0.19% and 0.18% respectively. Java stands at 0.15%, while the rest hover around the 0.10-0.12% range. While it makes sense that these “intra-similarity” values don’t really have any bearing on how the language compares to other languages, they are interesting to look at nonetheless. These results imply that C and C++ (and to a lesser extent Java) developers are more likely to use the same kind of words and structured language in their commit messages than developers using other languages.

The commonly used words show us some interesting patterns as well. For example, the word “merge” appears as a top 100 common word in every language. It occurs in increasing order of frequency: C and C++, Python, Shell, PHP, TypeScript, C#, Ruby, JavaScript, and Java. This could imply that the branching and merge features of git are used in increasing frequency in that order across these languages. Similar patterns for other keywords can be observed in the data set. For example, the word “test” appears much more frequently in Java (number 8) than in the other languages (in the range 12-17). This could imply that Java developers are more likely to implement testing into their code bases. “Linux” is relatively common in C and C++ (both number 23) with decreasing frequency for TypeScript, Python, Shell, Ruby and PHP (in order: 26, 39, 47, 93). “Linux” does not appear in the top 100 words for C#, or Java. This supports what we would expect, since Linux platforms are more likely to be targeted by C and C++, and not by C#, which traditionally targets Windows.

The clustering results showed us some surprising groupings. C and C++ having very similar vector representations was unsurprising to us, since the languages are very close to each other it would make sense that their commit messages would have similar contents. Python being clustered in with them was a little surprising, since traditionally C and C++ syntax is so different from Python’s. Python is heavily used in scientific and mathematical computation. C and C++ could be another great option for these fields, because of their speed. We think that perhaps this overlap could explain some of the similarity between these languages’ commit messages.

Another cluster contains TypeScript, C#, Ruby, Java, and PHP. These languages (apart from TypeScript) are traditionally used for development of native applications or backend server development. We think that this could contribute to their being so close together.

Finally, JavaScript and Shell are in their own clusters as outliers from the other two clusters. Shell is the one language on this list that is used more as a tool for conducting tasks across programs or strictly in a computer’s own file system. We think that this difference in use case could affect how the commit messages are written compared to the other languages. JavaScript is an extremely versatile language that can be used for anything from website development, backend development, or native applications. It is also famous for having a very large repository of 3rd party libraries written for it. We think that this inherent generality of scope could cause JavaScript to have such varied commit messages that it became the outlier we see in the above graph.

This project gave us each a chance to learn and see how data mining gives results that would be almost impossible to accomplish without the tools and knowledge we gained from this class. Finding the underlying patterns in the data that group together programming languages was a very captivating idea for us when we started this project, and seeing the data laid out and the goal accomplished is very rewarding. Analyzing the commit message revealed some interesting information, but we hardly scratched the surface of the data available around GitHub. More analysis examining the commit’s file contents and author information would be a great facet to explore using these same techniques we have practiced through this project.

## APPENDIX

### Table of Work

Evan -

| | |
|---|---|
| BigQuery data collection | 50% |
| Development of data cleaner | 100% |
| Development of PCA / Mean-Link Clustering | 100% |
| Development of CountMinSketch | 60% |
| Development of FastMinHash | 40% |
| Papers | 33% |
| Total Contribution | 40% |

Sierra -

| | |
|---|---|
| Development of FastMinHash | 60% |
| Papers | 33% |
| Consulting Development of Algorithms | 50% |
| Total Contribution | 30% |

Austin -

| | |
|---|---|
| Development of CountMinSketch | 40% |
| Papers | 33% |
| BigQuery data collection | 50% |
| Consulting Development of Algorithms | 50% |
| Total Contribution | 30% |

## 100 Most Frequent Words per Language

C:

| | |
|---|---|
| off | 5540012 |
| signed | 5349419 |
| com | 5177128 |
| fix | 3992167 |
| add | 3798336 |
| org | 3743876 |
| that | 2785870 |
| added | 2641871 |
| not | 2520351 |
| commit | 2433151 |
| merge | 2388353 |
| update | 2206155 |
| use | 2175155 |
| new | 1978735 |
| test | 1771063 |
| change | 1728610 |
| bug | 1711648 |
| code | 1660526 |
| file | 1660426 |
| fixed | 1559700 |
| remove | 1538699 |
| linux | 1486244 |
| patch | 1440566 |
| into | 1395147 |
| branch | 1373564 |
| kernel | 1329884 |
| db | 1314096 |
| support | 1312823 |
| make | 1251493 |
| now | 1242923 |
| more | 1223157 |

