

Github Events Data

Thomas Gersdorf

December 16, 2013

1 Fundamental and Derived Measures

1.1 Terminology

Issue An issue is a generalized form of everything which should be improved on a piece of software, like for instance bug reports or new feature requests.

Commit A commit is the action of submitting new code to the repository. Typically, commits are made **decentralized**, which means that the developer is typing "commit" on his own computer. One or several commits can be uploaded via a Push.

Push A push is a the action of uploading committed source code to the repository. It's one of the key features of Git's decentralized environment. During a push event, editing conflicts can emerge and then need to be resolved.

Gist Gists are snippets of code to be shared with others, interdependent of a repository. Can be cloned, forked etc. like a repository.

Fork Forks are clones/copies of a repository or Gist into a User's own account. The can be used to parallelize development of the project, the feedback to the original project can be done via *PullRequests*.

PullRequest

Payload Typically describes the contents of an action (like the source code itself, the comments itself, all accompanying information but not the meta-data.)

Detailed information on Github's Event API can be found in <http://developer.github.com/v3/activity/events/types/> .

1.2 Users

Event type	Description	Contents	# for all users in example dataset
IssuesEvent	An issues is opened or closed.	Opened & Closed Issues, Assignee (API)	29368
CreateEvent			116683
PushEvent	Source code is pushed into repository.	Size, SHA Hash	298729
GistEvent			26267
Issue CommentEvent			45166
Commit CommentEvent			7827
ForkEvent			19867
WatchEvent			65193
PullRequestEvent			20818
PullRequest ReviewCommentEvent			3726
FollowEvent			17049
GollumEvent	Update of Wikipages, such as release notes	Page name (e.g. "Release Notes") , Action (created, edited), URL	11781
DeleteEvent	Either a branch or tag is removed.	Delete Type (Branch or Tag), Name of Deleted Object	3785
MemberEvent			2764
DownloadEvent			2576
PublicEvent			553
TotalEvents	Counts all of the above.	-	672152

1.2.1 Constructed measures

Issues_Opened # of issues the user has opened

Issues_Closed # of issues the user has closed

Event_Continuity → TODO

Events_Own Events on user's repository

Events_Active_Involvement Measure how much the user engages in projects of other user or organizations, events on other repositories

Events_Passive_Involment Measures how much other users engage in projects of the user, events on the user's repository by other users

1.3 Repositories

1.3.1 Constructed measures

Distinct_Contributing_Users Number of different users contributing via IssuesEvent, CreateEvent, PushEvent, PullRequestEvent

2 Constructed Leadership Measures

2.1 Envisioning and Goal Orientation

$$LS_{\text{milestone}} = \frac{\# \text{ issue directly assigned to a milestone or release}}{\# \text{ total issues opened}} \quad (1)$$

2.2 Leading People

2.2.1 Open Issue and Assign User/Milestone

Measure:

$$LS_{\text{assign}} = \frac{\# \text{ issue directly assigned to other person}}{\# \text{ total issues opened}} \quad (2)$$

Reasoning: The more people directly assign users to issues, the more they know about the people's capabilities and empower other users to solve this issue.

Control for:

- Total number of issues
- Size of project

Assignee currently only via API from <https://api.github.com/repos/thgersdorf/github-play/issues/1>, see reference.

2.3 Empowering

2.4 Project Governance

2.4.1 Closing issues by other people's commits — Gatekeepers

Measure:

$$LS_{\text{assign}} = \frac{\# \text{ issue directly assigned to other person}}{\# \text{ total issues opened}} \quad (3)$$

Control for:

3 Performance Measures

3.1 Contributors

3.2 Release, Commits, Issue closures

$$\text{Issue closing} = \quad (4)$$

Issue User Effectiveness:

$$IUE = \frac{\frac{\Delta \# \text{ Issues Open}}{\# \text{ Total Issues}}}{\Delta \# \text{ User contributing}} \quad (5)$$

