千葉工業大学 社会システム科学部

プロジェクトマネジメント学科

平成25年度 卒業論文

オープンソースソフトウェア開発における役割分担の実態調査

Survey of roles in open source software development

プロジェクトマネジメントコース

矢吹研究室

1042067 関口元基／Genki SEKIGUCHI

|  |  |
| --- | --- |
| 指導教員印 | 学科受付印 |
|  |  |

目次

[第1章 緒論 2](#_Toc378772899)

[1.1 背景 2](#_Toc378772900)

[1.2 目的 2](#_Toc378772901)

[1.3 方法 3](#_Toc378772902)

[1.4 プロジェクトマネジメントとの関連性 3](#_Toc378772903)

[第2章 オープンソーススフトウェアの概要 6](#_Toc378772904)

[2.1 オープンソースとは何か 6](#_Toc378772905)

[2.2 オープンソーススフトウェアを定義する 7](#_Toc378772906)

[2.3 オープンソーススフトウェア誕生の背景 10](#_Toc378772907)

[2.4 ソフトウェア産業におけるオープンソース 11](#_Toc378772908)

[第3章 オープンソーススフトウェアプロジェクトの現状 14](#_Toc378772909)

[3.1オープンソースソフトウェアの種類 14](#_Toc378772910)

[3.2商用ソフトウェアとの比較 15](#_Toc378772911)

[3.3 オープンソースを活用したプロジェクトのメリット 17](#_Toc378772912)

[3.3.1開発者側のメリット，デメリット 17](#_Toc378772913)

[3.3.2 ユーザ側のメリット，デメリット 19](#_Toc378772914)

[第4章 ホスティングサイト 24](#_Toc378772915)

[4.1 ホスティングサイトとは 24](#_Toc378772916)

[4.2 匿名性とプロジェクト参加 24](#_Toc378772917)

[4.3 ツールが一通り揃ったホスティングサイト 24](#_Toc378772918)

[4.4 ホスティングサイトを選ぶ 26](#_Toc378772919)

[4.5 Githubとは何か 27](#_Toc378772920)

[4.5.1バージョン管理システム「Git（ギット）」 29](#_Toc378772921)

[4.5.2 履歴を管理する「リポジトリ」 30](#_Toc378772922)

[4.6.3 変更を記録する「コミット」 30](#_Toc378772923)

[4.5.4 リポジトリに変更を記録する「ワークツリー」と「インデックス」 31](#_Toc378772924)

[4.5.5 リポジトリの共有「プッシュ」と「クローン」と「プル」 31](#_Toc378772925)

[4.5.6 変更履歴の統合「変更履歴のマージ」 32](#_Toc378772926)

[4.5.7 変更履歴の統合「競合の解決」 32](#_Toc378772927)

[5. API 34](#_Toc378772928)

[5.1 APIについて 34](#_Toc378772929)

[5.1.1 APIを公開する・公開されるメリット・デメリット 34](#_Toc378772930)

[5.1.2 現在公開されているAPI 34](#_Toc378772931)

[5.2 GitHubのAPI 35](#_Toc378772932)

[第6章 実態調査 38](#_Toc378772933)

[6.1 調査目的 38](#_Toc378772934)

[6.2 手法 38](#_Toc378772935)

[6.3 調査対象 38](#_Toc378772936)

[6.4 調査環境 39](#_Toc378772937)

[6.5 活動ログの収集 39](#_Toc378772938)

[6.5.1 手入力で個人のログを収集した場合 40](#_Toc378772939)

[6.5.2 APIで個人のログを収集した場合 41](#_Toc378772940)

[6.5.3 手入力とAPIで比較 43](#_Toc378772941)

[6.5.4 使用したツール 44](#_Toc378772942)

[6.5.5 LearnBoost 47](#_Toc378772943)

[6.5.6 Bower 47](#_Toc378772944)

[6.5.7 Adobe Systems 49](#_Toc378772945)

[6.5.8 grunt 50](#_Toc378772946)

[6.5.9 Resque 51](#_Toc378772947)

[6.5.10 jekyll 52](#_Toc378772948)

[6.6 活動ログの分析 52](#_Toc378772949)

[6.6.1 Rとは 52](#_Toc378772950)

[6.6.2 主成分分析とは 53](#_Toc378772951)

[6.6.3 使用したツール 54](#_Toc378772952)

[6.6.4 LearnBoost 58](#_Toc378772953)

[6.6.5 Bower 60](#_Toc378772954)

[6.6.6 Adobe Systems 62](#_Toc378772955)

[6.6.7 grunt 64](#_Toc378772956)

[6.6.8 Resque 66](#_Toc378772957)

[6.6.9 jekyll 68](#_Toc378772958)

[第7章 結論 71](#_Toc378772959)

[7.1 結果・考察 71](#_Toc378772960)

[7.2 今後の課題と発展 72](#_Toc378772961)

[7.3 結論 72](#_Toc378772962)

**第1章**

**緒論**

# 第1章 緒論

## 1.1 背景

オープンソースソフトウェア(以下 OSS)を利用したプロジェクトが増えてきている[1]．オープンソースソフトウェアとはソフトウェアの設計図にあたるソースコードを，インターネットなどを通じて無償で公開し，誰でもそのソフトウェアの改良，再配布が行えるようにすることでありそのようなソフトウェアの名称である．オープンソースのソフトウェア開発のためには，バージョン管理システムやWiki，バグ追跡システム，メーリングシステムがよく利用される．これらのツールをまとめて提供するのがホスティングサイトであり，その代表的なものにGitHubがある[2][3]．

GitHubとはプログラマのためのソーシャルネットワーキングサイトともいい，今まで主流であったリポジトリホスティングとは大きく異なり，プロダクトの世界だけに情報が閉じていたが，開発者が公開しているソースコードはすべて閲覧できるようになっており，どのリポジトリに誰が興味を持っているのを把握できるようになった．これにより，人に目を向けられるようになったことがGitHubの大きな特徴である．GitHubのAPIを使用することによりプロジェクトの活動の様子やプロジェクトメンバの活動ログを収集することができる．

GitHubのAPIには以下のようなものがある[4]．

* ForkEvent : フォークを行ったイベント．自分のアカウント内に既存のリポジトリの複製をつくったという活動のログ．
* DeleteEvent : デリートを行ったイベント．プロジェクトで行われていたイベントを削除したという活動のログ．
* IssuesEvent : Issuesを行ったイベント．プロジェクトメンバに限らず，第三者もIssuesを発行したという活動のログ．
* IssueCommentEvent : Issuesにコメントを行ったイベント．プロジェクトメンバに限らず，第三者もIssuesにコメントしたという活動のログ．
* PushEvent : プッシュを行ったイベント．変更履歴をアップロードした活動のログ．
* PullRequestReviewCommentEvent : プルリクエストにコメントを行ったイベント．管理者がプルリクエストにコメントをした活動のログ．
* WatchEvent : スター（お気に入り）をしたイベント．自分が気になっているリポジトリにスターを付けたという活動のログ．

このようなGituHubのAPIを使い，活動ログを取得・解析し，結果を解釈することによって，今まで明らかになっていなかったOSSプロジェクトにおける各メンバの役割の分担状況を明らかにすることが期待できる．

## 1.2 目的

GitHubなどのネット上で公開されているオープンソースソフトウェアを調査し，ソフトウェア開発の実態を明らかにするとともに，GitHub上で実際に公開されているOSSを調べることで，プロジェクトメンバの役割分担の実態を明らかにする．

## 1.3 方法

以下の手法を用いる．

1. APIを使用し，GitHub上で行われているプロジェクトメンバ全員の活動ログを収集する．
2. 各イベントが何回行われているかの活動ログを，プロジェクトメンバごとの一覧表にまとめる．
3. 統計解析ツールRを使用し，②で得られた一覧表のデータを主成分分析し，結果を解釈する．

## 1.4 プロジェクトマネジメントとの関連性

本研究では実際にGitHub上で行われているOSSプロジェクトの役割分担の実態を調べる研究である．OSSプロジェクトにおけるプロジェクトマネージャーは，各プロジェクトの役割分担について深く理解しておくことが望ましく，本研究の成果はそれに貢献することが期待される．

参考文献

[1] 経済産業省. OSSの現状と今後の課題について. 2011. http://www.meti.go.jp/kohosys/press/0004397/1/030815opensoft.pdf.

[2] GitHub Social Coding. https://github.com,2013.09.11.

[3] 大塚弘記. ソーシャルコーディングの世界. WEB+DBPRESS. 2012, vol. 69, p. 18.

[4] GitHub Developer http://developer.github.com/v3/activity/events/types/#gollumevent, 2014.1.11

**第2章**

**オープンソーススフトウェアの概要**

# 第2章 オープンソーススフトウェアの概要

## 2.1 オープンソースとは何か

オープンソースソフトウェア，と聞いてその特徴，性質を正しく捉え，他人に対して詳細な説明をすることができる人はそう多くないだろう．だが，それは無理もない．このオープンソースソフトウェアという言葉がこの世に誕生したのはわずか数年前，1998年のことである．しかし，名前の響きこそ新しいものだが，オープンソースソフトウェアの持つ性質，概念といったものは遥か昔，それこそコンピュータ創世記から脈々と流れてきているものである．字面を見れば「ソースがオープンになっているソフトウェア」であるということは漠然と理解できるだろうが，その先の段階に理解が進む人は稀であろう．

では，そもそもソースとは何か．これはプログラミング言語と呼ばれる高水準の言葉で書かれている．これはあらゆるソフトウェアを作るにあたり，その設計図とでも言うべきものである．ソフトウェアを作るにはそのプログラムを作らなければならない．しかし，人間とコンピュータでは理解できる言語が異なり，例えば英語や日本語でプログラムを入力してもそれをコンピュータに理解させることは不可能である．では，どうすればいいのかというと，ここでプログラミング言語が使われる．プログラミング言語とは人間の理解できる言語とコンピュータが理解できる言語の中間に位置する言葉であり，主なものとして C 言語や JavaScript といったものがある．プログラマはこのプログラミング言語を使いプログラムを作成する．この段階ででき上がる物がソースコードと呼ばれるものである．このソースコードはプログラミング言語を読むことができない人には何が書いてあるのか分からないだろうが，その言語を学んだ人であれば手間と時間をかければ解読することができる．つまり，このソースコードは一見コンピュータに読ませるための言語に見えるが，まだ人間向けの言語であり，コンピュータがそれを理解することはできない．そこで，このソースコードをコンパイルすることでやっとプログラムとして完成するのである．

コンパイルとは翻訳することを指し，人間向けの言語で書かれたソースコードをコンピュータが理解できる言語に変換することである．このコンパイルを経てソースコードはバイナリコード呼ばれるものに変化する．バイナリコードとは実行プログラムのことである．ソフトウェアを購入するなりダウンロードするなりして手に入れた場合，大概のソフトにはセットアッププログラムがあり，それを起動することでソフトウェアはコンピュータにインストールされる．この際に使われるセットアッププログラム，これがバイナリコードである．このバイナリコードは通常，「0」と「1」の羅列，つまり2進法により記述されており，この段階になりやっとコンピュータに理解できる状態になる．これらを踏まえた上でオープンソースソフトウェアの説明をすると，オープンソースソフトウェアとは「設計図たるソースコードが公開されたソフトウェア」ということになる．ここで注目する点はソースコードを公開する，という点である．前述のとおり，ソースコードとはソフトウェアの設計図にあたる．このソースコードを手に入れた人は技術さえあればそのコードを改変することで元の機能を持った類似製品を低コスト，短時間で作成したり，またはそれを元にしたまったく新しいソフトウェアを作成したりすることができる．つまり，このソースコードは開発者側から見れば本来死守すべきものなのである．

では，なぜこのソースコードを公開するのだろうか．ソースコードを公開することには開発者側，ユーザ側双方にメリットがある．ユーザ側のメリットをあげるとおおよそ三つのメリットがある．プログラム保守，プログラム移植，機能追加の三点である．第一にプログラム保守とは，バグ等を発見した際，開発者の修正を待たずに自身の手でその修正をすることができる．第二にプログラム移植とは同じソフトウェアを他の OS 環境にあるコンピュータに移植する場合，例えば Windows環境から Mac環境へ移植する場合に通常ならそれぞれの機種に対応しているソフトウェアを別途購入しなければいけないが，ソースコードがあれば必要な部分のみ訂正すれば新たに購入する必要はない．第三に機能追加はユーザによる自由なカスタマイズを可能にするためである．

## 2.2 オープンソーススフトウェアを定義する

オープンソースという言葉を最も広義に使う場合，これはソースコードの公開の有無のみを指し，それ以外の条件付けはされていない．逆に，最も狭義の意味で使われる場合にはオープンソースの定義が正しく守られるように管理しているオープンソースイニシアティブ（Open Source Initiative 以下，文中ではOSIと表記）[1]による定義（Open Source Definition以下，文中では OSD と表記）[2]による「OSI認定マーク」の付いたソフトウェア，または OSI認定のライセンスを使用しているソフトウェアのことを言う．また，これ以外にも OSIの認定は受けていないがこれら「オープンソースの定義」に準拠しているソフトウェアもある．このように，様々な解釈のされているオープンソースだが，本論文中でオープンソース，またはOSSという言葉を使った際にはこの最も狭義の意味で使うものとする．

では，OSI による OSD とはどのようなものなのか．これは現在では全十か条の定義となっており，

1. Free Redistribution（自由な再頒布）

2. Source Code（ソースコードの開示）

3. Derived Works（派生著作物に関するライセンス）

4. Integrity of The Author's Source Code（著作者のソースコードの完全性）

5. No Discrimination Against persons or Groups（個人やグループに対する差別禁止）

6. No Discrimination Against Fields of Endeavor（利用分野に対する差別禁止）

7. Distribution of License（ライセンスの継承）

8. License Must Not Be Specific to a Product（特定製品に特化したライセンスの禁止）

9. License Must Not Restrict Other Software（他のソフトウェアを制限するライセンスの禁止）

10. License Must Be Technology-Neutral（テクノロジーニュートラルなライセンス）

となっている．

以下では上記の各項目について説明をしていく．

1. Free Redistribution（自由な再頒布）

この項目では完全に自由な再頒布を許諾することで，オープンソースソフトウェアの普及を目指している．そのため，販売することも許諾しなければならないが，手元を離れたオープンソースソフトウェアに対して更なる使用料や報酬を求めてはならない．

2. Source Code（ソースコードの開示）

OSIライセンスではすべての製品に対して適切な価格での提供を定めている．これに加え，オープンソースソフトウェアを頒布する際には必ずソースコードを含まなければならないということと，ソースコードでの頒布の許可を認めることが必要である．また，この他にもソースコードを意図的にわかりにくくすることは許されていない．これはオープンソースソフトウェアの普及という考えが背景にある．改変を困難にすることでこの普及が妨げられることを防ぐためである．

3. Derived Works（派生著作物に関するライセンス）

OSIライセンスでは以下の項目を遵守することを前提にソフトウェアの改変と派生ソフトウェアの再頒布を認めている．それはソフトウェアに対する自由な改変の許諾，及び派生ソフトウェアに対してもとのソフトウェアと同条件での頒布を許諾しなければならない，というものである．これはソースコードの公開のみではオープンソースとは言えず，自由改変，自由再頒布を認めなければならない，という前提があるためである．

4. Integrity of The Author's Source Code（著作者のソースコードの完全性）

OSIライセンスでは著作者はオリジナルのソースコードと派生版との見分けをつけられるように，「ソースコード+パッチファイル」という形で頒布を認める場合はオリジナルのソースコードと派生版との相違点を明確にし，製作者の名前を記述し， 元のソフトウェアが何かということを分からせなければならない，というものである．これはオリジナルと同名のソフトウェアが出回ると区別がつきにくくなってしまうからである．これにより改悪版などが作られた場合，著作者の評判を落とすことを防ぐことができる．

5. No Discrimination Against persons or Groups（個人やグループに対する差別禁止）

国家によっては外国為替法によりソフトウェアの輸出制限を行っている場合があるが，OSIライセンスではそのような制限をかけることは許されない．これはオープンソースの進化のためには多くの人が触れる必要があり，それを締め出すことは禁止されている．

