

Twitter 発言の分析による Web サービス障害の影響調査

岩瀬 翔[†] 矢吹 太朗

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[‡]

1 序論

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 サービスを使うことがある。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 [1]。その GitHub のサーバーが 2016 年 1 月 28 日にダウンした。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」がというサービスがあるが、これも 2017 年 11 月 1 日にダウンした。Web サービスの停止は、それを利用しているプロジェクトに大きな影響を与えると思われる。実際、GitHub ダウン時には、そのことに関する多くのツイートが Twitter 上に投稿されていた [2]。Web サービスのサーバーがダウンしてしまった事例から、Web サービスの障害がソフトウェア開発プロジェクトの進捗に影響を及ぼすのではないかと考えた。したがって、本研究では、Web サービスにおけるサーバーダウン等の障害によるサービス停止について、Twitter を用いた影響調査をする。

2 目的

ソフトウェア開発プロジェクトで使われる Web サービスに障害が発生した場合、どのような影響がどのくらいの頻度で発生するのか調査する。また、その影響がどれほどの人に及ぶのか調査する。

3 手法

調査方法は Twitter で投稿されている GitHub の障害発生に関するツイートをデータとして収集する。Twitter では API が提供されているが、仕様によって過去 1 週間以上前のツイートが検索によって取得するため以下の手順で行う。

1. Twitter から API を使わずにデータを取得するためのプログラムを作成する。
2. サービスの停止から復旧までに投稿された 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。

インターネットブラウザ上の Twitter では、Twitter API による検索で取得できない期間やいいね・RT の数で絞り込むといった高度な検索が利用できる。高度な検索とは、検索結果を特定の期間や特定のユーザーなどに絞り込むことができ、探しているツイートも見つけやすくなる。検索結果は下にスクロールすることで古いものが読み込まれていく。このブラウザを使った検索を利用し、データを収集する [3]。

ブラウザの Twitter 検索結果からデータを取得するために 2 つのプログラムを作成する。プログラムの作成には Python を用いた。1 つ目のプログラムではスクロール作業と全体の HTML を保存する作業を自動化する。ブラウザのスクロール作業を自動で行う必要があるため、ブラウザの自動操作ができるライブラリである「Selenium WebDriver」を使用し、検索結果をブラウザに全て表示させてから HTML を保存する。2 つ目のプログラムでは、保存した HTML ファイルからツイートの本文と時間のみ抽出するため、HTML を解析・スクレイピングを行うことのできるライブラリである「BeautifulSoup4」を使ってデータを抽出する。

Twitter 検索で使用する検索ワードは「GitHub, 言語, 日付, 期間」を指定する。例えば、序論で述べた 2016 年 1 月 28 日のツイートを検索する場合、「GitHub lang:ja since:2016-01-28_00:00:00 until:2016-01-29_00:00:00_JST」となる。データを取得する日を特定するため、GitHub に関連するすべてのサービスを継続的に状況監視している「GitHub Status」の「Status Message」を参照し、2016 年で主要なサービスが停止、復旧したとアナウンスされている時間を調べる。その日のツイートを作成したプログラムを使って検索し、ツイートの時間と本文のみを抽出する。これを各障害の発生日ごとに繰り返す。

Investigation of the influence of failure of web service based on tweet analysis.

[†] Sho IWASE(s1442012ap@s.chibakoudai.jp)

[‡] Department of Project Management, Faculty of Social Systems Science, Chiba Institute of Technology.

4 結果

GitHub Status の Status Message を参照に調べたところ、2016 年の GitHub におけるサービス停止回数は 14 回であった（ただし、10 月 21 日の障害は同時に Twitter も障害が発生しており、どちらのサービスも不安定な状態だったため対象に加えない）。

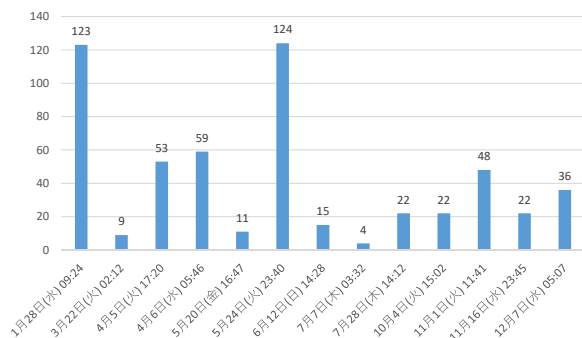


図 1 サービス停止から復旧までの間隔

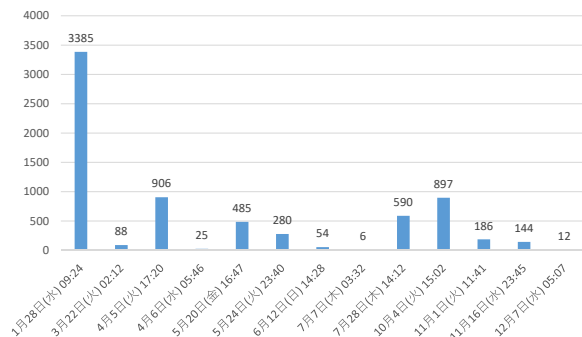


図 2 サービス停止中に投稿されたツイートの数

13 回分の各障害のサービス停止から復旧までの間隔を分単位でグラフにしたものが図 1 である。

そして、各障害のサービス停止中に投稿されたツイート数をグラフにしたものが図 2 である。

5 考察

調査した 13 回分の障害を比較すると、サービス停止から復旧までの間隔が同じくらいでも、1 日の時間帯によってツイート数に違いがあった。特にツイート数が多

かったのは平日の日中で、中でも会社への出勤や退勤にあたる時間帯だった。この時間帯に Web サービスが停止してしまうと例え数分の停止でもツイート数が多く、1 日のタスクが確認できなかったり、チーム内でのコミュニケーションが取れなかったりしてしまう。

平日の日中であればツイート数が多いのはもちろんだったが、3 月 22 日（火）の深夜 2 時台に発生した約 9 分間の障害でも 88 件のツイートが集まった。日中の障害でも 6 月 12 日（日）と 7 月 28 日（木）では同じ 14 時台の障害で停止時間に約 7 分の差はあるが 500 ツイート以上の差があった。これは日曜日に起きた障害であり、普段仕事をしている人々が休暇中だったため少なかったのではないかと考えられる。

また、GitHub Status の Status Message を参照してツイート取得する日を決定していたが、Status Message に記されているサービス停止や復旧のアナウンスよりも、サービス停止や復旧したという旨のツイートの方が平均して約 8 分ほど早かった。これは Twitter の速報性が優れていると言える。

6 結論

本研究では Web サービスの障害発生について調査した。Web サービスは利便性が高く、それだけに依存してしまうこともある。そのサービスが停止してしまうことは頻繁に有ることではなく、一時的なもののだが、その Web サービスを使用しているユーザーの仕事やコミュニケーションに少なからず影響を及ぼしてしまう。したがって、ソフトウェア開発プロジェクトなどで Web サービスを使用する場合は、サーバーダウン等の障害が発生するリスクを考慮する必要がある。

参考文献

- [1] 千葉祥太郎. 開発現場でホントに役立つツール (第 1 回) バージョン管理ツール「git」ソースコードは個人で管理 柔軟にワークフローを構築. 日経 SYSTEMS, No. 252, pp. 92–97, apr 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://www.atmarkit.co.jp/ait/articles/1601/28/news126.html> (2016.06.29 閲覧).
- [3] 鳥海不二夫. Twitter 上のビッグデータ収集と分析. 組織科学, Vol. 48, No. 4, pp. 47–59, 2015.