

Web サービスの障害がプロジェクトに与える影響を Twitter を活用して調査する方法

PM コース 矢吹研究室 1442012 岩瀬翔

1. 研究の背景

複数のメンバが同時に開発を行うソフトウェア開発プロジェクトでは、Web サービスが使われることがある。例えば、チーム内でファイルのバージョンを管理することのできる「GitHub」というサービスがある。他にもチーム内でコミュニケーションを取るためのチャットツール「Slack」やユーザレビューなどを行うために使用する「Skype」、チームで作成したファイルを共有できる「Google Drive」などがある。これらの Web サービスに障害が発生した場合、プロジェクトの進捗に影響が出るのではないかと考えた。

上記に述べた 4 つの Web サービスの中で、GitHub は企業においてもオープンソースソフトウェアとしてソースコードを公開するため利用されている。理由は不正なプログラムや脆弱性などの確認（監査）を行う事ができ、ソフトウェア自身の信頼性を判断する事ができるからなどである [1]。その GitHub のサーバーが 2016 年 1 月 28 日にダウンし、インターネット上で話題になった。実際に GitHub のサーバーダウンについて、Twitter を用いて調べてみたところ「仕事にならない」「卒論が書けない」などといった反応が多く見られた [2]。

GitHub がサーバーダウンしてしまった事例から、ソフトウェア開発プロジェクト進行中に GitHub で障害が発生した場合どのような影響が生じ、その対策のために何かできることはないのかと考え研究することとした。

2. 研究の目的

本研究の目的は次の 2 点を目標とする。

- ソフトウェア開発プロジェクトで使用される Web サービスに障害が発生した場合、どのような影響が発生し、その影響がどれほどの人に及ぶのか調査する。
- Web サービスの障害発生に対するリスク対策案を考案する。

上記の目標を達成することを目的とし研究する。

3. プロジェクトマネジメントとの関連

本研究はプロジェクトマネジメントにおける 10 個の知識エリアのうち、リスクマネジメントに関連付けることができる。理由として、Web サービスの障害はソフトウェア開発プロジェクトの進捗に関わる明確なリスクだからである。

4. 研究の方法

本研究は、ソフトウェア開発プロジェクトにおいて使用する代表的な Web サービスである、GitHub の障害発生時の状況について調査する。調査方法は、Twitter で投稿されている障害発生に関するツイートをデータとして収集する。データ収集の手順は以下の通りである [3]。

1. HTML ファイルからツイートの時間と本文のみを抽出するためのプログラムを作成する [4]。
2. GitHub に関連するすべてのサービスを断続的に状況監視している「GitHub Status」を参照し、2016 年で障害が発生、復旧したと記録されている時間を調べる。その時間前後を Twitter で「GitHub, 言語, 日付, 時間」を指定し検索する（例：GitHub lang:ja since:2016-10-19 until:2016-10-21_20:30:00_JST）。その後、障害発生した旨のツイートから、復旧したという旨のツイートを表示する。

3. 作成したプログラムをを使用して検索結果からデータを抽出する。
4. 2～3 の手順を障害発生時間ごとに繰り返す。
5. 障害発生から復旧までの「GitHub」に対するツイート数と、どのくらいの時間で復旧が完了するのかを調べる。
6. テキストマイニングにより頻出単語の抽出を行い、どのような影響が発生しているのか、どのような反応があるのか分析する。
7. 分析結果を基に、リスク対策案を考察する。

5. 現在の進捗状況

GitHub Status を参照に調べたところ、2016 年の GitHub の障害発生回数は 13 回であった。発生した障害それぞれについてデータの収集を完了した。下記のグラフは、各障害の発生から復旧までに投稿されたツイートの数のグラフ、各障害の発生から復旧までの時間のグラフである。

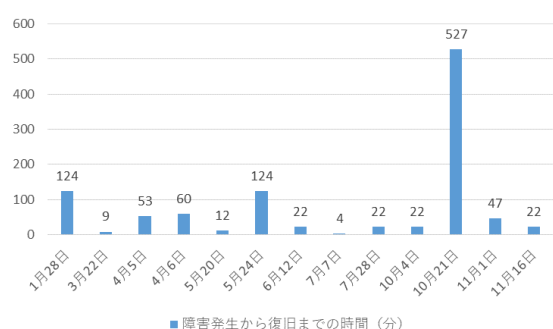


図 1 各障害の発生から復旧までの時間

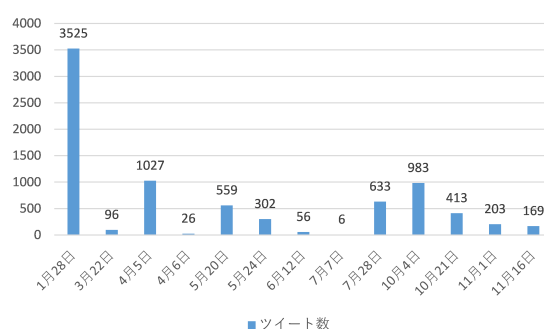


図 2 各障害発生時のツイート数

13 回の障害を見比べてみると、障害発生の曜日、時間帯等によって取得したツイートの数に幅があることが分かった。例えば、日中の約 20 分間の障害でも祝日と平日では 900 ツイート以上の差がある。さらに、平日の日中であればツイート数が多いのはもちろんだが、深夜 2 時に発生した約 10 分間の障害でも 100 件近いツイートが集まった。また、障害の発生報告と復旧報告が GitHub Status よりも平均約 7.7 分ほど速いということが分かった。

6. 今後の計画

以下のように研究を進める計画である。

1. 時間帯、曜日を加味して分析を行うことも検討する。
2. 集めたツイート本文のデータからテキストマイニングにより頻出単語を抽出し、多くツイートされている単語を分析する。
3. 背景で述べた「Slack」「Skype」「Google Drive」の障害発生についても同様の研究を行う。
4. 4 つの Web サービスのデータを調査した上で、障害発生に対するリスク対策案を考察する。

参考文献

- [1] まつもとゆきひろ監修. ネットを支えるオープンソース. KADOKAWA, 第 1 版, 2014.
- [2] 岩城俊介, @ IT. Github がダウン、「ぜんぶ のせいだ」のような惨状 今後の課題も. <http://breaking-news.jp/2016/01/28/022573> (2016.06.29 閲覧) .
- [3] 辻岡大知. ソーシャルメディアにおける炎上案件に関わるユーザの動向調査. 課題研究, 千葉工業大学, 2015.
- [4] 佐々木拓郎, るびきち. Ruby によるクローラー開発技法. SB クリエイティブ, 第 2 版, 2014.