6. No Discrimination Against Fields of Endeavor（利用分野に対する差別禁止）

オープンソースコミュニティでは学術・商業利用のためのユーザの参加も奨励している．このため，ライセンスで使用分野に制限をかけることを禁止している．

7. Distribution of License（ライセンスの継承）

OSIライセンスによるソフトウェアに付随する権利は再頒布された人すべてに対して適用されなければならない．

8. License Must Not Be Specific to a Product（特定製品に特化したライセンスの禁止）

ここでは特定の頒布物に対して特化した特長をつけることを禁止している．これは Linuxベースのマシンではオープンソースになるが，その他の OSマシンでは使えない，ということを防ぐためである．これは，オープンソースはあらゆる環境においてオープンソースでなければならないからである．

9. License Must Not Restrict Other Software（他のソフトウェアを制限するライセンスの禁止）

これはオープンソースとその他ソフトウェアを同時に利用した場合，その他ソフトウェアの機能を制限するようなことがあってはならない，ということである．

10. License Must Be Technology-Neutral（テクノロジーニュートラルなライセンス）

OSIライセンスではいかなる項目もその技術に対して制限を設けることはできない．これは FTPダウンロードや CD-ROM での頒布，ミラーサイトからの頒布といったクリック・ラッピングとの相性の悪い頒布方法あり，またこのライセンス方法を共用することで再頒布を足止めしてしまう結果を生む危惧が発生してしまう．そのため，OSIに準拠するライセンスは特定の技術に固執することなく，技術的に中立なライセンスを保持する必要があるのである．

さて，ここまで OSD の十項目のライセンスを提示してきたが，ここで注目すべき点が一つある．これは，Linux カーネルなどを筆頭に一般に思われている「オープンソース」像を崩すものである．それは，この OSD は「頒布に際してそのソフトウェアは無料でなければならない」という項目を持たない，という点である．オープンソースソフトウェアの定義の中には有償か無償かという考えは含まれてはいないのである．事実，Linux カーネルをはじめとした多くのオープンソースソフトウェアは無償で頒布されている．しかし，これは「無償で頒布しなければならないもの」だからではなく，頒布元や再頒布元の判断で無償での頒布を行う，と決めているためである．つまり，逆に言えば頒布元，再頒布元が「有償での頒布を行う」と判断を下せば有償頒布を行うことができ，また，今現在では無償頒布されているソフトウェアが将来的に有償になる可能性も常に秘めているのである．また，この Linux カーネルにしてもそのベースとなるオリジナルのものは無償頒布されているが，Red Hat社をはじめとする多くの企業がディストリビュータとして有償による頒布を行っている．これらの企業から頒布を受ける場合には Windows OSとさして変わらない料金を提示されることになる．これは，そのディストリビュータはあくまでもビジネスとして頒布を行い，本来は自らで行う必要があるバージョンアップや修正パッチの確認を企業側が行い，それに付随して各種ソフトウェアとの組み合わせ販売や独自のサポートを行うことで付加価値をつけ，有償での頒布を行える道を作ったのである．

以上のことから，ユーザはオープンソースソフトウェアを無償で手に入れる確証を得ているわけではない，ということになる．

## 2.3 オープンソーススフトウェア誕生の背景

独特の性質をもつ オープンソースソフトウェアだが，その成り立ちはどのようなものであったのだろうか．オープンソースソフトウェアの誕生の背景を語るうえで欠かすことができないのが，フリーソフトウェアの存在である．コンピュータがごく一握りの人々にしか使われていなかった 1970 年代までは，ソフトウェアの権利は至極曖昧なものであった．当時は，ソフトウェアは共有し，相互利用されるのが一般的で，特に，研究者の間では相互利用が盛んに行われていた．しかし 1980 年代に入ると，ソフトウェアを専有化し，そこから収益を得ようという流れが誕生する．そうした商業ソフトウェアの登場により，ソフトウェアの所有権は明確なものとなり，ソースコードも秘匿され，自由な配布や改変を行うことは不可能なものとなった．そうした状況で誕生したのが，先に挙げたフリーソフトウェアというものだ．フリーソフトウェアは，ソフトウェアの専有化の潮流とは逆に，コンピュータ黎明期のような相互利用を促すためにつくられた概念である．

フリーソフトウェアは，ソースコードを公開し，自由に配布や改変を行うことができる．さらに，フリーソフトウェアを改変して生まれたソフトウェアも，フリーソフトウェアでなければならない，というコピーレフトの性質をもっている．このフリーソフトウェアの概念をつくったのが，フリーソフトウェア財団の代表であるリチャード・ストールマンである．ストールマンは The Open Source Definition を提唱しているオープンソースイニシアティブの創設者でもある．フリーソフトウェアは派生ソフトウェアも必ずフリーソフトウェアとしなければならないという，厳密なコピーレフトの性質を持ち，当時の商業ソフトウェアのビジネスモデルとの間に大きな軋轢を生んでいた．そこでストールマンは，フリーソフトウェアの定義をゆるめ，権限の移譲の方式に多様性を認めた．そうして新たにうまれた概念がオープンソースソフトウェアである．オープンソースソフトウェアは，2.2章のとおり，著作者の考え方に合ったライセンスを選ぶことで，コピーレフトをどこまで厳密に採用するかを決定することができる．

このように，オープンソースソフトウェアは，商業ソフトウェアとフリーソフトウェアとの折り合いをつけるために生まれた．

## 2.4 ソフトウェア産業におけるオープンソース

これまではオープンソースソフトウェアの定義と歴史を洗い，その概要に触れてきた．では，その実態はどのようなものであるのだろうか．2.3 章においても触れたように，OSS はそもそも個人が作成し，公開していたフリーソフトウェアを起源としているから，一般的に企業との関係性は薄かった．しかしながら，OSS の発展が進むにつれ，OSS を自らのビジネスに取り入れようとする企業が登場する．その先駆けとなったのが Red Hatである．Red Hat は Linux ディストリビューションである Red Hat Linux の提供を行なった．Linux ディストリビューションとは，オープンソースで開発されている OSである Linux を，一般的な利用者が導入・利用しやすいように様々なソフトウェアを組み合わせ，パッケージングしたもののことを言う．Linux ディストリビューションには様々なものがあるが，Red Hat Linux は，ネットワークサーバやワークステーションといった，主に企業を対象とした構成になっている．Red Hat Linux の最も特徴的なところは，そのビジネスモデルにある．構成するソフトウェアのほとんどが OSSである Red Hat Linux は，基本的に無料での頒布が行われた．その一方で，サポートを有料にし，そこから収益を生み出していた．90 年代の後半では，こうした，OSSの関連商品・サービスを販売して収益を得るというビジネスモデルが多く存在した．その一方で，2000 年代中頃に入ると，オープンソースの開発形態を自社製品に取り込むことで，イノベーションコストを削減することを目的とする企業が登場した．その最も顕著な例が Google である．Google は検索事業からの広告収入を主な収益源としているが，その主幹事業を間接的に支えるために，インターネットブラウザであるGoogle Chrome やモバイルオペレーションシステムである Android といったソフトウェアの開発・提供を行っている．ブラウザや OS といった汎用的なソフトウェアは，多様なユーザニーズに逐一応えていると非常にコストがかさんでしまう．そこでGoogle は，この 2 つのソフトウェアの一部をオープンソースとすることで，ユーザ自身が追加機能の開発や，カスタマイズを行えるようにし，多様なユーザニーズに対応した．

次章ではオープンソーススフトウェアプロジェクトの現状をまとめる．

参考文献

[1] The Open Source Initiative. Attribution. http://www.opensource.org. 2013.9.11

[2] TheOpenSourceDefintion. Attribution. http://www.opensource.org/docs/definition.php. 2013.9.11

**第3章**

**オープンソーススフトウェアプロジェクトの現状**

# 第3章 オープンソーススフトウェアプロジェクトの現状

## 3.1オープンソースソフトウェアの種類

現在，オープンソースソフトウェアの数は増加の一途にある．しかし，オープンソースソフトウェアの開発者は自らの興味や目的によって開発が行われるため，ソフトウェアの種類や分野には多くの偏りがある．そのため，ユーザは自らの判断でニーズにあったソフトウェアを見つけなければならない．この章ではこの数あるソフトウェアの中からメジャーなソフトウェアをいくつか紹介する．

1. Linux

Linuxは1991 年，フィンランドの学生であった Linus Torvaldsを中心に開発されたオープンソースのUNIX 系カーネルである．このLinux はインターネットサーバ用 OSとして大きなシェアを占めており，頒布条件はGNU GPLによって公開されている．Linuxの特徴はバザール方式を始めて採用して開発されたオープンソースソフトウェアである，ということである．Linuxはこの開発方式によって多くのハッカを集めることができ，またそのハッカから提供されたドライバソフト（周辺機器を利用できるようにするのに必要なソフトウェア）を組み込むことができたため，急速に普及していったのである．

2. FreeBSD

FreeBSD はオープンソース UNIX 系 OSとして1993年に発表された．カリフォルニア大学（UC）のBerkeley 校で開発された長い歴史を持つ UNIXのBSD（Berkeley software distribution）版を Intel社のマイクロプロセッサ搭載パソコンで動作するように開発された OSである．日本では当初，IBM互換機でしか動作しないLinuxよりもBSD系OSのほうが普及していた．この BSDはソフトウェアの商用利用に関する制限が最も少ないライセンスの一つであり，カスタムソフトウェアを商用ソフトウェアとして販売することも可能である．また，この FreeBSDは Apple社の最新 OSである Mac OS Xのベースになったことも知られている．

3. Apache

Apacheは現在世界で最も高いシェアを持っている Webサーバである．このオリジナルは NCSA（National Center for Supercomputing Applications）[1]が開発したフリーソフトウェアである．1995 年にこの主要な開発スタッフが Netscape 社に移籍後，オリジナル版のユーザであったハッカが開発を続け，1999 年に Apache Software Foundation [2]を設立．オープンソースソフトウェアの成功例として現在に至っている．

現在，オープンソースソフトウェアの数は増加の一途にある．しかし，オープンソーンソースコミュニティの賛同を得ることができず，開発は大きく遅れることとなった1998年のMozillaプロジェクトの発表から4年がたった2002年になってMozilla 1.0がリリースされた．

## 3.2商用ソフトウェアとの比較

オープンソースソフトウェアと商用ソフトウェアという選択肢がユーザに用意されている．これらの違いは開発工程やソースコードの公開の有無だけではない．使用に当たっての責任の所在や導入に当たってのコストなど，双方の間には大きな違いがある．ここではその相違点を挙げる．

1. 無保証の提供条件

オープンソースソフトウェアは基本的に無保証である．著作権者のみならず，頒布を行っている関係者も含め，ソフトウェアの利用に関して一切の保証を付けずに提供している．

2. 環境への適合性への保証

オープンソースソフトウェアは開発者側で一定の環境の元，動作チェックは行っているが，それはユーザに対して保証するものではない．

3. 品質や性能に関する責任

開発者はオープンソースソフトウェアの品質や性能に関する責任を負わない．ソフトウェアに欠陥があったとしても著作権者は修正の義務を負わず，ユーザが自ら修正や訂正を行うことが求められている．また，そのためにかかるコストはすべてユーザの負担である．

4. 使用結果の責任

ソフトウェアを利用した結果，ユーザに何らかの損害があったとしてもそのことに対する保証は一切行われない．この場合の損害とはデータの損失や他プログラムとの相性問題などを指す．これに対し，商用ソフトウェアはライセンシーをとっての販売を行う代わりにマニュアルに沿った行動に対しての品質保証を行っている．その保証内容は以下のとおりである．

* マニュアルに記載のある内容に従って動作することを保証
* マニュアルに記載のある内容のサポートを行うことの保証
* ソフトウェアが保証規定を満たさない場合は製品の交換，保証，代金の交換のいずれかを行う．

これらの内容は記載のある事柄に関しての保証を行うことを確約するものではあるが，それは裏を返せばマニュアルに記載のないことに関しては一切の保証を行うものではない，ということを意味している．これは商用ソフトウェアに関する問題のひとつとなっている．また，そのソフトウェアを利用して事業を行った結果いかなる損失を企業が負ったとしてもその補償は一切ない．しかし，例外条件として法律にのっとり損失の一部を補填するシステムは存在する．オープンソースソフトウェアと所要ソフトウェアを比較する場合，真っ先に出てくるのはコストの問題である．オープンソースソフトウェアには無償でなければならないという規定はないが，多くのソフトウェアは無償で行われている．例えば雑誌の付録やボランティアのディストリビュータによるダウンロード頒布などである．これに比べ，商用ソフトウェアは有償での販売を行っている．しかし，TCO（Total Cost of Ownership 購入から保守・運用までにかかる総コスト）から見た場合，オープンソースソフトウェアは本当に割安なのだろうか．オープンソースソフトウェアを使用する上で発生するコストにはどのようなものがあるのだろうか．

ここではサーバシステムの構築を例に挙げる．

1. システム構築にかかるコスト

サーバを構築するに当たってはアクセス権の限定や不要ソフトウェアのインストールを行わないなどの注意が必要であり，これらを行うためには初期段階での厳密なシステムの設計が必要となる．ここでそのためのコストがかかってくるのである．

2. インストールコスト

Linuxは Windowsに比べてメーカによる動作確認情報が不足している．この ため，Linux をインストールする場合，使用するパソコンにインストールできるのか，また何かしらの設定を行う必要があるのか，といったことを自ら調べないとならないのである．さらにいえば Linux関係の技術者は雇用コストが一般の技術者に比べ高くつくことが多い．

3. バージョンアップ時のコスト

オープンソースソフトウェアの基本は「早めに公開，頻繁に公開」にある．そのため，ユーザにも頻繁なバージョンアップが要求される．これはつまり，その度に技術者に対する人件費がかかる，ということになる．また，バージョンアップ後にしっかりとソフトウェアが起動するかを確かめる動作確認にもコストがかかる．また，これらの費用以外にも外部サポートを採用している場合にはそのための費用もかかり，さらに言えば企業では社員が Linux環境になれていない場合には社員教育のための費用もかかることになる．では，これらの点を踏まえた上でユーザはオープンソースソフトウェアと商用ソフトウェアのどちらを選ぶのが有益なのだろうか．FSFの Richard M Stallman を始めとした「フリーソフトウェア」信者は「商用ソフトウェア＝悪」，「フリーソフトウェア＝正義」という信念を持っている．彼らはソースコードの公開を行わない商用ソフトウェアを販売している企業を「悪の権化」とし，OS環境をすべてフリーソフトウェアで統一しようとしている．彼らはあらゆる商用ソフトウェアの開発手段を批判している．しかし，この手の批判は穴が多く，的確なものではない．例えば商用ソフトウェアにはバグが多く，ソースコードが複雑である，という批判がある．しかし，まずバグに関しては現在のコンピュータ工学の技術ではバグを完全になくすことは不可能とされている．またソースコードが複雑である，という点に関しては商用ソフトウェアにはバックワードコンパチビリティ（前バージョンとの互換性を持たせること）の必要があるため，ソースコードが複雑にならざるを得ない．つまり，ユーザ本位の開発を行えば，必然的にソースコードは複雑になってしまうのである．

このように「反商用ソフトウェア」派の意見は容易に反論ができる．また，一般的なユーザにとってはソースコード公開の有無はソフトウェアの有益性を語る上で重要視されることはなく，Windows 環境における動作保証がなされ，購入当初の契約でサポートがしっかりと約束されている商用ソフトウェアのほうが安心して利用できる，という面がある．このため，現実的な側面から見ればオープンソースソフトウェアやフリーソフトウェアよりも，商用ソフトウェアを選んだほうが有益であるだろう．

## 3.3 オープンソースを活用したプロジェクトのメリット

ほとんどの人がそれなりにはわかっているようですが，きちんと理解していないためにトラブルになるケースもあるようです．ここでは，オープンソースソフトウェアのメリット，デメリットを説明し，導入するための参考にしていただきたいと思います．