C#:

| | |
|---|---|
| added | 939968 |
| add | 818558 |
| fix | 807518 |
| merge | 763511 |
| update | 731888 |
| branch | 531739 |
| com | 523774 |
| llvm | 493548 |
| fixed | 483640 |
| commit | 476378 |
| new | 475130 |
| test | 409673 |
| that | 395384 |
| not | 394932 |
| use | 389158 |
| https | 374726 |
| updated | 364157 |
| master | 343687 |
| code | 339750 |
| file | 338280 |
| remove | 314011 |
| readme | 310492 |
| into | 303199 |
| tests | 300812 |
| change | 277146 |
| now | 275713 |
| cs | 264803 |
| request | 263265 |
| org | 261818 |
| github | 258356 |
| version | 257042 |

| | |
|---|---|
| function | 1221141 |
| error | 1210665 |
| no | 1197827 |
| all | 1159721 |
| http | 1150510 |
| cd | 1109883 |
| version | 1083119 |
| build | 1046894 |
| but | 1041200 |
| only | 1041149 |
| tests | 1035911 |
| reviewed | 1022650 |
| fixes | 1014859 |
| src | 1012741 |
| net | 992011 |
| files | 985801 |
| https | 985146 |
| up | 983939 |
| have | 970444 |
| master | 958253 |
| get | 946693 |
| check | 933589 |
| also | 931915 |
| gmail | 926242 |
| updated | 922826 |
| path | 914431 |
| data | 900555 |
| changes | 877206 |
| request | 871700 |
| html | 869001 |
| trunk | 847784 |
| after | 809026 |
| should | 802719 |

| | |
|---|---|
| md | 257000 |
| more | 255002 |
| pull | 250370 |
| src | 246931 |
| build | 246125 |
| support | 238536 |
| bug | 238288 |
| files | 233889 |
| changes | 233066 |
| trunk | 228437 |
| fixes | 222766 |
| rename | 220400 |
| make | 214400 |
| all | 206397 |
| removed | 204575 |
| revert | 199332 |
| revision | 193313 |
| core | 193076 |
| error | 191164 |
| initial | 187881 |
| function | 180368 |
| up | 177612 |
| former | 176307 |
| type | 176094 |
| no | 173694 |
| issue | 173655 |
| project | 173163 |
| name | 171561 |
| db | 168645 |
| path | 168457 |
| check | 163872 |
| data | 160786 |
| changed | 160637 |

| | |
|---|---|
| instead | 797124 |
| work | 795562 |
| used | 793614 |
| has | 779862 |
| removed | 779251 |
| list | 772559 |
| pull | 772257 |
| revision | 769955 |
| don | 762531 |
| using | 750586 |
| dev | 745770 |
| chromium | 738211 |
| upstream | 738078 |
| llvm | 735798 |
| name | 722719 |
| arm | 720923 |
| type | 718054 |
| size | 715782 |
| time | 708418 |
| config | 707134 |
| cpp | 702526 |
| include | 695256 |
| xhtml | 686573 |
| setmaxreader | 683726 |
| call | 677694 |
| before | 675995 |
| default | 675843 |
| device | 669603 |
| memory | 665452 |
| mode | 665057 |
| case | 664732 |
| intel | 660766 |
| debug | 659570 |

| | |
|---|---|
| server | 159047 |
| only | 157791 |
| sql | 157621 |
| get | 155424 |
| class | 154780 |
| list | 152197 |
| lazygen | 151612 |
| patch | 150930 |
| hg | 149244 |
| have | 147732 |
| instead | 144356 |
| release | 143995 |
| scripts | 143801 |
| work | 143774 |
| create | 138659 |
| default | 137501 |
| also | 137050 |
| using | 136649 |
| don | 135875 |
| log | 135133 |
| message | 134278 |
| cpp | 134205 |
| cd | 133317 |
| updates | 133009 |
| but | 132521 |
| config | 131027 |
| method | 130995 |
| move | 127565 |
| net | 126875 |
| sc | 126010 |
| reviewed | 124369 |
| script | 123736 |
| should | 123534 |