3.3 Downloads, Views etc.

4 Reference

4.1 Assignee from API

Example URL: <https://api.github.com/repos/thgersdorf/github-play/issues/1>

```
1 {
2   "url": "https://api.github.com/repos/kevans91/Chatters/issues/5",
3   "labels_url": "https://api.github.com/repos/kevans91/Chatters/issues/5/labels{/↵
4     name}",
5   "comments_url": "https://api.github.com/repos/kevans91/Chatters/issues/5/comments↵
6     ",
7   "events_url": "https://api.github.com/repos/kevans91/Chatters/issues/5/events",
8   "html_url": "https://github.com/kevans91/Chatters/issues/5",
9   "id": 4059710,
10  "number": 5,
11  "title": "Locked Menu Entries",
12  "user": {
13    "login": "kevans91",
```

```

12     "id": 646318,
13     "avatar_url": "https://gravatar.com/avatar/8d4c2928fe43bb196970867f59ab6b93?d=↵
        https%3A%2F%2Fidenticons.github.com%2F5b22ea77797a2c6e6e0859b128be4d94.png&↵
        r=x",
14     "gravatar_id": "8d4c2928fe43bb196970867f59ab6b93",
15     "url": "https://api.github.com/users/kevans91",
16     "html_url": "https://github.com/kevans91",
17     "followers_url": "https://api.github.com/users/kevans91/followers",
18     "following_url": "https://api.github.com/users/kevans91/following{/other_user}"↵
        ,
19     "gists_url": "https://api.github.com/users/kevans91/gists{/gist_id}",
20     "starred_url": "https://api.github.com/users/kevans91/starred{/owner}/{/repo}",
21     "subscriptions_url": "https://api.github.com/users/kevans91/subscriptions",
22     "organizations_url": "https://api.github.com/users/kevans91/orgs",
23     "repos_url": "https://api.github.com/users/kevans91/repos",
24     "events_url": "https://api.github.com/users/kevans91/events{/privacy}",
25     "received_events_url": "https://api.github.com/users/kevans91/received_events",
26     "type": "User",
27     "site_admin": false
28 },
29 "labels": [
30   {
31     "url": "https://api.github.com/repos/kevans91/Chatters/labels/enhancement",
32     "name": "enhancement",
33     "color": "84b6eb"
34   }
35 ],
36 "state": "open",
37 "assignee": null,
38 "milestone": null,
39 "comments": 0,
40 "created_at": "2012-04-11T06:43:08Z",
41 "updated_at": "2012-04-11T06:43:08Z",
42 "closed_at": null,
43 "pull_request": {
44   "html_url": null,
45   "diff_url": null,
46   "patch_url": null
47 },
48 "body": "Locked menu entries currently clutter up the menus at the top pretty ↵
        badly. They either need to be removed or used.",
49 "closed_by": null
50 }

```

4.2 Raw JSONs

4.2.1 IssuesEvent

```

1  "repository":{"homepage":"http://tdurand.github.com/mapafortaleza",
2  "has_downloads":true,"has_issues":true,"master_branch":"gh-pages","forks":5,"↵
    language":"JavaScript","fork":false,"has_wiki":true,"url":"https://github.com/↵
    tdurand/mapafortaleza","created_at":"2011/09/29 03:27:42 -0700","size":132,"↵
    private":false,"name":"mapafortaleza","description":"Mapa de Fortaleza","owner"↵
    :{"name":"tdurand","open_issues":9,"watchers":13,"pushed_at":"2012/02/29 16:43:30 -0800"↵
    }},
3  "actor_attributes":{"name":"Thibault Durand","company":"@tibbb","gravatar_id":"↵
    cedb88f236afaeed4caea39917ebd0a7","blog":"http://www.thibault-durand.fr","type"↵
    :{"name":"User","login":"tdurand"}},
4  "created_at":"2012/04/11 16:01:27 -0700","public":true,"actor":"tdurand","payload"↵
    :{"number":12,"action":"opened","issue":4075341},"url":"https://github.com/↵
    tdurand/mapafortaleza/issues/12","type":"
5  IssuesEvent"}

```

4.2.2 PushEvent

```

1 {"repository":{"homepage":"","has_downloads":true,"has_issues":true,"
2 forks":1,"fork":false,"has_wiki":true,"url":"https://github.com/brunolopesjn/JConvexHull","created_at":"2012/04/10 18:37:34 -0700","size":208,"private":false,"name":"JConvexHull","description":"Programa Java Swing que resolve o problema do caixeiro viajante (TSP) utilizando o Convexo de Hull","owner":{"name":"brunolopesjn","open_issues":0,"watchers":1,"pushed_at":"2012/04/11 15:01:36 -0700"},"actor_attributes":{"name":"Bruno Lopes Alcantara Batista","gravatar_id":"854bf813e2a302dcc4a059b89c502424","blog":"kariridev.blogspot.com","type":"User","login":"brunolopesjn","email":"brunolopesjn@gmail.com"},"created_at":"2012/04/11 15:01:37 -0700","public":true,"actor":"brunolopesjn","payload":{"head":"8469f67ff919f2a776336bfe61f5ae7f40e76c72","size":1,"shas":["8469f67ff919f2a776336bfe61f5ae7f40e76c72","brunolopesjn@gmail.com","Criado a interface grafica base, de desenho dos pontos, o desenho\ndo plano cartesiano e a ica de abertura de arquivos.","Bruno Lopes Alcantara Batista",true]],"ref":"refs/heads/master"},"url":"https://github.com/brunolopesjn/JConvexHull/compare/ea77edbd70...8469f67ff9","type":"PushEvent"}