### 3.3.1開発者側のメリット，デメリット

開発者にとってオープンソースのメリットはソフトウェアを自由にカスタマイズできることや，目的にあった環境に移植できることにある．これにより，システム構築やカスタマイズソフトウェアの開発などのビジネスも可能となる．このオープンソースを使ったソフトウェア開発を行うに当たって有効な点が二点ある．まず一点目がオープンソースソフトウェアを使ったソフトウェア開発が行えること，二点目に既存のソフトウェアを二次利用できる，という点である．既存のソフトウェアを利用したいと思った場合，そのソフトウェアが自分の持っている環境に合わなければ使うことは不可能であった．Macintosh と Windows の互換性の問題などがこれにあたる．Macintoshは Windows が普及する以前から GUIに適応した環境を提供していたため，DTP（Desk Top Publishing）やデザインなどの業界で多く利用されていた．こういったソフトウェアは多くが Macintosh用として開発されたため，Windowsでは利用できない，ということが数多くあった．逆に，ワープロや表計算ソフトといったオフィスソフトは Windows が主流であったため，Macintosh では利用できなかった．このため，コンピュータ技術者はこれらのソフトウェアを双方に利用できるものに作り変えたいと思っても，商用ソフトウェアは使用が認められているのはバイナリコードのみであり，ソースコードは非公開であったため，改変をすることは非常に困難であった．さらにもうひとつの大きな障害が著作権の問題である．1980年代に各国でソフトウェアに著作権が認められたため，特定の条件以外では第三者は勝手に改変や再頒布を行うことはできなくなっていたのである．しかし，これがオープンソースソフトウェアの場合，この問題は解決する．オープンソースソフトウェアはソースコードが公開され，最初から移植を含むあらゆる改変の自由が認められている．商用ソフトウェアでは開発会社以外にできない移植もープンソースソフトウェアでは決して簡単な作業ではないが，技術と時間があれば誰でも自由に行うことができるのである．また，まったく新しいアプリケーションソフトウェアを開発したとして，そのソフトウェアを Macintoshないし Windows上で開発した場合，そのソフトウェアを提供したとしてもユーザはそのソフトウェアに対応した OS を購入する必要がある．しかし，元から Linux 上で開発をしていれば，そのソフトウェアと Linux をセットにして頒布することが可能であるため，ユーザに対して余計な費用負担を強いることはない．開発者，ユーザ双方にとって思い通りの環境が低コストで用意することが可能となる．

このように，自由な改変やバグの修正が自らで行えるといった利点がコンピュータ技術者に受け入れられ，利用者が広がる中で商用ソフトウェアを抜きディファクトスタンダードとなったソフトウェアも少なくない．特に，Linux に代表されるサーバコンピュータ用のソフトウェアはオープンソースソフトウェア抜きには考えることができない規模にまで浸透している．特にシェアの大きなソフトウェアとしては Web サーバソフトウェアである「Apache」は全体の 66.04％（2002 年），電子メール配信ソフトウェアである「Sendmail」は全体の 42％（2001 年），DNS（Domain Name System）用ソフトウェアである「BIND」にいたっては全体の 95％（2000 年）と，他の追随を許さないほど普及している．また，Linux 自体も Windows の 49.6％に対して 29.6％（2002年）のシェアを持っている[1]．これらのソフトウェアは現在では Linux のディストリビューションによって Linuxのカーネルとともに頒布されているが，元をただせばこれらのソフトウェアは Linux 用に開発されたものではない．元々は UNIX 用ソフトウェアとして開発されたものがLinux用に移植され，Linux とともに普及をしていったのである．

では，この Linux が普及していった敬意はどのようなものなのか．Linux は現在普及している OS の中では特に歴史の浅い OS である．この歴史の浅い OS が現在の地位に上り詰めた主要な要因は二つあげることが出来る．一つは多くのボランティアの協力がある．プロジェクトに協力したハッカは様々なドライバを開発し，Linux は多くの環境に適応可能な OS となった．この結果，Linux はFreeBSD などそれまで主流であった UNIX 系 OS を押しのけ，UNIX 系 OS のディファクトスタンダードとなった．二つはそれまでの GNUや UNIXの資産と相乗効果が挙げられる．これはすでに多くのユーザの支持を得ていた GCCや Apache，Sendmail やBIND といったソフトウェアを利用できたため，サーバ用 OS としての地位を得た．また，Linuxが普及したことによって付随する形でこれらのソフトウェアもディファクトスタンダードとして広がっていったのである．

現在，FSF では Linux のことを「GNU／Linux」と呼んでいる．これは，Linux のカーネル部分と Linuxパッケージを区別するための名称であり，「GNU／Linux」とはツール群を含むパッケージ全体のことを指している．次に既存ソフトウェアの二次利用であるが，代表的なものではやはり Linux やApache，またデータベースソフトウェアの MySQL や PostgreSQL などがある．これらのソフトウェアは商用ソフトウェアと比べて決して見劣りするものではなく，また市場シェアからみても大きなシェアを持っている．

このような安定性の高いソフトウェアを使えるシステムでは初期コストを大きく削減することができるというメリットがある．また，特定の目的に適合したカスタムソフトウェアを開発する際にも同様の機能を持ったオープンソースソフトウェアを利用することによって開発期間，及びコストを大幅に削減できるのである．この開発期間の削減とはシステム設計，デバッグ，そしてメンテナンスの期間を短縮することが出来るのである．しかし，開発者が気を付けなければならない点としてオープンソースソフトウェアには動作保証がない，ということである．商用ソフトウェアであれば当然指定された環境であれば動作保証があるが，オープンソースソフトウェアにはこういった保証は一切ない．オープンソースソフトウェアはユーザが自由に利用できる代わりに一切の動作保証はなく，ソフトウェアのメンテナンスはすべてユーザが負うことになる．また，中にはいかなる環境においても動作保証自体をまったくしていない，というものもある．このため，開発者は完成したソフトウェアに何らかの問題が起きた場合はその責任すべて開発者に覆いかぶさるのである．

### 3.3.2 ユーザ側のメリット，デメリット

オープンソースソフトウェアに対して一般ユーザが感じるメリットは開発者が感じるメリットと大きくかけ離れていることが多い．これは基本的に一般ユーザに対しソフトウェアを元のまま利用し自ら手を加える，といったことを行わないからである．彼らがオープンソースソフトウェアを利用する理由はソースコードが公開されているからでもなく，また改変の自由が認められているからでもない．商用ソフトウェアに匹敵する高性能ソフトウェアを無償で利用できるからである．

これまでにも述べたようにオープンソースソフトウェアは必ずしも無償で頒布されているわけではないため，「オープンソース＝無償のソフトウェア」という考えは誤りである．しかし，多くのユーザから見ればオープンソースソフトウェアは無償頒布されていて当たり前，と認識されているのである．確かに入手からセッティングまでの肯定を考えればそれまでにかかるコストはディストリビュータに支払うソフトウェアの代金のみであり，企業単位での使用の際もひとつ分の料金を支払えば後は複製の自由があるため商用ソフトウェアに比べはるかに安価ですむ．しかし，後々の保守・運営のことまで考えるとそのコストはユーザの利用目的や技術者の技術によって異なる．ユーザによってはオープンソースソフトウェアを利用することによって余計な手間とコスト負担が増える可能性があるのである．このため，オープンソースソフトウェアを利用するに当たっては初期導入コストのみではなく，その後のサポートコストを含む調査の必要がある．では，ユーザがオープンソースソフトウェアを利用する際にある無償以外のメリットは何があるだろうか．それでは以下にメリット，そしてデメリットを挙げる．

・メリット

1. コストを節約できる．特にデータベースでは大きく節約できる

2. 安定性がある（メジャソフトウェアに限る）

3. セキュリティ・パッチ対応が早い．ソースコードが公開されているため，いざというときは自分で対応できる

4. 機能の追加を自由に行える

・デメリット

1. 自分でサポートを行う場合，技術が伴わない場合がある

2. マイナソフトウェアの場合，開発者がサポートを行わなくなると自分で修正を行わなければならなくなる

3. ディストリビュータが販売している製品まで無償と誤解されることがある

4. Linuxなどに詳しいエンジニアが必要

これを見ると，メリットとしてはコストの削減，安定性の確保，ソースコードの公開と多方面から見たメリットが挙げられる．しかし，逆にデメリットを見ると3.のオープンソースに対する誤解を除けば残りはサポートに関するものばかりである．このことから，目先のコスト削減やソースコードの公開といった自由度の高さを理由に採用し，安易に採用した結果サポート作業に苦心する，といったことが多いようである．また，個人ユーザに関していえば一般的にソースコードを使用することはない．彼らがオープンソースソフトウェアを使用する最大の理由はひとえに「無償」という点にあり，自らソフトウェアを改変し，カスタムソフトウェアを作る，といった高い知識を必要とすることを行うユーザは極めてまれである．

そんななか，個人ユーザがソースコードを利用せざるを得ない機会がひとつだけある．それは，ソフトウェアのアップデートの際である．ソフトウェアをアップデートするには通常，パッチとして頒布されるバイナリコードを利用する．しかし，場合によってはこのバイナリコードの使用ができないことがある．Linuxを例にすると，GNU／Linuxではオープンソースプロジェクトが推奨しているディレクトリ構造とディストリビューションのディレクトリ構造が異なる場合がある．このような場合，バイナリコードを使ったアップデートは行えず，ユーザがソースコードを入手し，自らそのソースコードをコンパイルしなければならないのである．しかし，通常は頒布元のディストリビュータが元のソースコードをコンパイルしたプログラムを提供しているので，やはり個人ユーザがソースコードを触る機会は滅多にないと考えて構わないだろう．

次章ではweb上でプロジェクトを運用するために必要なツールが揃うホスティングサイトを述べる．

参考文献

[1] NCSA. National Center for Supercomputing Applications.

http://www.ncsa.uiuc.edu, 2013.10.22

[2] Apache Software Foundation .

http://www.apache.org, 2013.10.22

**第4章**

**ホスティングサイト**

# 第4章 ホスティングサイト

## 4.1 ホスティングサイトとは

技術的な観点からプロジェクトのウェブサイトを立ち上げることについては， それほど語ることはない．ウェブサーバを起動し，ウェブページを書くことはかなり単純な仕事である．ウェブサイトの主な機能は，明快にプロジェクトの概要を提供し，(バージョン管理システム, バグ追跡システムなどの)他のツールをウェブサイトと結びつけることだ．たとえウェブサーバを設定して起動する技量がなくても，その作業をやってくれる人を探すことは普通難しくない．とはいえ，時間と労力を節約するために，プロジェクトを運営するためのツールが一通り揃っているホスティングサイトがよく好んで使われる．

## 4.2 匿名性とプロジェクト参加

厳密にはホスティングサイトに限った問題ではないが，ホスティングサイトで最もよく見られるのが，ユーザログインの機能に関する苦情だ．ログイン機能自体は十分単純でウェブサイトでは，訪問者が自分のユーザ名とパスワードを登録することができる．登録するとユーザのプロフィールが保存され，プロジェクトの管理者は，ユーザにリポジトリへのコミット権限のような特定の権限を与えることができる．この機能は非常に有用で，ホスティングサイトの主な利点のひとつ．問題は，未登録のユーザにも許可されるべきタスク，特にバグ追跡システム内のファイルアップロードや，既存の問題にコメントをつけるときに，往々にしてログインが必要になってしまっている点にある．こうしたタスクにログインを必要としてしまうと，本来迅速で便利であるべきタスクに参加する敷居が高くなってしまう．勿論，バグ追跡システムにデータを登録した人に連絡を取りたい人もいるが，その場合は(本人が望んだ場合に)電子メールアドレスを入力できるフィールドを設けておけば可能．新しいユーザがバグを発見して報告したいと思ったとして，バグ追跡システムに入力する前にアカウント作成フォームを入力しないといけないとわかれば難しい．ユーザを管理する利点は，通常は欠点に勝るものだが，ユーザを匿名で行動させる選択肢があるなら，全ての読み取り専用のアクションだけでなく，特にバグ追跡システムや，持っているならwikiページのデータ入力についても，ログインしていないユーザに許可するように心がけた方がよい．

## 4.3 ツールが一通り揃ったホスティングサイト

一通りのものが揃ったホスティングサイトには，主に二つの利点がる．ひとつめは，サーバのディスク容量とネットワーク帯域の太さ．つまり，超高速なネットワーク上に，複数のサーバマシンが巨大なラックに収納されている．プロジェクトがどれだけ成功しても，ディスク容量を使い切ったり，ネットワーク接続が使い物にならなくなることはないだろう．ふたつめは,サイトの維持が簡単なところである．ホスティングサイトは，バグ追跡システム，バージョン管理システム，メーリングリスト管理システム，アーカイバや,プロジェクトのウェブサイトを運営するのに必要ものを全て選べる．また，それらは既に設定済みであり，蓄積される全てのデータのバックアップにも注意を自動的に払ってくれるので，多くの決断をする必要はなくなる．フォームに入力し，ボタンを押しさえすればよいのだ．そうすれば巨大なプロジェクト用ウェブサイトを容易に運用することができるようになる．

しかし，他のツールでよいものがあったとしても，ホスティングサイトが選択したツールや設定を受け入れなければならないという欠点もあり，ホスティングサイトは，設定できるパラメータの範囲を普通狭くしているので柔軟性がない．自前でウェブサイトを立ち上げてサーバへの完全な管理権限を持っていた場合にできるなど，きめの細かい制御は厳しい．このことのよい例が，自動生成されるファイルの扱いだ．あるプロジェクトのウェブページは，自動生成されたファイルかもしれない．たとえば，FAQのデータを編集しやすいマスターフォーマットに保存し，それからHTMLやPDF，その他表示用のフォーマットを生成するシステムがあるとする．自動生成されたフォーマットではなく，マスターファイルだけをバージョン管理したいと考える．しかしウェブサイトが他人のサーバに置いてある場合は，マスターファイルが変更された場合にFAQのHTML版を再生成するカスタムフックを設定できないかもしれない．唯一の回避策は，ウェブサイトで表示できるように自動生成されたフォーマットもバージョン管理することだ．もっと大きな影響があるかもしれない．自身が望むほどウェブサイトの見た目を変える権限がない可能性もあり，ホスティングサイトの中には，ウェブページをカスタマイズすることを許可しているが 結局はデフォルトの配置がうんざりするようなやり方で表示されてしまうものもある．

たとえば，SourceForgeというホスティングサイトがある．ここでのプロジェクトの中には，完全にカスタマイズしたホームページがあるが，詳しい情報を参照させるために "Sourceforge上のページ" に開発者を誘導しているものがある．SourceForge のページは，プロジェクトのホームページになるはずのものだが，ユーザに自前のホームページを使わせないようにしています．SourceForge のページには，バグ追跡システムやCVSリポジトリ，ダウンロードサイトへのリンクなどがあるが，不幸なことに，SourceForge のページにはたくさんのプロジェクトとは無関係なリンクも含まれている．ページの一番上部にはバナー広告があるが，アニメーション画像であることもよくある．左側にはプロジェクトに興味がある人には殆ど関係がないリンクが垂直に配置されている．右側にも別の広告がよく配置されており，ページの中央部分だけが本当にプロジェクトに特有の事項に専用の場所になっているが， これらもわかりにくい方法で配置されているので，訪問者が次にどこをクリックしたらいいのかわからなくなってしまうことがよくある．SourceForge のページ設計の裏には，きっと広告のように，SourceForgeの立場からすればまっとうな理由があるのだろう．しかし，プロジェクトの立場から見れば，その結果が理想のウェブページとはかけ離れたものになるかもしれない． SourceForgeを非難するつもりで言っているのではなく，似たような懸念は多くのホスティングサイトにも当てはまる．重要なのは，トレードオフが存在するということであり，プロジェクトのウェブサイトを維持するための技術的な重荷から解放されるが，他人のやり方を受け入れることとひきかえにはじめて恩恵を受けられるのだ．ホスティングサイトがプロジェクトに最適かどうかを決められるのは自分自身であり，仮にホスティングサイトを選んだ場合，プロジェクトの"ホームとなるURL"に自前のドメインを使うことで，自前のサーバに移行する余地を残しておくことも重要である．そうすることで自前のURLをホスティングサイトに転送することもでき，公開されているURLに完全に手を加えたホームページを置くことで，ユーザを洗練された機能を持ったホスティングサイトに誘導することもできる．ウェブサイトのホスティングについて後に別の解を選んだとしても，プロジェクトのURLは変えないように確実に準備をしておくようにしたい．

## 4.4 ホスティングサイトを選ぶ

2013年現時点で，自由につかえるホスティングサイトとして定評のあるのは次の三つだ．これらはすべて，オープンソースライセンスで公開するプロジェクト用に無料で使うことができ，バージョン管理システムやバグ追跡システム，そして Wiki も用意されている (バイナリのダウンロード機能など，その他の機能を用意しているところもある)．

・GitHub

対応しているバージョン管理システムはGit だけだがGitHubはGit を使ったプロジェクトの世界の中心となっており，各種サービスもGitと統合されている．また，GitHub には，プログラムからそのサービスを利用するためのAPIも完備されている．メーリングリストはないが，メーリングリストのサービスは他にもいろいろあるのでたいした問題にはならない．