| | |
|---|---|
| lazygen | 658777 |
| just | 655129 |
| driver | 652655 |

| | |
|---|---|
| http | 122088 |
| api | 120799 |
| microsoft | 120373 |

C++:

| | |
|---|---|
| off | 5278135 |
| signed | 5095415 |
| com | 4957214 |
| fix | 3697383 |
| add | 3505808 |
| org | 3362927 |
| that | 2605247 |
| added | 2595762 |
| commit | 2541497 |
| not | 2340569 |
| merge | 2269196 |
| update | 2116475 |
| use | 1989411 |
| new | 1805308 |
| change | 1664123 |
| test | 1646267 |
| bug | 1596311 |
| code | 1562630 |
| file | 1559656 |
| fixed | 1539522 |
| remove | 1421461 |
| linux | 1399063 |
| patch | 1358474 |
| into | 1340367 |
| branch | 1312897 |
| db | 1273597 |
| kernel | 1235654 |

Java:

| | |
|---|---|
| merge | 2420757 |
| added | 2350049 |
| add | 2233418 |
| fix | 2043601 |
| update | 1935792 |
| com | 1366245 |
| test | 1274455 |
| commit | 1253104 |
| new | 1207386 |
| branch | 1160457 |
| fixed | 1094124 |
| java | 1045827 |
| not | 1022644 |
| into | 1019823 |
| bug | 976713 |
| use | 931915 |
| org | 921769 |
| that | 918546 |
| change | 904944 |
| master | 891931 |
| request | 835333 |
| version | 834233 |
| code | 826494 |
| https | 811006 |
| tests | 807466 |
| release | 796899 |
| remove | 796553 |

| | |
|---|---|
| now | 1166584 |
| support | 1166265 |
| function | 1159539 |
| make | 1147017 |
| more | 1132572 |
| no | 1114783 |
| error | 1109625 |
| all | 1092600 |
| cd | 1084974 |
| reviewed | 1005126 |
| build | 990669 |
| tests | 976003 |
| but | 974276 |
| only | 971572 |
| https | 957515 |
| src | 944444 |
| master | 941628 |
| fixes | 932372 |
| net | 929652 |
| files | 921834 |
| up | 917298 |
| have | 907322 |
| gmail | 874345 |
| version | 872961 |
| also | 870182 |
| get | 864701 |
| http | 863496 |
| data | 861121 |
| check | 860567 |
| updated | 842034 |
| request | 840516 |
| changes | 825591 |
| cpp | 805429 |

| | |
|---|---|
| file | 785820 |
| pull | 759936 |
| readme | 759479 |
| updated | 709890 |
| src | 689815 |
| now | 685550 |
| github | 654770 |
| support | 619305 |
| more | 617015 |
| build | 608436 |
| db | 607061 |
| md | 600733 |
| plugin | 590865 |
| removed | 588389 |
| changes | 567726 |
| chromium | 531888 |
| files | 526673 |
| all | 511230 |
| http | 508823 |
| make | 508057 |
| am | 491862 |
| class | 487116 |
| cd | 484001 |
| maven | 482506 |
| initial | 478493 |
| ffa | 475307 |
| fixes | 466772 |
| data | 464572 |
| no | 462262 |
| platform | 457607 |
| error | 447850 |
| up | 444079 |
| name | 436257 |

| | |
|---|---|
| path | 781403 |
| after | 769688 |
| work | 760272 |
| removed | 754991 |
| should | 749988 |
| pull | 745420 |
| used | 745257 |
| has | 736796 |
| llvm | 733625 |
| instead | 732271 |
| chromium | 727485 |
| trunk | 717079 |
| list | 714316 |
| dev | 713238 |
| using | 712014 |
| former | 701886 |
| don | 691901 |
| type | 691484 |
| include | 688521 |
| arm | 682300 |
| size | 680720 |
| upstream | 675248 |
| time | 662874 |
| name | 655840 |
| config | 655158 |
| readme | 652851 |
| revision | 645128 |
| intel | 638029 |
| call | 633800 |
| device | 633588 |
| kroah | 632163 |
| before | 631974 |
| memory | 628696 |