```

4.2.3 IssueCommentEvent

4.2.4 CommitCommentEvent

4.2.5 ForkEvent

4.2.6 FollowEvent

4.2.7 GollumEvent

```

1 {"repository":{"homepage":"http://www.sisodb.com","has_downloads":true,"has_issues":true,"forks":4,"language":"C#","fork":false,"has_wiki":true,"url":"https://github.com/
2 danielwertheim/SisoDb-Provider","created_at":"2011/01/31 05:33:16 -0800","size":3472,"private":false,"name":"SisoDb-Provider","description":"SisoDb - Simple Structure Oriented Db","owner":{"name":"danielwertheim","open_issues":0,"watchers":52,"pushed_at":"2012/04/11 14:52:42 -0700"},"actor_attributes":{"name":"Daniel Wertheim","gravatar_id":"274ce0291206c9d27635a865ac48b5c4","location":"Sweden","blog":"http://daniel.wertheim.se","type":"User","login":"danielwertheim"},"created_at":"2012/04/11 15:01:37 -0700","public":true,"actor":"danielwertheim","payload":{"pages":[{"sha":"909e1049863468ad462655505441bf7bf39dc739","title":"Release notes","action":"edited","page_name":"Release notes","summary":null,"html_url":"https://github.com/danielwertheim/SisoDb-Provider/wiki/Release-notes"}]},"url":"https://github.com/danielwertheim/SisoDb-Provider/wiki/Release-notes","type":"GollumEvent"}

```

4.2.8 DeleteEvent

```

1 {"repository":{"homepage":"","has_downloads":true,"has_issues":false,"master_branch":"master","forks":0,"language":"Shell","fork":true,"has_wiki":true,"url":"https://github.com/travp624/android_device_moto_shadow","created_at":"2012/04/03 01:32:24 -0700","size":6060,"private":false,"name":"android_device_moto_shadow","description":"","owner":{"name":"travp624","open_issues":0,"watchers":1,"pushed_at":"2012/04/11 15:02:07 -0700"},"actor_attributes":{"name":"teamicemods","gravatar_id":"1eb4d77c7c79eccde05a7047fbae4195","location":"wisconsin home of cheese baby","blog":"","type":"User","login":"travp624","

```

```
email":"travp624@gmail.com"},"created_at":"2012/04/11 15:02:07 -0700","public":↵
true,"actor":"travp624","payload":{"ref_type":"branch","ref":"ics"},"url":↵
https://github.com/","type":"DeleteEvent"}
```

4.3 SQL Queries Temp

4.3.1 Generic Event Queries

```
1  # total events involvements:
2  SELECT * FROM `events` WHERE `actor` = 'mitar' AND `repository_owner` != 'mitar'
3  SELECT * FROM `events` WHERE `actor` = 'mitar' AND `repository_owner` = 'mitar'
4
5  # specific Issues properties
6  SELECT payload_number, payload_action, payload_issue from events where type = '↵
IssuesEvent' LIMIT 30;
7
8
9  # user profiling
10 INSERT INTO eng_users (user, TotalEvents, IssuesEvent , CreateEvent , PushEvent , ↵
GistEvent , IssueCommentEvent , CommitCommentEvent , ForkEvent , WatchEvent , ↵
PullRequestEvent , PullRequestReviewCommentEvent , FollowEvent , GollumEvent , ↵
DeleteEvent , MemberEvent , DownloadEvent , PublicEvent, Issues_Opened, ↵
Issues_Closed, Events_Own, Events_Active_Involvement, Events_Passive_Involvement↵
) SELECT ' ' + user_i[0] + ' ' , COUNT(*), SUM(Type = 'IssuesEvent'), SUM(Type ↵
= 'CreateEvent'), SUM(Type = 'PushEvent'), SUM(Type = 'GistEvent'), SUM(Type = ↵
'IssueCommentEvent'),SUM(Type = 'CommitCommentEvent'),SUM(Type = 'ForkEvent'),↵
SUM(Type = 'WatchEvent'),SUM(Type = 'PullRequestEvent'),SUM(Type = '↵
PullRequestReviewCommentEvent'),SUM(Type = 'FollowEvent'),SUM(Type = '↵
GollumEvent'), SUM(Type = 'DeleteEvent'), SUM(Type = 'MemberEvent'), SUM(Type↵
= 'DownloadEvent'), SUM(Type = 'PublicEvent'), SUM(type = 'IssuesEvent' AND ↵
payload_action='opened'), SUM(type = 'IssuesEvent' AND payload_action='closed')↵
, SUM(actor = ' ' +
11 user_i[0] + ' ' AND `repository_owner` = ' ' + user_i[0] + ' '), SUM(actor = ' ' + ↵
user_i[0] + ' ' AND `repository_owner` != ' ' + user_i[0] + ' ') , SUM( actor != ' ' ↵
+ user_i[0] + ' ' AND `repository_owner` = ' ' + user_i[0] + ' ' ) FROM events ↵
WHERE actor_attributes_type = 'User' AND actor_attributes_login = ' ' + user_i↵
[0] + ' ' ON DUPLICATE KEY UPDATE user = ' ' + user_i[0] + ' ' ;
12 # passive involvement:
13 UPDATE eng_users SET Events_Passive_Involvement = ( SELECT COUNT( * ) FROM `events↵
WHERE `repository_organization` IS NULL AND `repository_owner` = '↵
torvalds' AND actor != 'torvalds' ) WHERE eng_users.user = 'torvalds'
```

5 Performance Optimization

Lessons learned:

-