・Google Code Hosting

対応しているバージョン管理システムはSubversionとMercurial (少なくとも現時点では Git は未対応) で，Wiki やダウンロードエリア，そしてよくできたバグ追跡システムも用意されている．また，API もある．バージョン管理システムには当然ながらそれ自身の API があり，問題追跡システムにも専用のAPI がある．Wiki のコンテンツは バージョン管理システムと連動しているので，スクリプトから編集することができる．また，ダウンロードエリアにもスクリプトによるアップロードという名のAPI が用意されている．また，メーリングリストとしてGoogle Groupsが使える．

・SourceForge

4.3章で述べたように，これは最も古くからあるホスティングサイトで，古くからあるだけあって今でも規模は最大である．非常に多機能であり，インターフェイスも使いやすいものだ．しかし，プロジェクトのページに表示される広告ブロックが気になるという人もいる．現時点では，ホスティングサイトの三巨頭の中で唯一，主要なバージョン管理システム (Git，Subversion，Mercurial そして CVS) のすべてに対応している．SourceForge にはメーリングリストサービスも用意されているが，もちろん他のメーリングリストサービスを使うこともできる．

Apache Software Foundation のように，自分達の任務と既にあるプロジェクトのコミュニティにうまく合っているオープンソースプロジェクトを無料でホスティングしている組織もある．もう少し詳しく言うと，もしバージョン管理システムに Git を使っているのなら GitHub を，それ以外を使っているのなら Google Code Hostingを使うと良いと考えられている．フリーなホスティングサイトの多くは，そのサイト自身を構成するソフトウェアをフリーソフトウェアのライセンスで公開していない(公開しているサイトとしては Launchpad やGitorious そしてGNU Savannah がある)．個人的には，サイト自身を構成するソフトウェアのすべてのコードにもアクセスできるのが理想だと考えている．自分のデータをエクスポートしたり，自分のデータの操作を自動化させたりできるということが重要であり，そのようなサイトであれば決してそこにロックインされることがなくなり，さらにプログラムを通じて拡張可能になる．ホスティングサイトを動かすためのすべてのコードをオープンソースで入手することにはそれなりの価値がある．とはいえ，実際にそのコードを運用するということはほとんどのユーザにとって難しい．これらのサイトには何台ものサーバやカスタマイズしたネットワーク，そして専任スタッフが必要になってくる．いずれにせよ，単にコードを持っているだけでは，サービスを複製したり「フォーク」したりするのは厳しいものがある．大事なのは，自分のデータがしばられてしまわないのを確実にするということだ．Wikipedia に ホスティングサービスの比較 (日本語) の記事がある．オープンソースプロジェクトのホスティング環境の最新情報については，まずここを見るとよい．また，Haggen So は，博士論文 Construction of an Evaluation Model for Free/Open Source Project Hosting (FOSPHost) sites の研究の一環としてさまざまなホスティングサイトを評価した．発表されたのは2005 年だが，2007 年に新たに更新されている．この評価基準は今後も長く使えるものである．

## 4.5 Githubとは何か

Andreessen Horowitzの発表によると，1億ドルという途方もない額がGitHubに投資された．GitHubはその金で一体何をするつもりか，Andreessen Horowitzにとって良い投資だったのか，そんな巨額の投資はGitHubにとって良いことなのか，さまざまな意見や憶測が，Webの至るところに出現した．しかしそもそも，GitHubとはどのようなものなのか，デベロッパにとってなぜそんなに人気があるのか，GitHubは「コードを共有したり公開するためのサービスだ」とか，「プログラマのためのソーシャルネットワーキングサイト」だ，として通説されているがいずれも，GitHubが特別な存在である理由を説明していない．

GitHubはその名のとおり，Gitを使うためのハブである．GitはLinuxの作者Linus Torvaldsが始めたプロジェクトで，GitHubの訓練士Matthew McCulloughの説明によると，Gitは昔からあるバージョン管理システム(version control system, VCS)の一種であり，プロジェクトの改訂履歴を管理し保存する．プログラムのコードに利用されることが圧倒的に多いが，実はWordのドキュメントでもFinal Cutのプロジェクトでも，どんなタイプのファイルでも管理できる．とにかく，コンピュータのプログラムにかぎらず，どんなドキュメントでも，すべての段階の草案やアップデート履歴を保存し管理できるファイルシステムなのだ．Git以前によく使われたCVSやSubversionといったバージョン管理システムでは，一つのプロジェクトの複数の参加者に対し，中央的な“リポジトリ”（repository, 保存庫）が一つだけあった．一人のデベロッパがどこかを書き換えると，その変更が直接，中央的リポジトリにも及ぶ．しかしGitは分散バージョン管理システムなので，一人一人がリポジトリ全体のコピーを保有し，そのコピーに対して変更を行う．そしてその変更がOKとなったら，中央的サーバにその変更を“チェックイン”する．コードを書き換えるたびにサーバに接続しなくてもよいから，これまでよりもきめ細かい変更とその共有が，やりやすくなる．

GitHubは，いろんなプロジェクトのためにGitのリポジトリをホスティングするサービスだ．独自の便利な機能がたくさんある．Gitはコマンドラインツールだが，GitHubはWeb上でグラフィカルなユーザインタフェイスを提供する．セキュリティのためのアクセス制御もあり，また各プロジェクトにwikiやベーシックなタスク管理ツールなど，コラボレーションのための機能も提供する．GitHubの最大の目玉機能が“フォーク(forking)”だ．フォークとは，食器のフォークの先端のように，一つのプロジェクトが複数に分派していくこと．それをGitHubでは，誰かのリポジトリをほかの人がコピーすることによって行う．オリジナルに対するライト(write, 書き込み)アクセスがなくても，それを自分のところで改変できる．自分が行った変更をオリジナルに反映したかったら，オリジナルのオーナーに“プルリクエスト(pull request)”と呼ばれる通知を送る．そしてオーナーは，ボタンをクリックするだけで，その人のリポジトリに対して行われた変更を自分のリポジトリにも取り入れることができる．人のコードを自分の（メインの）コードに導入することを，デベロッパ界隈の用語でマージする(merge, 併合)と言う．マージは複数のデベロッパが関わるプロジェクトにおいて最重要な工程であり，GitHubではそれが安全確実迅速にできる．これら三つの機能フォークとプルリクエストとマージがあるために，GitHubはきわめて強力なサービスなのだ． GitHub以前には，オープンソースのプロジェクトに寄与貢献しようと思ったら，まずそのプロジェクトのソースコードを手作業でダウンロードし，変更をローカルに行い，そして“パッチ(patch, つぎあて)”と呼ばれる変更内容の一覧をプロジェクトのメンテナにメールする．メンテナはそのパッチを評価し，OKとなったらそれらの変更をマージする．パッチは，まったく未知の人から突然送られてくることも多い．しかし，GitHubではネットワーク効果が絶大な威力を発揮する，とPollackは説明を続ける．プルリクエストを送ると，プロジェクトのメンテナは自身のプロフィールを見る．そこには，自身がGitHub上で行った寄与貢献(contributions, コントリビューション)がすべて書かれている．自身のパッチが受け入れられたら，自身はオリジナルのサイト上で信任を取得し，そのことがプロフィールにも載る．メンテナがプロフィールを見るときそれは，自身の評判を知るための履歴書のようなものだ．GitHub上にたくさんの人間とたくさんのプロジェクトがあればあるほど，プロジェクトのメンテナが優れた寄与貢献者(contributors, コントリビュータ)をより見つけやすくなる．これぞまさに，ネットワーク効果の典型だ．パッチは，特定のメンテナの手に渡るだけでなく，公開の場で議論することもできる．GitHubのインターフェイスを使わないメンテナにとっても，GitHubの存在はコントリビューションの管理を容易にしてくれる．オープンソースのサーバサイドJavaScriptプラットホーム，Node.jsのメンテナIsaac Schlueterによると，最終的にはとりあえずパッチをダウンロードしてコマンドラインでマージしている．GitHubのマージボタンは使わないため，敷居を低くして誰もが気軽に参加できるため，オープンソースの開発がより民主化され，若いプロジェクトの成長を助けることが期待される．

GitHubには，一般に公開されるオープンソースのリポジトリ集という顔のほかに，企業にプライベートのリポジトリを売ったり，そのソフトウェアの企業向けのオンプレミスのインスタンスを売るビジネスとしての顔がある．これらのソリューションにはもちろん，オープンなGitHubのネットワーク効果というアドバンテージはないが，コラボレーションの機能はそっくりそのままある．ただしビジネスとしてのそれには，敵もいる．Atlassianが2010年に買収したBitBucketも，競合製品の一つだ（VCSないしSCMS(source code management system)としてGitだけでなく﻿同じく好評なMercurialも使える）．今年の初めにAtlassianが立ち上げたStashは，プライベートでオンプレミスなGitリポジトリをBitBucket的/GitHub的なコラボレーション環境で設営利用できる製品だ．Atlassianはこのほか，バグトラッカーのJiraやwikiソフトのConfluenceなど，デベロッパのためのコラボレーションツールも売っている．Atlassianは2010年にAccel Partnersから6000万ドルを調達している．この6000万ドルという数字を見れば，GitHubの1億ドルの意味もおよそ分かるだろう．同社が未来に向けて何を志向しているのかも分かる．しかし，おそらく特に重要なのは，GitHubがコードに関して図書館のようになっていることだ．Gitを使うときめ細かい粒度の小さい変更の記録ができるので，まったくの初心者でもあるいはエキスパートでも，世界のもっとも偉大なデベロッパたちの足跡を踏みしめることができ，彼らがプログラミングの難題をどうやって解いたかを知ることができる．今回の投資の結果がどう出ようとも，それこそが，GitHubが今後の世界の歴史に残す，驚異的な遺産である．

以降はGitHubの代表的な機能を詳しく述べていく．

### 4.5.1バージョン管理システム「Git（ギット）」

Github上で主に使用する管理システムの一つである．Gitとは，分散型バージョン管理システムの一つで，もともとLinuxのソースコードを効果的に管理するために開発された．

Gitでは，ファイルの状態を好きなときに更新履歴として保存しておくことができる．そのため，一度編集したファイルを過去の状態に戻したり，編集箇所の差分を表示したりすることができる．また，古いファイルを元に編集したファイルで，他人の編集した最新ファイルを上書きしようとすると，サーバにアップロードした時に警告が出るため，知らず知らずのうちに他人の編集内容を上書きしてしまうといったリスクは起こらない．そのためGitでファイルを管理すると，更新履歴がGitに保存されて行く．手元にバックアップ用のファイルのコピーは置く必要がない．

### 4.5.2 履歴を管理する「リポジトリ」

リポジトリとは，ファイルやディレクトリの状態を記録する場所である．保存された状態は，内容の変更履歴として格納されている．変更履歴を管理したいディレクトリをリポジトリの管理下に置くことで，そのディレクトリ内のファイルやディレクトリの変更履歴を記録することができる．まず，Gitのリポジトリは，リモートリポジトリとローカルリポジトリの2種類に分けられる．

・リモートリポジトリ

専用のサーバに配置して複数人で共有するためのリポジトリである．

・ローカルリポジトリ

ユーザ一人ひとりが利用するために，自分の手元のマシン上に配置するリポジトリだ．リポジトリをリモートとローカルの2種類に分けることで，普段の作業はローカルリポジトリを使って全て手元のマシン上で行うことができる．自分のローカルリポジトリで作業した内容を公開したい時は，リモートリポジトリにアップロードして公開し，また，リモートリポジトリを通してほかの人の作業内容を取得することもできる．

### 4.6.3 変更を記録する「コミット」

ファイルやディレクトリの追加・変更を，リポジトリに記録するにはコミットという操作を行う．コミットを実行すると，リポジトリの内では，前回コミットした時の状態から現在の状態までの差分を記録したコミット(またはリビジョン)と呼ばれるものが作成される．このコミットは，時系列順につながった状態でリポジトリに格納されている．このコミットを最新の物から辿ることで，過去の変更履歴やその内容を知ることができるようになっている．これらのコミットには，コミットの情報から計算された重複のない英数字40桁の名前が付けられており，この名前を指定することで，リポジトリの中からコミットを指定することができる．バグ修正や機能追加などの異なる意味を持つ変更は，できるだけ分けてコミットするよう心掛けたい．後から履歴を見て特定の変更内容を探す時に探しやすくなるからだ．また，コミットの実行時にはコミットメッセージの入力を求められる．コミットメッセージは必須となっているため，空のままで実行するとコミットが失敗する．コメントは，他の人がコミットの変更内容を調べる場合や，自分で後から履歴を見直す際に大切な情報となるので，変更内容のわかりやすいコメントを書くようにすることが重要になってくる．

### 4.5.4 リポジトリに変更を記録する「ワークツリー」と「インデックス」

Gitでは，Gitの管理下に置かれた，実際に作業をしているディレクトリのことをワークツリーと呼ぶ．そして，Gitではリポジトリとワークツリーの間にはインデックスというものが存在している．インデックスとは，リポジトリにコミットする準備をするための場所のことを指す．

Gitでは，コミットを実行した時にワークツリーから直接リポジトリ内に状態を記録するのでなく，その間に設けられているインデックスの設定された状態を記録するようになっている．そのため，コミットでファイルの状態を記録するためには，まずインデックスにファイルを登録する必要がある．インデックスを間に挟むことで，ワークツリー内の必要ないファイルを含めずにコミットを行ったり，ファイルの一部の変更だけをインデックスに登録してコミットすることができる．

### 4.5.5 リポジトリの共有「プッシュ」と「クローン」と「プル」

ここまでは，ローカルリポジトリの基本的な使い方を説明してきた．ここからは，リモートリポジトリを使ってローカルリポジトリの変更履歴を共有する方法について説明する．

・Push

リモートリポジトリで自分の手元のローカルリポジトリの変更履歴を共有するには，ローカルリポジトリ内の変更履歴をアップロードする必要がある．そのために，Gitではプッシュ(Push)という操作を行う． Pushを実行すると，リモートリポジトリに自分の変更履歴がアップロードされて，リモートリポジトリ内の変更履歴がローカルリポジトリの変更履歴と同じ状態になる．

・Clone

誰かの変更履歴の入ったリモートリポジトリがあれば，それをまるごと複製して自分の手元で作業をすることができる．リモートリポジトリを複製するには，クローン(Clone)という操作を行います．クローンを実行すると，リモートリポジトリの内容をまるまるダウンロードしてきて，別のマシンにローカルリポジトリとして作成できる．クローンしたローカルリポジトリは変更履歴も複製されているので，元々のリポジトリと全く同じように履歴の参照やコミットをすることができる．

・pull

リモートリポジトリを共有して複数人で作業すると，リモートリポジトリにプッシュしていく．すると，自分のローカルリポジトリに，ほかの人がプッシュした変更内容を取り込む必要がある．リモートリポジトリからローカルリポジトリを更新するにはプル(Pull)という操作を行う．Pullを実行すると，リモートリポジトリから最新の変更履歴をダウンロードしてきて，自分のローカルリポジトリにその内容を取り込むことができる．

### 4.5.6 変更履歴の統合「変更履歴のマージ」

最後にpullを実行してから次のpushをするまでの間に，ほかの人がpushをしてリモートリポジトリを更新してしまっていた場合には，自分のpushが拒否されてしまう． この場合，マージという作業を行なって他の履歴での変更を取り込むまで自分のpushは拒否される．これは，マージを行わないまま履歴を上書きしてしまうと，ほかの人がpushした変更が失われてしまうからだ．マージでは，Gitが変更箇所を自動的に統合してくれるが，自動では統合出来ない場合もある．次の4.5.7では，手動で統合する方法を説明する.