| | |
|---|---|
| issue | 435674 |
| method | 435147 |
| webcore | 422628 |
| expected | 422506 |
| xml | 422068 |
| list | 421327 |
| changed | 419437 |
| edef | 415790 |
| api | 415692 |
| reviewed | 412526 |
| off | 406082 |
| create | 391621 |
| dev | 389699 |
| only | 388531 |
| main | 384259 |
| type | 379915 |
| url | 371113 |
| html | 369807 |
| message | 364839 |
| get | 364255 |
| project | 363825 |
| signed | 359939 |
| check | 356295 |
| adding | 355668 |
| also | 348826 |
| instead | 347682 |
| work | 345049 |
| default | 344324 |
| log | 341287 |
| cpp | 339995 |
| have | 338116 |
| usermoduledownloader | 334315 |
| using | 332789 |

| | |
|---|---|
| case | 624899 |
| platform | 624794 |
| mode | 623159 |
| default | 621550 |
| driver | 619586 |
| debug | 619208 |
| first | 617689 |

| | |
|---|---|
| user | 329486 |
| former | 328766 |
| after | 325509 |
| should | 324942 |
| review | 321224 |
| function | 318983 |
| feature | 318899 |

JavaScript:

| | |
|---|---|
| update | 8613319 |
| add | 6491657 |
| merge | 6191864 |
| fix | 5692279 |
| added | 5131286 |
| com | 4044482 |
| request | 3105385 |
| branch | 2988407 |
| pull | 2916442 |
| commit | 2768428 |
| new | 2624670 |
| test | 2542281 |
| fixed | 2374425 |
| master | 2370755 |
| js | 2360914 |
| https | 2313154 |
| data | 2237946 |
| github | 2214937 |
| into | 2114317 |
| updated | 2101277 |
| readme | 2094422 |
| version | 2062970 |
| remove | 2059044 |

PHP:

| | |
|---|---|
| update | 2165712 |
| fix | 1977942 |
| added | 1929198 |
| add | 1928928 |
| merge | 1825727 |
| com | 1456265 |
| branch | 1160659 |
| org | 1075027 |
| php | 1064526 |
| commit | 1035091 |
| new | 966870 |
| fixed | 943237 |
| test | 898496 |
| not | 894821 |
| that | 886168 |
| bug | 847452 |
| https | 826668 |
| off | 807072 |
| file | 797828 |
| use | 796462 |
| request | 774458 |
| signed | 761405 |
| master | 755517 |

| | |
|---|---|
| use | 2046864 |
| not | 2008115 |
| file | 1985823 |
| md | 1938300 |
| that | 1911038 |
| change | 1791864 |
| off | 1758295 |
| bug | 1700875 |
| code | 1683380 |
| signed | 1638853 |
| tests | 1622057 |
| org | 1617774 |
| page | 1588453 |
| more | 1436864 |
| now | 1385143 |
| json | 1380939 |
| html | 1379619 |
| fixes | 1349195 |
| files | 1340051 |
| changes | 1273537 |
| initial | 1207134 |
| support | 1190691 |
| removed | 1174800 |
| all | 1157618 |
| no | 1139427 |
| build | 1136880 |
| error | 1118007 |
| up | 1110517 |
| user | 1088433 |
| db | 1087332 |
| make | 1077186 |
| release | 1065234 |
| http | 1000263 |

| | |
|---|---|
| remove | 728835 |
| into | 723324 |
| change | 709217 |
| pull | 699342 |
| trunk | 692967 |
| code | 674781 |
| updated | 647051 |
| tests | 630934 |
| db | 623129 |
| readme | 622707 |
| http | 602315 |
| fixes | 589481 |
| github | 556207 |
| patch | 553922 |
| version | 546422 |
| chromium | 541727 |
| now | 525163 |
| llvm | 524792 |
| more | 523432 |
| files | 514470 |
| html | 504369 |
| page | 497185 |
| cd | 489448 |
| no | 486202 |
| src | 484501 |
| changes | 484423 |
| removed | 466665 |
| support | 465572 |
| error | 456064 |
| hipl | 448371 |
| path | 446883 |
| all | 444475 |
| webcore | 437583 |

| | |
|---|---|
| name | 991978 |
| changed | 990594 |
| issue | 981552 |
| api | 979672 |
| package | 975264 |
| index | 971228 |
| config | 967643 |
| create | 962044 |
| css | 957942 |
| adding | 953421 |
| list | 951785 |
| function | 911609 |
| chore | 873260 |
| move | 864309 |
| only | 857567 |
| get | 851407 |
| feature | 841282 |
| link | 839477 |
| former | 838395 |
| default | 835739 |
| src | 835484 |
| path | 831406 |
| plugin | 819313 |
| url | 805587 |
| class | 804434 |
| php | 796697 |
| updates | 792961 |
| first | 792551 |
| dev | 786691 |
| work | 783375 |
| trunk | 782115 |
| node | 775418 |
| have | 767574 |

| | |
|---|---|
| md | 435613 |
| expected | 427450 |
| platform | 422410 |
| make | 419662 |
| move | 404344 |
| js | 404151 |
| url | 403767 |
| name | 400861 |
| user | 400801 |
| css | 398861 |
| revision | 388829 |
| updates | 383100 |
| up | 381593 |
| class | 381532 |
| initial | 376833 |
| data | 375681 |
| dev | 374370 |
| function | 368305 |
| cpp | 367015 |
| config | 363599 |
| list | 361424 |
| only | 360819 |
| issue | 358007 |
| former | 355976 |
| default | 347041 |
| api | 346377 |
| have | 342983 |
| changed | 335390 |
| get | 334608 |
| but | 332613 |
| type | 325241 |
| create | 323420 |
| also | 321295 |

| | |
|---|---|
| type | 763039 |
| log | 756471 |
| cd | 753774 |
| source | 751266 |
| patch | 746266 |
| check | 738778 |
| instead | 735694 |
| developmentkit | 733891 |
| using | 732260 |
| but | 729423 |
| text | 725587 |

| | |
|---|---|
| build | 321254 |
| check | 320972 |
| linux | 319951 |
| auto | 319132 |
| core | 309410 |
| adding | 307137 |
| about | 305713 |
| instead | 305506 |
| reviewed | 304587 |
| form | 303726 |
| disable | 300913 |

Python:

| | |
|---|---|
| com | 5955902 |
| off | 5558730 |
| add | 5516950 |
| signed | 5372963 |
| fix | 5223149 |
| update | 4681376 |
| merge | 4397191 |
| added | 3812603 |
| org | 3611774 |
| commit | 3455537 |
| that | 3089004 |
| not | 2817190 |
| test | 2536323 |
| use | 2503156 |
| new | 2320439 |
| change | 2241767 |
| branch | 2220695 |
| file | 2040692 |
| into | 2002408 |

Ruby:

| | |
|---|---|
| add | 2515291 |
| update | 2375670 |
| merge | 2118037 |
| fix | 1930579 |
| added | 1700363 |
| com | 1500616 |
| request | 1167814 |
| pull | 1085429 |
| branch | 982834 |
| new | 962079 |
| test | 907765 |
| version | 899931 |
| use | 893243 |
| commit | 870472 |
| updated | 851494 |
| md | 845051 |
| that | 822515 |
| remove | 821349 |
| not | 793216 |

| | |
|---|---|
| bug | 1980557 |
| remove | 1955418 |
| fixed | 1954438 |
| code | 1906416 |
| request | 1899104 |
| pull | 1756555 |
| master | 1669092 |
| tests | 1599964 |
| more | 1549869 |
| patch | 1531901 |
| now | 1528194 |
| md | 1496323 |
| https | 1487378 |
| support | 1464754 |
| version | 1442824 |
| py | 1436981 |
| make | 1435984 |
| error | 1401683 |
| linux | 1390084 |
| all | 1381409 |
| readme | 1361786 |
| db | 1352451 |
| no | 1333805 |
| updated | 1328911 |
| function | 1317569 |
| data | 1287762 |
| files | 1280640 |
| fixes | 1271794 |
| kernel | 1252426 |
| up | 1222688 |
| changes | 1192427 |
| github | 1185602 |
| only | 1172442 |