### 4.5.7 変更履歴の統合「競合の解決」

4.5.6で説明したように，マージを行うとGitが変更箇所を自動的に統合する．しかし，自動で統合できない場合もある．それは，リモートリポジトリとローカルリポジトリでファイル内の同じ箇所を変更していた場合で，この場合，どちらの変更を取り込むか自動では判断できないのでエラーが発生してしまう．競合が発生した箇所は，Gitがファイルの内容を修正するので，手動で修正する必要がある．

次章ではGitHubのAPIについて述べる．

**第5章**

**API**

# 5. API

## 5.1 APIについて

API（Application Programming Interface，アプリケーション・プログラミング・インターフェース）とは，コンピュータプログラム(ソフトェア)の機能や管理するデータなどを、外部の他のプログラムから呼び出して利用するための手順やデータ形式などを定めた規約のこと。主に，ファイル制御，ウィンドウ制御，画像処理，文字制御などのための関数として提供されることが多い．

　つまり，アプリケーションをプログラムするにあたって，プログラムの手間を省くために，もっと簡潔にプログラムができるように設定されたインタフェースのことである．

### 5.1.1 APIを公開する・公開されるメリット・デメリット

Web製作者にとっては，他社の膨大なデータベースや機能を無料で利用できるため，Webサイトの開発のコストを大幅に削除できるので，効率的に制作できる．このため，個人や小資本の会社でも，人気のWebサイトを低コストで効率的に作成されることが可能になったのである．また，API提供会社にとっては，自社のみでは考え付かないようなWebサービスや自社のみでは実現できないようなWebサービスを外部の人間が作成してくれるため，API利用サイトから自社のサイトへユーザが流入してくるのを期待できる．さらに，自社のデータベースや機能を間接的に自社のサイトを利用してくれるユーザが増加し，それが結果としてデータを集積するスピードが速まり，競争優位に立つことを期待できるのである．だが，デメリットも発生する．既存のWebサイトにパブリックAPIを追加することは、多くのリスクを伴う．またOSSのようなオープンな環境での編集ポリシーはそのリスクを更に高めてしまう．例えばサードパーティのAPIに頼ったシステムを作るということは，その企業が方針を変えただけでシステムが動かなくなる危険性がある．最近では個人が開発した地図アプリがGoogle Maps APIの突然の仕様変更により，地図の表示がされなくなったり，登録されている場所がずれたりしている．また，Webサービスを公開する背景にはビジネス的な戦略が見え隠れするため，今後APIを使用したサービスが広まった後に，突然広告が表示されるようになってしまうようなユーザにとってもデメリットが発生してしまうことがある．

### 5.1.2 現在公開されているAPI

2013年に入り，官公庁の統計データやNHKの番組情報など、今までなかなか利用できなかったデータがAPIとして扱えるようになった．その中で，特に人気の高いAPIを紹介する．

* 総務省 次世代統計利用システム

国勢調査，人口推計，就業構造，企業統計，物価統計などを公開している．

* NHK番組表

NHKが放映しているテレビジョンの番組表などを公開している．

* 郵便専門ネット

郵便番号から住所検索，郵便番号の簡易存在チェックなどできる．

* GithubAPI

バージョン管理システムとして多く利用されており，イベントを公開している．

* Google Calendar Data API

AtomやRSSを基盤とするGDataと呼ばれるデータフォーマットを利用してGoogle Calendarのカレンダー情報を外部から操作できる．

* Google Maps API

Google Mapsで使用されている地図情報システムを利用し、独自の地図サービスを構築できる．

* Yahoo!検索API

Yahoo!の検索機能をHTTPリクエストを通して実行し，その結果をXML形式で取得できる．

* ニコニコ動画API

ニコニコ動画が提供している機能Getthumbinfo，thumb，getrelationなど

このように現在様々な業種のAPIが多く公開されている．それにより，複数の異なる提供元の技術やコンテンツを複合させて新しいサービスを形作ることができる．

## 5.2 GitHubのAPI

GitHubのAPIとはGitHubの機能を抽出する時に必要なコードであり，GitHubのAPIは機能ごとに設定されている．本研究のため使用するイベントは以下になる．

表1-1 本研究で使用するGitHubのAPI

|  |  |
| --- | --- |
| イベント名 | 意味 |
| CommitCommentEvent | コミットにコメントを行ったイベント． |
| CreateEvent | イベントオブジェクトを行ったイベント．待機関数で実行したプログラムを待機させたという活動のログ． |
| DeleteEvent | デリートを行ったイベント．プロジェクトで行われていたイベントを削除したという活動のログ． |
| IssueCommentEvent | Issuesにコメントを行ったイベント．プロジェクトメンバに限らず，第三者もIssuesにコメントしたという活動のログ． |
| IssuesEvent | Issuesを行ったイベント．プロジェクトメンバに限らず，第三者もIssuesを発行したという活動のログ． |
| PullRequestEvent | プルリクエストを行ったイベント．管理者に更新を依頼する活動のログ． |
| PullRequestReviewCommentEvent | プルリクエストにコメントを行ったイベント．管理者がプルリクエストにコメントをした活動のログ． |
| PushEvent | プッシュを行ったイベント．変更履歴をアップロードした活動のログ． |
| WatchEvent | スター（お気に入り）をしたイベント．自分が気になっているリポジトリにスターを付けたという活動のログ． |
| ForkEvent | フォークを行ったイベント．自分のアカウント内に既存のリポジトリの複製をつくったという活動のログ． |
| GollumEvent | Wikiを作成したイベント．プロジェクトのWikiを作成，更新したという活動のログ． |

以上の11項目をAPIで取得する．このAPIを使用することでメンバがどのような活動をいつ，誰が，誰に，何のプロジェクトに，何をしているのか，などを明確にすることができる．

次章では実際にGitHubで行われているプロジェクトの役割分担を調査する．

**第6章**

**実態調査**

# 第6章 実態調査

## 6.1 調査目的

GitHub上で公開されているOSSをAPIや統計解析ツールのRを使用することで，プロジェクトメンバの役割分担の実態を明らかにする．

## 6.2 手法

以下の手法を用いる．

1. APIを使用し，GitHub上で行われているプロジェクトメンバ全員の活動ログを収集する．
2. 各イベントが何回行われているかの活動ログを，プロジェクトメンバごとの一覧表にまとめる．
3. ②で得られた一覧表のデータを主成分分析し，結果を解釈する．

## 6.3 調査対象

・「LearnBoost」

2011年5月1日から2013年12月24日までのデータを調査の対象に設定．

・「Bower」

2012年6月19日から2013年12月24日までのデータを調査の対象に設定．

・「Adobe Systems」

2010年9月24日から2013年12月24日までのデータを調査の対象に設定．

・「grunt」

2012年5月4日から2013年12月24日までのデータを調査の対象に設定．

・「Resque」

2013年1月11日から2013年12月24日までのデータを調査の対象に設定．

・「jekyll」

2011年10月19日から2013年12月24日までのデータを調査の対象に設定．

## 6.4 調査環境

本研究は以下の調査環境で行っている．

・VirtualBox

・Ubuntu 12.04

・Python

・api.py

・R x64 3.0.2

## 6.5 活動ログの収集

今回，Github上で行われているプロジェクトのデータを収集するためにAPIを活用する．APIの詳しい説明は第5章で述べる．

本研究に取り組む前の練習実験として，特定の個人を選んでの活動（コミット，Issueの操作）のログをとって実態に触れるとともに，手入力により得たデータとAPIででたデータを見比べ，互いのデータがあっているか，どちらが性格なのかを検証する作業を行う．また，そのログを使いExcelのピポットテーブルで表にしCSV形式に直せるようまとめる．特定の個人は矢吹太郎准教授のGitHub内でのtaroyabukiの活動のログyabuki-20131119.txtを対象とし，調査項目はcommit，Issuesとする．活動期間は2013年7月21日から2013年11月9日までとする．

### 6.5.1 手入力で個人のログを収集した場合

表1-1 手入力で得られた活動ログをエクセルで表示した場合

|  |  |  |
| --- | --- | --- |
| 日時 | 何をしたか | commit,Issueの題名 |
| 2013/7/21 | Commit | Initial commit |
| 2013/7/23 | Commit | フォルダ名・ファイル名の修正 |
| 2013/7/23 | Commit | 2012年度課題研究 |
| 2013/7/23 | Commit | 課題研究（伊藤） |
| 2013/7/23 | Commit | 課題研究（増田） |
| 2013/7/29 | Commit | PM演習の枠組み |
| 2013/7/29 | Commit | 野口の研究の全体像を記述するファイル |
| 2013/7/30 | Commit | GitHubアカウントのリスト |
| 2013/7/30 | Commit | Create 2013年度プログラム言語とプログラミング |
| 2013/7/30 | Commit | Create 長期的な課題.md |
| 2013/7/30 | Commit | Update 長期的な課題.md |
| 2013/7/30 | Commit | Update 長期的な課題.md |
| 2013/8/1 | Commit | Update 長期的な課題.md |
| 2013/8/7 | Commit | Create 参考文献リストのテンプレート.md |
| 2013/8/7 | Commit | Update 参考文献リストのテンプレート.md |
| 2013/8/7 | Commit | Update README.md |
| 2013/8/7 | Commit | Update README.md |
| 2013/8/7 | Commit | フォルダ名変更（卒論→卒業論文） |
| 2013/8/26 | Commit | アプリのランキングデータ |
| 2013/9/10 | Commit | Create GitHub APIの試し方.md |

・commitとIssueのログを収集できた．

・commitの総数は80で，Isuuesは114だった．

・commitは矢吹先生がコミットした日時を集めた．

・Issueは矢吹先生がクローズした日時を集めた．

・すべてのログを収集するまでに2時間40分かかった．

検証結果：得られたデータをそのままエクセルに打ち込んでいたためデータは整理され，正確な時間，commit,Isuueの題名は収集することができたが，手入力なので時間と労力を使ってしまった．

### 6.5.2 APIで個人のログを収集した場合

まずyabukilabのcommitの履歴をubuntで確認するため，eventsはper\_page=30固定，トータル300件しか取れないが，矢吹准教授の総コミット数は目視で300以下であったのでeventsを使う．

Pythonapi.pyhttps://api.github.com/users/taroyabuki/events>events-yabuki-20131119.txt

でデータを読み込み，個人の活動の日時とタイプを一覧表示するので，

./jq '.created\_at,.type' events.txt | awk '{ printf($0); if (NR % 2 == 0) printf("\n"); else printf(","); }' > events-yabuki-20131119.txt

図1-1のような生成されていない生のままのデータを収集する．

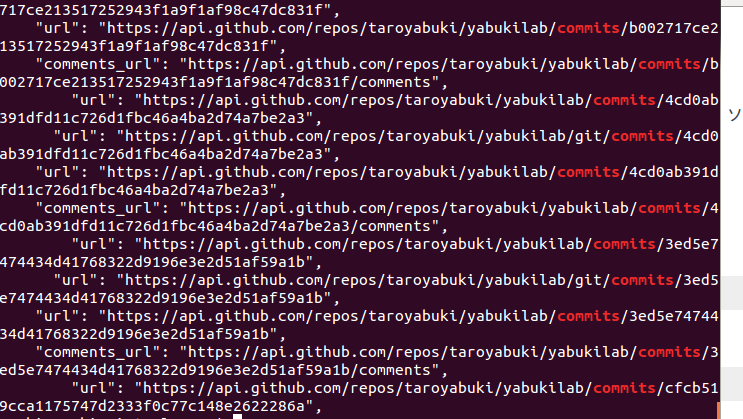


図1-1 APIで得られた生のcommitのデータ

この生のデータを，

pythonapi.py"https://github.com/taroyabuki/yabukilab/blob/master/library/github/events-yabuki-201311191800.txt" > events-yabuki.txt

としてテキストに保存する．

次にyabukilabのIsuueの履歴を確認する．先のcommitと同様にプログラムを書く．

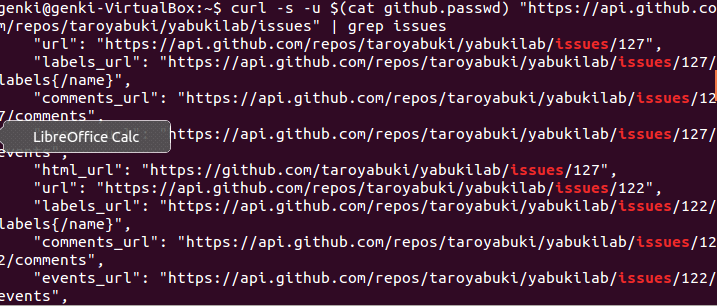


図1-2 APIで得られた生のIssuesのデータ

上記で得た二つのデータをエクセルに表示する．

表1-2 APIで得られた活動ログをエクセルに表示した場合

|  |  |
| --- | --- |
| 2013-11-19T08:10:14Z | PushEvent |
| 2013-11-19T06:26:31Z | PushEvent |
| 2013-11-19T05:59:05Z | PushEvent |
| 2013-11-19T05:17:49Z | PushEvent |
| 2013-11-19T03:00:23Z | PushEvent |
| 2013-11-19T02:47:49Z | PushEvent |
| 2013-11-15T08:49:19Z | PushEvent |
| 2013-11-15T06:27:01Z | PushEvent |
| 2013-11-15T06:25:32Z | PushEvent |
| 2013-11-14T09:39:13Z | PushEvent |
| 2013-11-14T09:37:55Z | PushEvent |
| 2013-11-14T09:21:47Z | PushEvent |
| 2013-11-14T07:50:32Z | PushEvent |
| 2013-11-12T07:21:01Z | PushEvent |
| 2013-11-12T07:19:08Z | PushEvent |
| 2013-11-12T06:27:47Z | PushEvent |
| 2013-11-12T06:07:33Z | PushEvent |
| 2013-11-12T04:53:32Z | PushEvent |
| 2013-11-12T02:37:38Z | PushEvent |
| 2013-11-12T02:16:05Z | PushEvent |
| 2013-11-12T01:57:15Z | PushEvent |
| 2013-11-11T14:38:06Z | IssueCommentEvent |
| 2013-11-10T14:27:42Z | PushEvent |

・10分ほどでデータの収集が完了．

・commit，Isueesの題名は取得できなかった．

検証結果：時間と労力を使うことなく素早くデータを収集できる．

### 6.5.3 手入力とAPIで比較

上記で得たデータをエクセルで比較してみる．

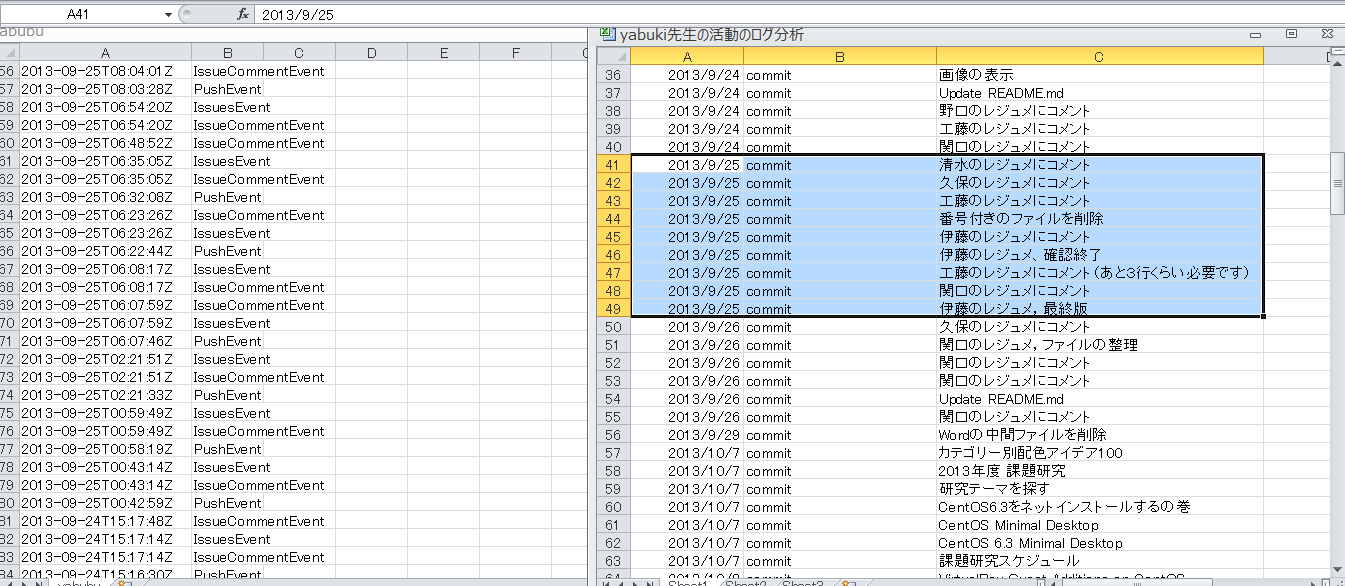


図1-3左がAPIを使用した活動ログ，右が手入力をした活動ログをそれぞれ表示した図

検証結果：手入力，APIともに総数や日時は同じであったが，本研究を進めるうえで，素早く正確なデータが得られることが検証できたので，本研究ではAPIを使用する．

### 6.5.4 使用したツール

実際に個人の活動ログを取得し，エクセルのピポットテーブルでまとめていく．例として，LearnBoostのGuillermo Rauch(guille)を対象とする．まず，GitHubからGuillermo RauchのGitHubでの全活動ログを取得しテキストに保存する．

python api.py "https://api.github.com/users/ guille /events" > events.txt

次に，個人の活動の日時とタイプを一覧表示し，リポジトリ名も取得するので，events.txtをjqで整形し，リポジトリ名が.repo.nameであることを突き止め，スクリプトを修正する．

./jq '.created\_at,.repo.name,.type' events.txt | awk '{ printf("%s", $0); if (NR % 3 == 0) printf("\n"); else printf(","); }' > events.csv

でCSV形式で保存すると以下のように表示される．

表1-3 Guillermo Rauchの活動ログ

|  |  |  |
| --- | --- | --- |
| 2013-09-04T20:30:35Z | LearnBoost/engine.io-client | IssueCommentEvent |
| 2013-09-06T02:01:05Z | LearnBoost/engine.io | IssueCommentEvent |
| 2013-09-06T17:22:37Z | visionmedia/page.js | PullRequestReviewCommentEvent |
| 2013-09-07T16:22:39Z | LearnBoost/socket.io | IssueCommentEvent |
| 2013-09-07T16:22:39Z | LearnBoost/socket.io | IssuesEvent |
| 2013-09-07T17:09:07Z | LearnBoost/engine.io | IssueCommentEvent |
| 2013-09-09T02:34:16Z | LearnBoost/socket.io | IssueCommentEvent |
| 2013-09-09T03:31:03Z | / | FollowEvent |
| 2013-09-14T18:05:11Z | mozilla/pdf.js | IssuesEvent |
| 2013-09-14T18:06:15Z | mozilla/pdf.js | IssueCommentEvent |
| 2013-09-14T23:23:12Z | LearnBoost/engine.io-client | IssueCommentEvent |
| 2013-09-16T04:12:48Z | / | FollowEvent |
| 2013-09-19T17:38:11Z | LearnBoost/engine.io-client | PushEvent |
| 2013-09-19T17:38:11Z | LearnBoost/engine.io-client | IssuesEvent |
| 2013-09-19T17:38:11Z | LearnBoost/engine.io-client | PullRequestEvent |
| 2013-09-19T17:38:27Z | LearnBoost/engine.io-client | IssueCommentEvent |
| 2013-09-21T17:53:10Z | bugsnag/bugsnag-js | IssuesEvent |
| 2013-09-21T17:53:41Z | bugsnag/bugsnag-js | IssueCommentEvent |
| 2013-09-23T02:13:01Z | LearnBoost/socket.io | IssueCommentEvent |
| 2013-09-25T04:07:13Z | / | FollowEvent |
| 2013-09-27T15:53:27Z | LearnBoost/engine.io | IssueCommentEvent |
| 2013-09-30T14:12:12Z | LearnBoost/engine.io-client | IssueCommentEvent |

上記のままではLearnBoost t以外のプロジェクトも表示されてしまうので，エクセルでLearnBoostのみを抽出する．抽出されたデータをエクセルのピポットテーブルを使いまとめる．

表1-4 Guillermo Rauchの活動ログのまとめ

|  |  |
| --- | --- |
| guille | (すべて) |
|  |  |
| **行ラベル** | **データの個数 / LearnBoost/** |
| **CommitCommentEvent** | **2** |
| **CreateEvent** | **12** |
| **DeleteEvent** | **5** |
| **IssueCommentEvent** | **97** |
| **IssuesEvent** | **17** |
| **PullRequestEvent** | **48** |
| **PullRequestReviewCommentEvent** | **5** |
| **PushEvent** | **39** |
| **総計** | **225** |

上記で述べた個人の活動ログを収集する手法を使い，プロジェクトメンバ全員のデータを抽出し，一覧表にまとめる．また，活動ログの収集結果一覧表のデータをCSV形式に変換しRで読み込み，主成分分析する際，グラフの値が読みにくくなってしまうので記号として以下のように置き換える．

表1-5 イベント名の変更

|  |  |
| --- | --- |
| イベント名 | 記号 |
| CommitCommentEvent | a |
| CreateEvent | b |
| DeleteEvent | c |
| IssueCommentEvent | d |
| IssuesEvent | e |
| PullRequestEvent | f |
| PullRequestReviewCommentEvent | g |
| PushEvent | h |
| WatchEvent | i |
| ForkEvent | j |
| GollumEvent | k |

### 6.5.5 LearnBoost

概要 : 小学校のクラス/児童管理をすべて一箇所でできる成績管理サービス．GitHub上でのプログラマの協力者は大，小規模合わせ1000名を超している．

プロジェクト期間：2011年5月1日～

メンバ数：6 人

・Guillermo Rauch (guille)

・MatthewMueller (Matthew Mueller)

・Nathan Rajlich (TooTallNate)

・Damián Suárez (retrofox)

・Carly Stambaugh (stambizzle)

・TJ Holowaychuk (visionmedia)

表1-6 LearnBoostの全メンバの活動ログの収集結果一覧表

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | a | b | c | d | e | f | g | h | i | j | k |
| guille | 2 | 12 | 5 | 97 | 17 | 48 | 5 | 39 | 0 | 0 | 0 |
| Matthew Mueller | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TooTallNate | 0 | 0 | 0 | 2 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| retrofox | 0 | 6 | 2 | 0 | 0 | 5 | 0 | 10 | 0 | 0 | 0 |
| stambizzle | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| visionmedia | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### 6.5.6 Bower

概要：バウアーは，Web用のパッケージマネージャで，問題への一般的なソリューションを提供している．フロントエンドパッケージ管理をよりビルドスタックによって消費することができるAPIを介してパッケージ依存関係モデルを明らかにさせておりGit上で動作し，パッケージに依存していない．

プロジェクト期間：2012年6月19日～

プロジェクトメンバ：13人

・Addy Osmani (addyosmani)

・Ben Schwarz (benschwarz)

・Chris Aniszczyk (caniszczyk)

・Dan heberden (danheberden)

・Alex MacCaw (maccman)

・Marco Oliveira (marcooliveira)

・Nicolas Gallagher (necolas)

・Paul Irish (paulirish)

・André Cruz (satazor)

・Sindre Sorhus (sindresorhus)

・Stephen Sawchuk (stephenplusplus)

・Sven Lito (svnlto)

・Mat Scales (wibblymat)

・データ収集結果

表1-7 Bowerの全メンバの活動ログの収集結果一覧表

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | A | b | c | d | e | f | g | h | i | j | k |
| Addyosmani | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Benschwarz | 0 | 2 | 0 | 55 | 9 | 6 | 10 | 26 | 0 | 0 | 0 |
| Caniszczyk | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Danheberden | 2 | 5 | 2 | 24 | 8 | 2 | 1 | 15 | 0 | 2 | 1 |
| Maccman | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Marcooliveira | 1 | 0 | 0 | 3 | 0 | 0 | 0 | 12 | 1 | 0 | 0 |
| Necolas | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Paulirish | 0 | 0 | 0 | 6 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
| Satazor | 0 | 3 | 3 | 50 | 15 | 9 | 1 | 11 | 0 | 0 | 0 |
| Sindresorhus | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Stephen Sawchuk | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| stephenplusplus | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Svnlto | 0 | 0 | 0 | 7 | 0 | 3 | 0 | 5 | 0 | 1 | 0 |
| Wibblymat | 2 | 6 | 2 | 88 | 23 | 33 | 7 | 28 | 3 | 0 | 0 |

### 6.5.7 Adobe Systems

概要：Adobeは，オープンソースへの取り組みに積極的に参加している．Adobeは，既存のプロジェクトに貢献し，オープンソースとしてコードを公開し，オープンアクセス機能を用いることでオープンな活動を支援している．HTML5，WebKitの，アパッチ， ZendのPHP，SQLliteに貢献している．

プロジェクト期間： 2010年9月24日～

メンバ：19人

・Alexandru Chiculita (achicu)

・Adam Lehman (adrocknaphobia)

・Alexander Macdonald (alexmac)

・Brian LeRoux (brianleroux)

・Christophe Coenraets (ccoenraets)

・Raymond Camden (cfjedimaster)

・Kevin Dangoor (dangoor)

・DmitryBaranovskiy (DmitryBaranovskiy)

・emalasky (emalasky)

・jacob(fat)

・Garth Braithwaite (GarthDB)

・harish-io (harish-io)

・Ian Wehrman (iwehrman)

・LarZ(larz0)

・mikechambers (mikechambers)

・Divya Manian (nimbupani)

・piatra (piatra)

・thibaultimbert (thibaultimbert)

・データ収集結果

表1-8 Adobe Systemsの全メンバの活動ログの収集結果一覧表

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | a | b | c | d | e | f | g | h | i | j | k |
| Achicu | 0 | 7 | 1 | 4 | 1 | 3 | 0 | 52 | 1 | 0 | 0 |
| Adrocknaphobia | 0 | 28 | 3 | 70 | 31 | 32 | 2 | 1 | 2 | 1 | 22 |
| Alexmac | 0 | 4 | 0 | 40 | 18 | 14 | 0 | 56 | 2 | 0 | 0 |
| Brianleroux | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ccoenraets | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cfjedimaster | 0 | 0 | 0 | 17 | 4 | 1 | 0 | 0 | 0 | 1 | 0 |
| Dangoor | 0 | 5 | 11 | 147 | 31 | 16 | 42 | 36 | 0 | 0 | 6 |
| DmitryBaranovskiy | 0 | 9 | 0 | 56 | 51 | 17 | 1 | 60 | 0 | 0 | 0 |
| Emalasky | 0 | 10 | 3 | 1 | 1 | 2 | 0 | 74 | 1 | 2 | 0 |
| Fat | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GarthDB | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| harish-io | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Iwehrman | 1 | 8 | 7 | 65 | 11 | 15 | 106 | 73 | 0 | 0 | 1 |
| Jhatwich | 0 | 0 | 0 | 10 | 0 | 12 | 1 | 0 | 1 | 3 | 2 |
| larz0 | 1 | 15 | 7 | 126 | 9 | 17 | 18 | 59 | 0 | 0 | 0 |
| Mikechambers | 1 | 1 | 0 | 37 | 26 | 5 | 4 | 9 | 1 | 1 | 0 |
| Nimbupani | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 6 | 0 | 0 | 0 |
| Piatra | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Thibaultimbert | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### 6.5.8 grunt

概要：タスクランナーであり，縮小，コンパイル，ユニットテスト，リンティングのような反復的なタスクを実行するときに作業を自動化させるリポジトリ．日々絶えることなくバージョンを更新しており，人気のリポジトリとして有名．

プロジェクト期間： 2012年5月4日～

メンバ：10人

・Ben Alman (cowboy)

・Dave Geddes (geddski)

・Isaac Durazo (isaacdurazo)

・Jarrod Overson (jsoverson)

・Jörn Zaefferer (jzaefferer)

・Kyle Robinson Young (shama)

・sindresorhus (sindresorhus)

・Tim Branyen (tbranyen)

・Tyler Kellen (tkellen)

・vladikoff (vladikoff)

・データ収集結果

表1-9 gruntの全メンバの活動ログの収集結果一覧表

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | a | b | c | d | e | f | g | h | i | j | k |
| Cowboy | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 13 | 1 | 0 | 0 |
| Geddski | 0 | 6 | 0 | 22 | 7 | 3 | 1 | 6 | 0 | 0 | 0 |
| Isaacdurazo | 0 | 0 | 1 | 1 | 2 | 1 | 0 | 13 | 2 | 0 | 0 |
| Jsoverson | 0 | 6 | 0 | 61 | 22 | 17 | 0 | 13 | 0 | 0 | 0 |
| Jzaefferer | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Shama | 1 | 4 | 1 | 93 | 29 | 21 | 1 | 14 | 0 | 0 | 0 |
| Sindresorhus | 1 | 1 | 0 | 7 | 3 | 01 | 1 | 3 | 0 | 0 | 0 |
| Tbranyen | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tkellen | 1 | 12 | 3 | 57 | 25 | 18 | 0 | 18 | 0 | 0 | 0 |
| Vladikoff | 0 | 4 | 5 | 41 | 18 | 16 | 3 | 21 | 0 | 1 | 0 |

### 6.5.9 Resque

概要：Railsアプリケーションの中で非同期処理（バッチ処理）を行う．非同期処理を行うためのRubyライブラリとしてはBackgrounDRbやdelayed\_jobなどに次ぎ有名なライブラリ．

プロジェクト期間：2013年1月11日～

メンバ：6人

・Matteo Centenaro (bugant)

・Eoin Coffey (ecoffey)

・Terence Lee (hone)

・steveklabnik (steveklabnik)

・Tony Arceieri (tarcieri)

・Ryan Biesemerye (yaauie)

・データ収集結果

表1-10 Resqueの全メンバの活動ログの収集結果一覧表

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | a | b | C | d | e | f | g | h | i | j | k |
| Bugant | 0 | 0 | 0 | 23 | 0 | 8 | 6 | 4 | 0 | 0 | 0 |
| Ecoffey | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| Hone | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Steveklabnik | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tarcieri | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Yaauie | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### 6.5.10 jekyll

概要：Webサイトやブログにユーザのプレーンテキストを変換するリポジトリ．シンプルなデータベース，コメントの管理が特徴．

プロジェクト期間： 2012年10月19日～

メンバ：4人

・Coby Chapple (cobyism)

・Matt Rogers (mattr-)

・Parker Moore (parkr)

・Antoal Broder (penibelst)

・データ収集結果

表1-11 jekyllの全メンバの活動ログの収集結果一覧表

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | a | b | c | d | e | f | g | h | i | j | k |
| cobyism | 0 | 0 | 0 | 8 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| mattr- | 0 | 3 | 3 | 65 | 7 | 19 | 13 | 31 | 0 | 0 | 0 |
| parkr | 3 | 16 | 11 | 81 | 8 | 25 | 11 | 46 | 1 | 0 | 0 |
| penibelst | 2 | 11 | 11 | 105 | 6 | 13 | 7 | 19 | 1 | 0 | 0 |

## 6.6 活動ログの分析

6.5で算出された，全メンバの活動ログの収集結果一覧表をもとにRで分析していく．

### 6.6.1 Rとは

Rは統計計算とグラフィックスのための言語・環境で，GNUプロジェクトの一つであり，AT&T(当時)のベル研究所でJohn M. Chambersと同僚により開発されたS言語・環境に似ている．RはSを別個に実装したものと考えられている．両者の仕様及び実装には幾つかの重要な違いがあるが，S用に書かれたコードの多くは変更無しでRでも実行できる．多様な統計手法(線形・非線形モデル・古典的統計検定・時系列解析・判別分析・クラスタリング・その他)とグラフィックスを提供し，広汎な拡張が可能であり，Sは高頻度で統計的手法による研究の手段の選択肢となるが，Rを使えばそうした活動にオープンソースの道から参加できるようになる．

Rの強みの一つは，巧くデザインされた出版物並みのプロットを容易に作成できる点であり，これには必要なら数学記号や式を含めることもできる．グラフィックスにおける細かなデザインの標準設定に，これまで周到な配慮が払われており，ユーザが完全に制御することもできる．Rはデータ操作・計算・そしてグラフィックス表示のためのソフト機能の統合された纏まりでもあり，次の要素がある．

* データを効率的に操作し保管する機能
* 配列(特に行列)の計算のための演算子のセット
* データ解析の媒介に使う道具（主成分分析や因子分析など）の大規模で一貫した集り
* データ解析のためのグラフィカルな機能・画面又は印刷物への出力
* 条件分岐・ループ・ユーザ定義の再帰的関数や入出力機能を含み開発の進んだ簡潔で効率的なプログラミング言語

### 6.6.2 主成分分析とは

主成分分析は，たとえばビーカに入った液体の成分を明らかにすることを意味している．ビーカの液体は無色だが，その成分は目で見ただけでは分からない．そこで化学の分野では液クロマトグラフといった装置を使用して，液体成分を分析するのである．主成分分析は，成分内容の不明な対象を明確にする方法である．統計学における主成分分析(PCA:PrincipalComponent Analysis)は，ある対象について観測された多数の項目(x1,x2,x3,….xp )を個々に分析するのではなく，それらの項目全体が織りなす意味合いを解釈するために使用する方法論である．項目を個々に分析する方法は一変量解析とよばれ，個々の項目ごとに平均や分散を求め各項目のもつ意味を解釈するが，主成分分析は多数の項目（人間の体が対象なら，身長，体重，ウェスト，ヒップ，足のサイズなど）が全体としてもっている意味合いを解釈するために使用するもので多変量解析の一方法論である．上記のデータを主成分分析することで，例えば体型について「大柄－小柄」，「がっしり型－スリム型」といった情報を抽出できるのである．主成分分析は数学的には固有値分解に過ぎないが，工学の分野ではスペクトル解析とよばれ波形分析などに応用されている．一般に言葉では次のように説明される．