| | |
|---|---|
| file | 779551 |
| off | 776749 |
| master | 775839 |
| https | 748091 |
| into | 743827 |
| signed | 727525 |
| org | 719813 |
| fixed | 718173 |
| github | 703295 |
| readme | 692951 |
| tests | 660678 |
| change | 626028 |
| code | 602636 |
| bug | 593073 |
| page | 558215 |
| more | 544001 |
| now | 500353 |
| html | 497011 |
| support | 477802 |
| changes | 458429 |
| files | 455972 |
| make | 447856 |
| db | 445544 |
| release | 441861 |
| up | 440569 |
| removed | 438242 |
| all | 434141 |
| webcore | 427355 |
| fixes | 426762 |
| platform | 424035 |
| error | 418248 |
| bump | 416368 |
| no | 415349 |

| | |
|---|---|
| get | 1155320 |
| but | 1128037 |
| http | 1105289 |
| cd | 1101949 |
| have | 1098500 |
| build | 1095492 |
| also | 1063847 |
| check | 1060995 |
| reviewed | 1033666 |
| name | 1007942 |
| removed | 1005692 |
| list | 989559 |
| dev | 975153 |
| net | 973394 |
| gmail | 966591 |
| config | 946725 |
| work | 936293 |
| python | 933511 |
| instead | 933210 |
| should | 916923 |
| using | 912604 |
| path | 911019 |
| after | 902857 |
| former | 897502 |
| link | 886929 |
| default | 884481 |
| used | 879620 |
| user | 869323 |
| issue | 864078 |
| has | 863835 |
| page | 853446 |
| revert | 847404 |
| chromium | 845450 |

| | |
|---|---|
| name | 410976 |
| data | 397556 |
| expected | 397511 |
| user | 397362 |
| blog | 396683 |
| config | 396194 |
| api | 389557 |
| initial | 375132 |
| class | 372947 |
| cd | 368557 |
| build | 360908 |
| js | 360596 |
| http | 359651 |
| default | 359100 |
| create | 355502 |
| only | 350500 |
| adding | 348575 |
| changed | 343540 |
| list | 339915 |
| get | 335694 |
| linux | 334054 |
| function | 333560 |
| have | 332954 |
| path | 324160 |
| show | 321970 |
| method | 321512 |
| index | 316830 |
| feature | 316447 |
| micro | 316439 |
| mirroring | 311157 |
| link | 306909 |
| instead | 304724 |
| src | 304361 |

| | |
|---|---|
| first | 843694 |
| src | 832368 |
| don | 824531 |
| time | 819235 |
| type | 817553 |
| changed | 817036 |
| initial | 802182 |
| move | 788902 |
| am | 787006 |
| sidis | 786726 |
| create | 782701 |
| adding | 777821 |
| api | 772520 |
| script | 770241 |
| release | 769124 |

| | |
|---|---|
| patch | 303754 |
| but | 298729 |
| text | 295562 |
| cpp | 295276 |
| check | 293925 |
| installfiltered | 293916 |
| css | 293769 |
| issue | 292734 |
| message | 292676 |
| using | 289629 |
| ruby | 288785 |
| post | 287430 |
| rails | 287000 |
| first | 282917 |
| move | 282342 |

Shell:

| | |
|---|---|
| add | 7424005 |
| fix | 7242102 |
| com | 7042376 |
| update | 6404732 |
| off | 6188351 |
| merge | 6146365 |
| signed | 5972214 |
| added | 5066512 |
| commit | 4013272 |
| org | 3964529 |
| that | 3647935 |
| not | 3553264 |
| use | 3347657 |
| new | 3279088 |
| test | 3263320 |

TypeScript:

| | |
|---|---|
| fix | 621144 |
| update | 605130 |
| add | 577730 |
| merge | 519678 |
| com | 323513 |
| added | 303575 |
| request | 300410 |
| pull | 285595 |
| https | 238185 |
| github | 228979 |
| branch | 220466 |
| test | 219585 |
| version | 192495 |
| remove | 186745 |
| chore | 184938 |