* 説明変量の圧縮
* 多変数から少数の総合的指標にする
* 多変数から少数の合成変数にする
* データの低次元化
* データの圧縮または縮約
* 情報の簡素化・要約

現在では主成分分析はその計算結果につづけて他の手法に使うのが主流で，こうしたその他の手法まで含めて主成分分析としてあつかわれている．具体的な手法として，

* 多変数から複数のメカニズムの分離（主成分スコアの計算と同じ利用目的が異なる）
* メカニズム提示（因子負荷量と相関方向）
* 主成分回帰（単純な曲線の重回帰）
* メカニズムの曲線方程式の算出（主成分層別によるメカニズム関連データ抽出と重回帰による方程式）が主な手法である．

### 6.6.3 使用したツール

本研究の収集されたデータを解析するためにRでツールを開発し導入した．

まず，6.5で算出された，全メンバの活動ログの収集結果一覧表CSV形式のデータを読み込む．

setwd("c:/cit")

myData <- read.csv("learnboost.csv")

head(myData)

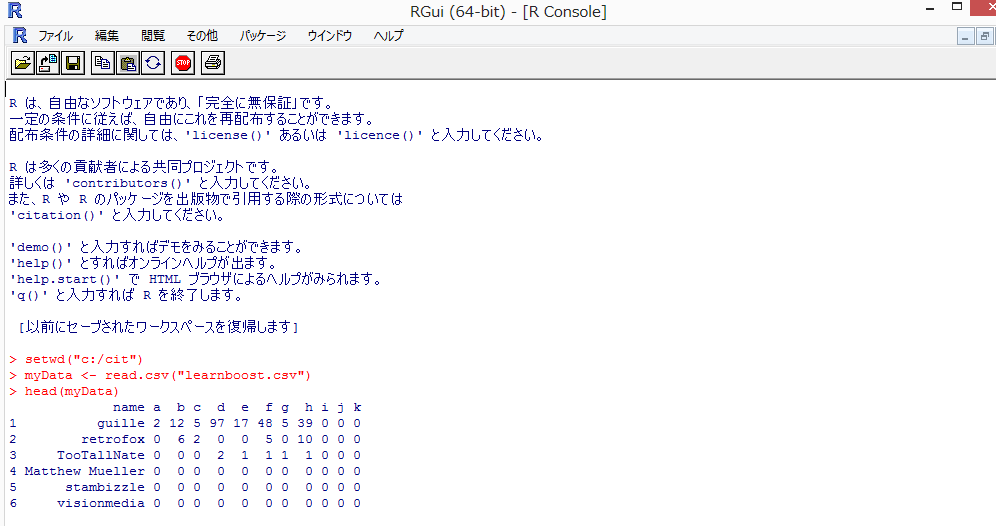


図1-4 RでCSV形式のエクセルデータを取り込んでいる場面

次に，行に名前が表示され結果を見やすくし，

#biplot

row.names(myData) <- myData$name

myData

これから主成分分析を行うと宣言する．

#主成分分析

myResult <-prcomp(myData[,-1]) #1列目（name）は不要

主成分分析をしたデータの概要を表示する．

#結果の概要

summary(myResult)

主成分分析をしたデータを固有ベクトルにする．

#主成分

myResult$rotation

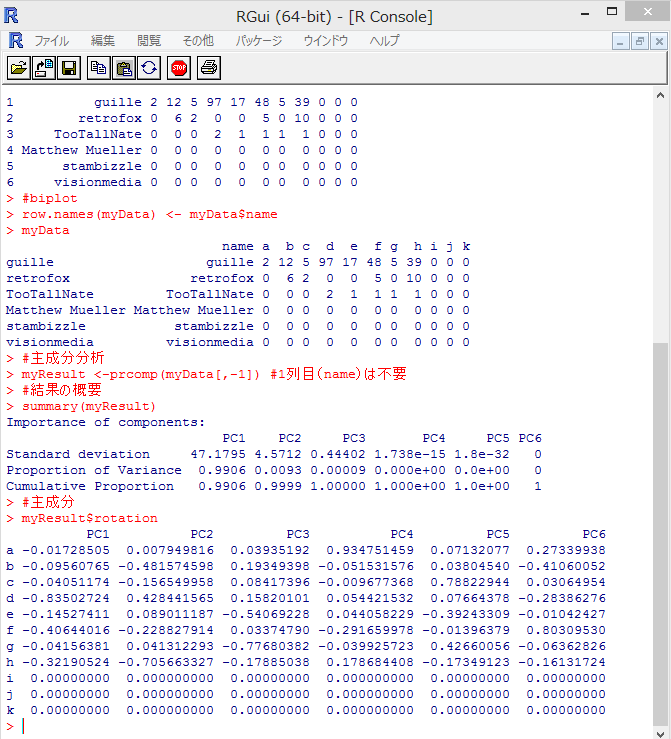


図1-5 Rで主成分分析を行い，結果を表示した場面

次に，主成分をグラフに図示していく．

第一主成分の表示

#主成分の図示

barplot(sort(myResult$rotation[,1]))

第二主成分の表示

barplot(sort(myResult$rotation[,2]))

上記の2つのコードで第一，第二主成分の明らかにさせる．また，各プロジェクトメンバのイベント数のばらつきを把握するため，主成分スコアを表示させる．

#主成分スコア

myResult$x

#主成分スコアの図示

biplot(myResult

上記のツールを使い，各プロジェクトの分析を行い結果を解釈する．その際，主成分を解釈するときには，たとえば第二主成分（以下PC2）までの累積寄与率が十分に高いならば，データの持つばらつきは第1主成分（以下PC1）・PC2の2つだけで十分に説明できると判断して，これら2つの主成分の解釈のみを行えばよい．累積寄与率の値によっては，第三主成分以下の解釈が必要になることもある．主成分の意味を解釈する際に注意すべきことは，主成分どうしは互いに直角に交わる（直交する）方向を向くように計算されて作られているという点である．つまり，主成分どうしは互いに無相関・無関係なのだから，このことを解釈にも反映させなければならない．すなわち，ある主成分の意味するものは，他の主成分が意味するものとは性質が違っているということである．たとえば，PC1を「重さ」を表すものと解釈したならば，PC2を「軽さ」を表すものと解釈するのは誤りである．なぜならば，もしPC1が重さを表すものであるならば，「軽い」ということは符号の違い（「重い」が正ならば「軽い」は負）で表されてしまい，PC1だけで説明がついてしまうからである．したがって，この場合，PC2は「軽い－重い」とは無関係なものに解釈しなければならない．たとえば「高い－低い」などのように，PC1とは無関係な要素に注目して解釈を考える必要がある．以上を踏まえ主成分分析を行い，解釈する．

主成分スコアでは，分布されている場所を解釈することでメンバがどのようなイベントを行っているかが分かるようになっている．メンバ同士が重なっている部分はイベントが被っており，作業の重複やどのイベントに重点を置いているかがわかる．また，メンバ同士が重なることなくバラバラに分布されている場合，メンバ一人ひとり別のイベントをこなしているので，プロジェクトにおいての役割分担がはっきりしていると考察できる．

### 6.6.4 LearnBoost

表1-10 LearnBoostのPC1

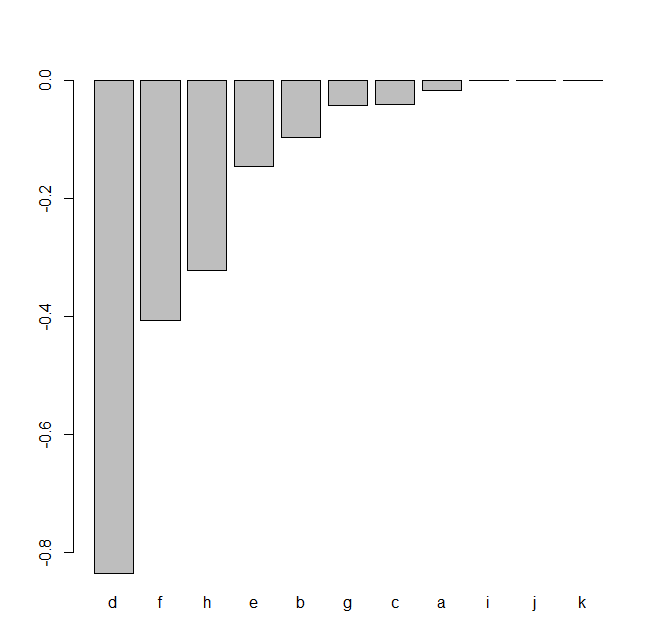


表1-11 LearnBoostのPC2

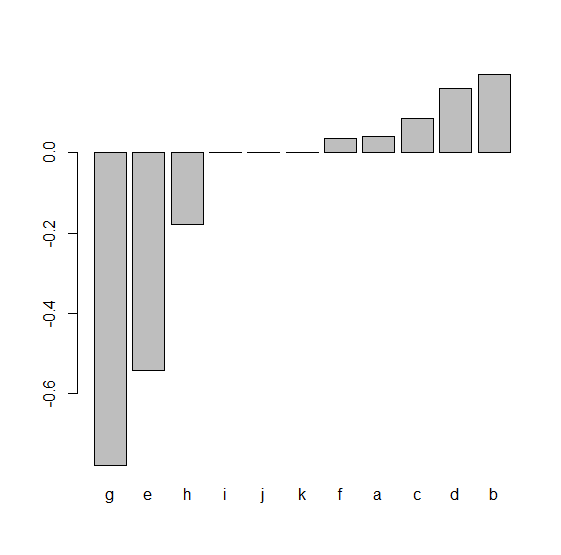
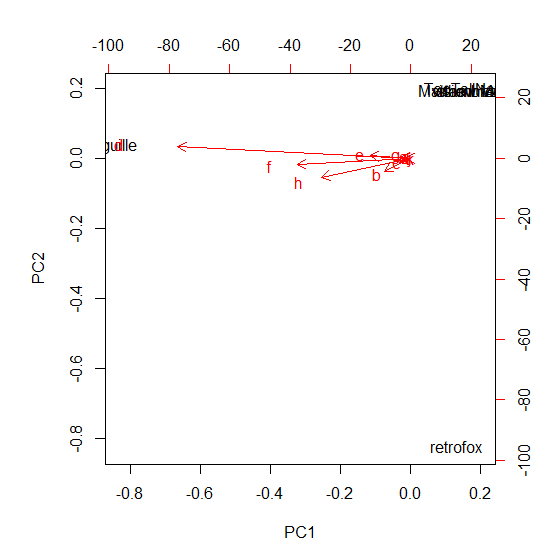


表1-12 LearnBoostの主成分スコア



分析結果の考察 :

・PC1ではIssueCommentEventが負の-0.8で，WatchEvent, ForkEvent,GollumEventが0.0に限りなく近く並ぶ結果となった．

・PC2ではPullRequestReviewCommentEventが負の-0.6を大きく下回り，CreateEventが正の0.0を上回った．

・主成分スコアでは，guilleとretrofox以外のメンバが一か所に固まっており，OSS上で活発に活動を行っているメンバは2人のみであることが考察できる．

・LearnBoostはOSS外での役割を担当しているメンバに重点をおいたプロジェクトだと考察できる．

### 6.6.5 Bower

表1-13 BowerのPC1

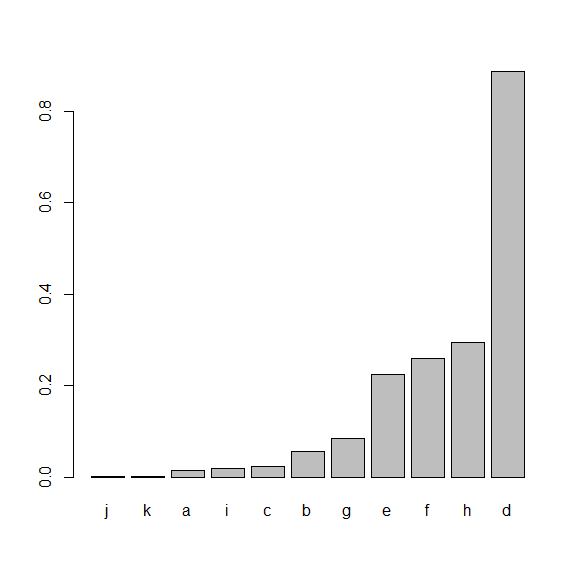


表1-14 BowerのPC2

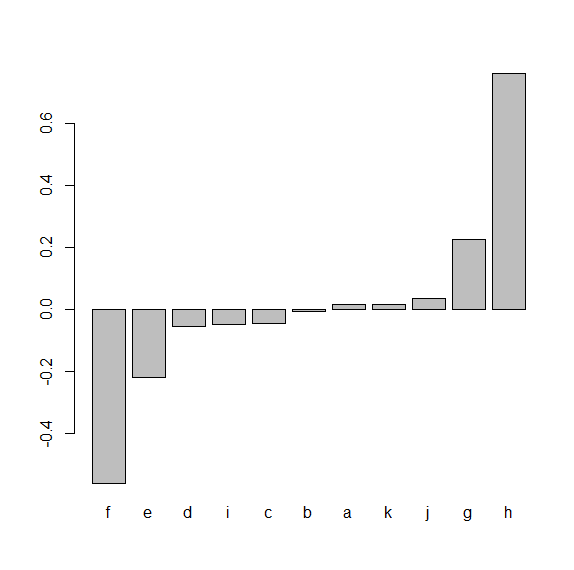
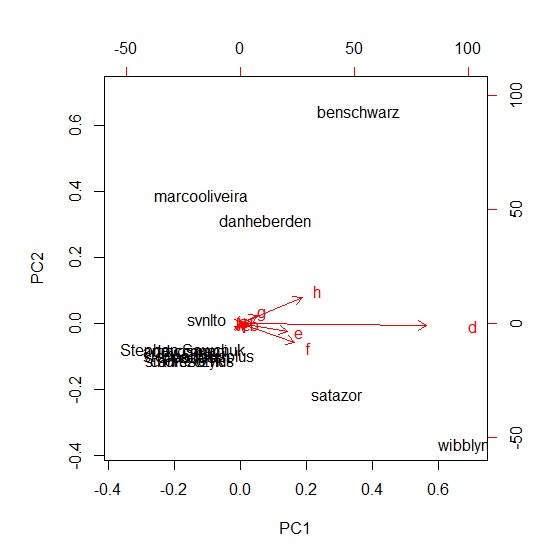


表1-15 Bowerの主成分スコア



分析結果の考察 :

・PC1ではWatchEventとGollumEventが0.0に限りなく近く，IssueCommentEventが正の0.8を大きく上回っている．

・PC2ではPullRequestEventが負の-0.4を下回り，PushEventは正の0.6である．

・主成分スコアでは，イベントを多くこなしているメンバとそうでないメンバは分かれている．役割分担がなされていると考察できる．

・Bowerは，13人中7人がOSS外での役割を担当し，6人がプログラマとしてGitHub上で活動していることが考察できる．

### 6.6.6 Adobe Systems

表1-16 Adobe SystemsのPC1

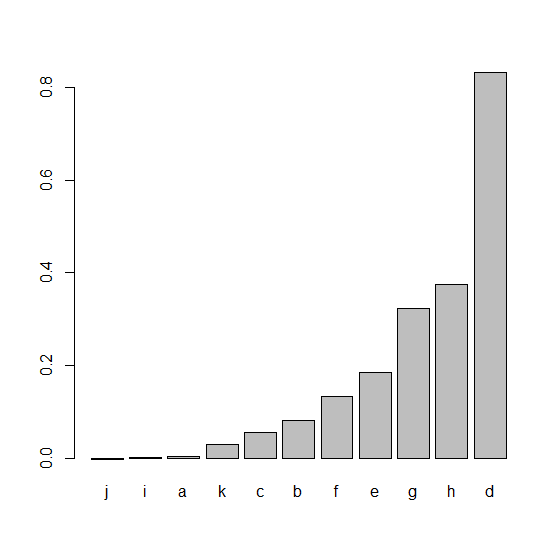


表1-17 Adobe SystemsのPC2

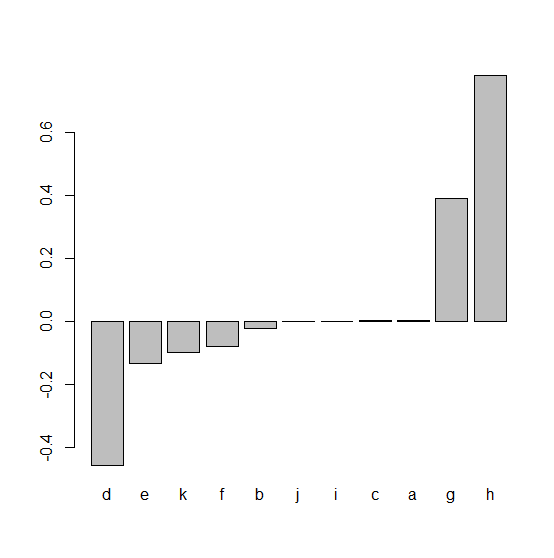
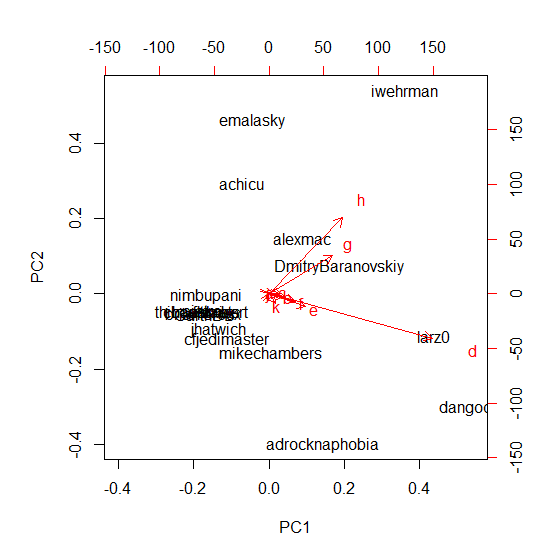


表1-18 Adobe Systemsの主成分スコア



分析結果の考察 :

・PC1ではForkEventが0.0に近く，IssueCommentEventが0.8であった．

・PC2ではIssueCommentEventが-0.4であり，PushEventが0.6を大きく上回った．

・主成分スコアでは，イベントを多くこなしているメンバとそうでないメンバは分かれている．役割分担がなされていると考察できる．

・Adobe Systemsは多くのプロジェクトメンバがIssueCommentEventを行っており，メンバ間でのやりとりも多く，OSS上の活動に重点をおいていると考察できる．

### 6.6.7 grunt

表1-19 gruntのPC1

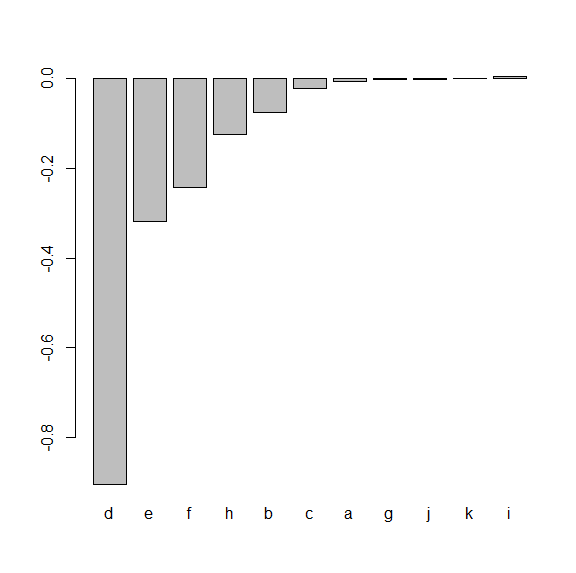


表1-20 gruntのPC2

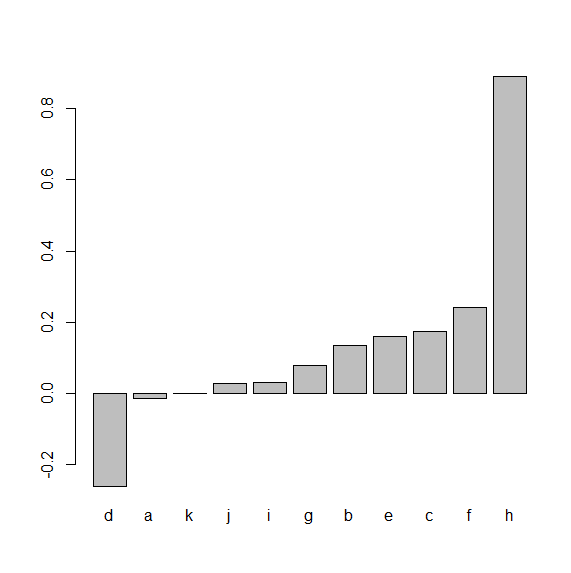
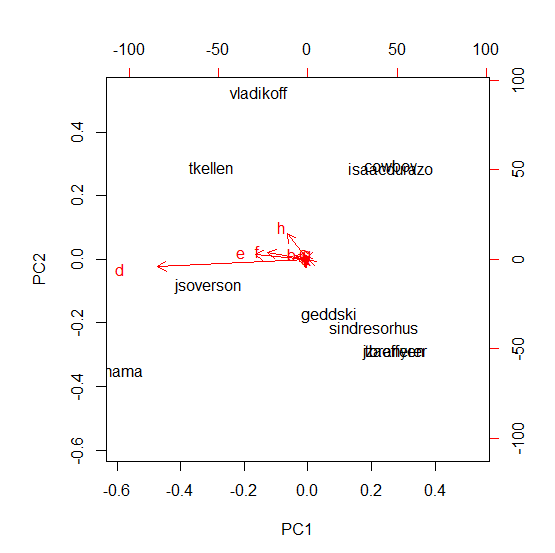


表1-21 gruntの主成分スコア



分析結果の考察 :

・PC1ではIssueCommentEventが負の-0.8であり，WatchEventが正の0.1であった．

・PC2ではIssueCommentEventが負の-0.2であり，PushEventが正の0.8であった．

・主成分スコアでは，イベントを多くこなしているメンバとそうでないメンバは分かれている．このプロジェクトは各メンバがそれぞれ別のイベントをこなしており，役割分担がしっかりされていると考察できる．

### 6.6.8 Resque

表1-22 ResqueのPC1

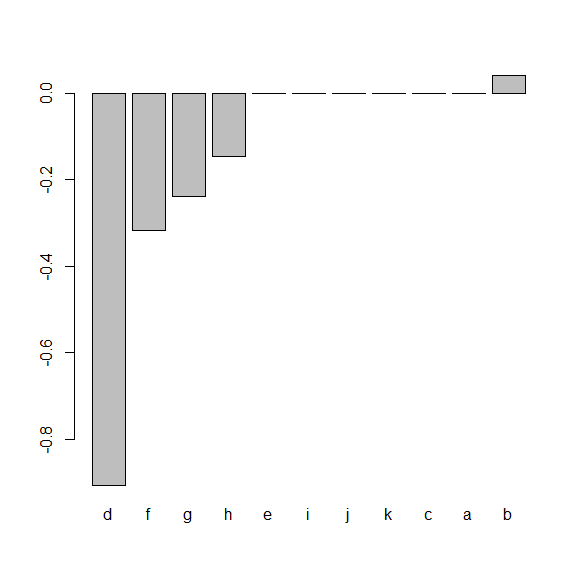


表1-23 ResqueのPC2

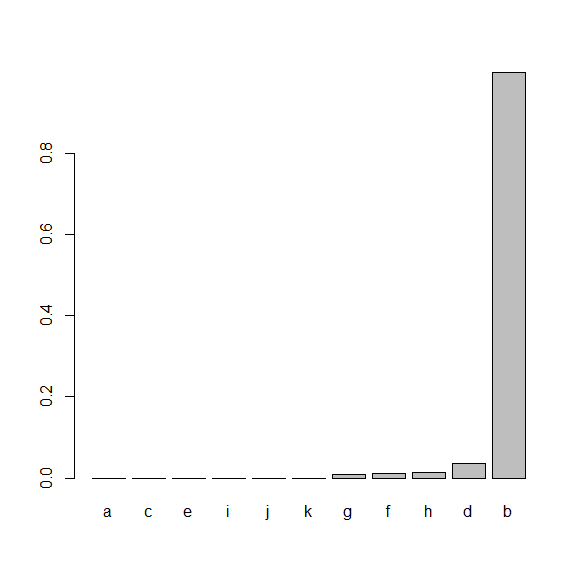
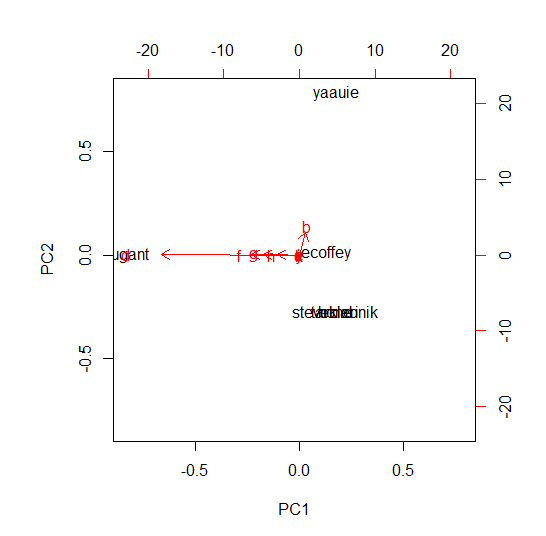


表1-24 Resqueの主成分スコア



分析結果の考察 :

・PC1ではIssueCommentEventが負の-0.8を大きく下回り，CreateEventが0.0を上回った．

・PC2ではCommitCommentEvent，DeleteEvent，IssuesEvent，WatchEvent，ForkEvent，GollumEventの6つのイベントが0.0であり，CreateEventが正の0.8を大きく上回った．

・主成分スコアでは，jgnatとecoffey以外のメンバが一か所に固まっており，OSS上で活発に活動を行っているメンバは2人のみであることが考察できる．

### 6.6.9 jekyll

表1-25 jekyllのPC1

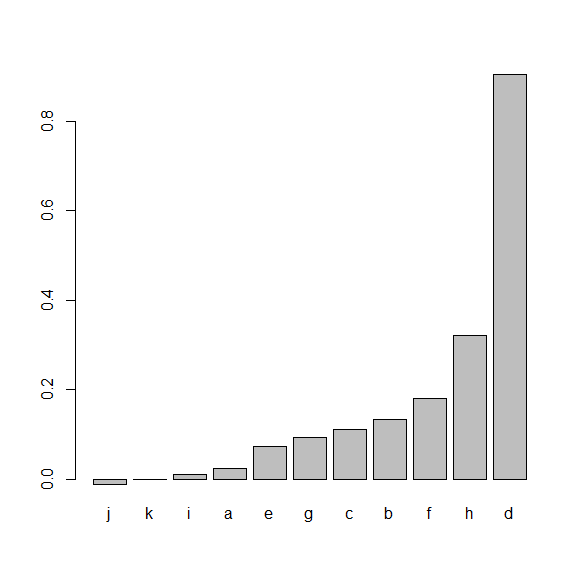


表1-26 jekyllのPC2

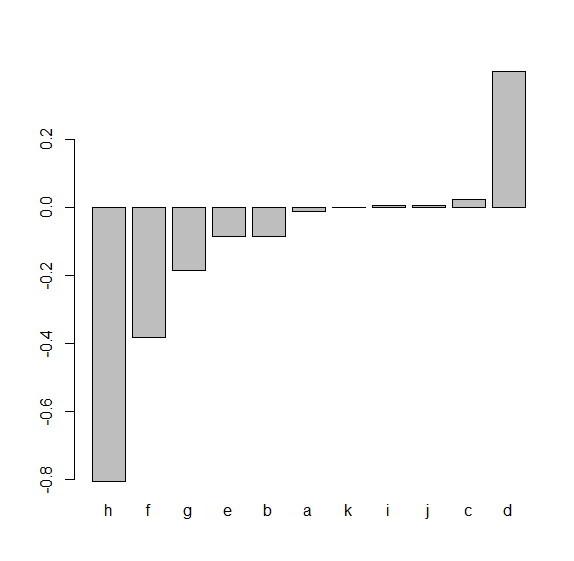
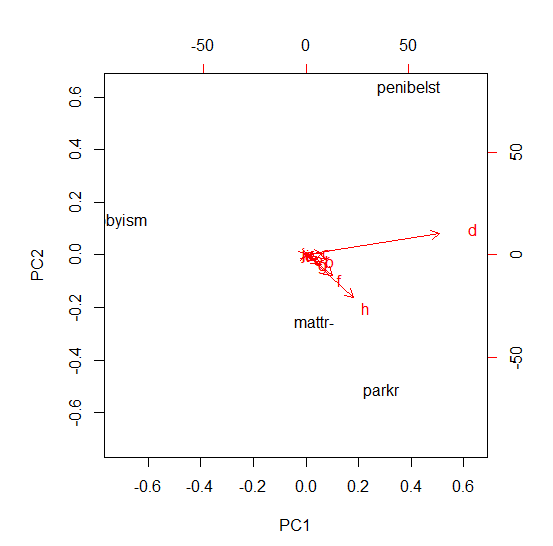


表1-27 jekyllの主成分スコア



分析結果の考察 :

・PC1ではWatchEventが0.0を少し下回り，IssueCommentEventが正の0.8を大きく上回った．

・PC2ではPushEventが負の-0.8であり，IssueCommentEventが正の0.2であった．

・主成分スコアでは，イベントを多くこなしているメンバとそうでないメンバは分かれている．このプロジェクトは各メンバがそれぞれ別のイベントをこなしており，メンバ同士重なることはなく，役割分担がしっかりされていると考察できる．

次章では6章で明らかになった各プロジェクトの役割分担の実態を総括し，結論としてまとめる．

**第7章**

**結論**

# 第7章 結論

## 7.1 結果・考察

6章ではプロジェクトごとの分析結果を考察してきた．7.1では全プロジェクトの活動ログのPC1，PC2の正負の最大値のイベントを下記の表にまとめる．複数イベントがある場合はイベント数が同じである．

表1-10 各プロジェクトのPC1，PC2の正負の最大値

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | PC1正 | PC1負 | PC2 正 | PC2 負 |
| LearnBoost | IssueCommentEvent | WatchEvent/  0.0 | PullRequestReviewCommentEvent | CreateEvent |
| Bower | IssueCommentEvent | WatchEvent/  GollumEvent | PushEvent | PullRequestEvent |
| Adobe Systems | IssueCommentEvent | ForkEvent/  0.0 | PushEvent | IssueCommentEvent |
| grunt | WatchEvent/  0.0 | IssueCommentEvent | PushEvent | IssueCommentEvent |
| Resque | CreateEvent/  0.0 | IssueCommentEvent | CreateEvent | WatchEvent/ ForkEvent |
| jekyll | IssueCommentEvent | WatchEvent | IssueCommentEvent | PushEvent |

以上から主成分は次のような傾向が多く見られた．

* PushEventとIssuesCommentEventの絶対値が大きく，正負が逆であることが分かった．つまり，PushとIssueにコメントをする行為は同じメンバが行っていることが多いことが考えられる．
* WatchEvent とIssueCommentEventの絶対値が大きく，正負が逆であることが分かった．つまり，リポジトリにスターを付ける行為とIssueにコメントをする行為は別々のメンバが行っていることが多いことが考えられる．
* OSS開発プロジェクトでは活動ログが無いメンバが存在することが分かった．このことから，活動ログの無いメンバは，OSS外でなんらかの役割を担当しているメンバだと考察できる．

## 7.2 今後の課題と発展

今回の研究で，あくまで役割が重複していないか，どこのイベントに重点をおいているかなどの実態は分かったが，プロジェクト内のプロジェクトリーダーやプロジェクトマネージャーなどの具体的な役割は不明確であった．今後，さらに詳しく役割を明らかにするために，PMBOKで定義されている，責任分担(RACI)マトリックスのようなレベルごとに役割を表す独自の調査項目を作り，APIなどで収集した活動ログを責任分担(RACI)マトリックスに自動で反映させるツールを作ることで，各プロジェクトの詳しい役割分担の実態を明らかにできるのではないかと期待できる．

また，本研究の手法で出た活動ログを参考にし，個人のコミットのパターンや時間変化の周期を自動で考えるコードを作り，一定時間コミットが無いと自動的に知らせが来るOSSプロジェクト上の心電図のような機能を利用することで，遠方にいるメンバの健康管理や遅延防止につながるのではないか．

## 7.3 結論

プロジェクトメンバの役割分担の実態を解明することができた．このような手法を活用することによって，OSS開発プロジェクトの実態を明らかにしたり，プロジェクトマネジメントの手法を導入したりすることが容易になると期待される．