| | |
|---|---|
| branch | 3104555 |
| request | 2834966 |
| change | 2755848 |
| remove | 2740073 |
| fixed | 2730003 |
| file | 2704262 |
| into | 2659806 |
| pull | 2628228 |
| master | 2443416 |
| code | 2437079 |
| bug | 2404965 |
| version | 2176603 |
| tests | 2115820 |
| https | 2086903 |
| more | 2027453 |
| updated | 1986279 |
| now | 1981401 |
| support | 1932597 |
| make | 1847945 |
| md | 1829224 |
| github | 1807991 |
| changes | 1802969 |
| patch | 1801423 |
| error | 1790433 |
| all | 1766248 |
| db | 1761907 |
| fixes | 1751224 |
| no | 1749639 |
| build | 1706011 |
| files | 1613049 |
| linux | 1592713 |
| up | 1572254 |
| readme | 1570403 |

| | |
|---|---|
| into | 180010 |
| new | 177891 |
| master | 174011 |
| tests | 168773 |
| use | 166918 |
| fixed | 164997 |
| readme | 152047 |
| change | 140159 |
| file | 137537 |
| linux | 135285 |
| not | 133051 |
| updated | 132452 |
| commit | 129823 |
| build | 128034 |
| package | 127724 |
| extra | 123554 |
| js | 123195 |
| md | 122916 |
| release | 122114 |
| code | 118317 |
| fixes | 109569 |
| more | 102986 |
| that | 101443 |
| support | 100074 |
| feat | 96332 |
| docs | 94692 |
| json | 93644 |
| now | 93621 |
| changes | 92287 |
| refactor | 90366 |
| error | 89530 |
| files | 87629 |
| bug | 87377 |

| | |
|---|---|
| function | 1558309 |
| data | 1544587 |
| dggxrv | 1483700 |
| only | 1471863 |
| kernel | 1433993 |
| get | 1382463 |
| cd | 1377940 |
| removed | 1373480 |
| but | 1358001 |
| have | 1356403 |
| path | 1353119 |
| check | 1351074 |
| name | 1346569 |
| release | 1321630 |
| config | 1308580 |
| time | 1305359 |
| src | 1300108 |
| also | 1281866 |
| page | 1269794 |
| list | 1255355 |
| instead | 1234222 |
| dev | 1220703 |
| http | 1209903 |
| user | 1194267 |
| gmail | 1191660 |
| work | 1184625 |
| default | 1184543 |
| issue | 1175204 |
| net | 1169628 |
| using | 1163955 |
| changed | 1147077 |
| type | 1145963 |
| should | 1142216 |

| | |
|---|---|
| bump | 83968 |
| api | 83374 |
| node | 83000 |
| microsoft | 82672 |
| up | 82027 |
| component | 81281 |
| types | 80890 |
| config | 79848 |
| issue | 79538 |
| all | 79321 |
| page | 79217 |
| app | 78555 |
| ts | 78011 |
| make | 77536 |
| typescript | 77174 |
| removed | 75875 |
| feature | 75355 |
| dependabot | 74422 |
| off | 74290 |
| dependency | 73599 |
| data | 72045 |
| dev | 69894 |
| no | 69577 |
| type | 69519 |
| initial | 69328 |
| signed | 69210 |
| plugin | 68882 |
| sources | 68824 |
| src | 68320 |
| name | 66733 |
| db | 66489 |
| ev | 66020 |
| components | 65686 |

| | |
|---|---|
| reviewed | 1136924 |
| after | 1122332 |
| link | 1120541 |
| former | 1097954 |
| used | 1081375 |
| don | 1070760 |
| adding | 1069626 |
| script | 1068321 |
| acceptedtaxon | 1052572 |
| usermoduledownloader | 1050095 |
| trunk | 1048645 |
| move | 1045792 |
| has | 1045058 |
| server | 1022253 |
| first | 1020263 |
| revert | 1020120 |
| api | 1010102 |
| revision | 1003284 |
| sidis | 992483 |

| | |
|---|---|
| deps | 65662 |
| nouveau | 65509 |
| list | 65253 |
| user | 65118 |
| fichier | 65099 |
| changelog | 64981 |
| angular | 64662 |
| create | 64113 |
| develop | 62967 |
| supprimé | 62892 |
| move | 61996 |
| default | 61374 |
| dependencies | 60622 |
| preview | 60025 |
| only | 57598 |
| react | 57366 |
| adding | 57312 |
| server | 55641 |
| npm | 55563 |