

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において、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「GitHub」や、コミュニケーションを取るためのチャットツール「Slack」である [1]。

これらのサービスの停止は、それを利用しているプロジェクトに大きな影響を与えられる。実際、2016 年 1 月 28 日の GitHub の停止時や、2017 年 11 月 1 日の Slack の停止時には、そのせいで仕事が進められなくなったというようなつぶやきが、Twitter 上で複数観測された [2]。

2 目的

Twitter の発言を収集するためのツールを開発し、それを用いてソフトウェア開発で利用される Web サービスの停止が開発に与える影響を調査する。

3 手法

本研究は以下の 2 段階で行う。

1. Twitter からツイートを収集するためのツールを開発する。
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。

初めに、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 が提供されており、その使用を試みたが、1 週間以上前のツイートは検索で取得できないという制限があった [3]。本調査では過去の障害発生時に発信されたツイートを収集しなかったため、API の仕様では実

現できなかった。そこで、インターネットブラウザ上の Twitter で使用できる「高度な検索」を利用した。高度な検索は、API による検索で取得できない期間やいいね・リツイートの数で絞り込むなどといったことが可能になっている。さらに、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。検索結果はブラウザを最下部まで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 Twitter の検索画面から Web スクレイピングをするためのツールを開発し、過去のツイート取得を可能にした。

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検索画面を最下部までスクロールする作業と全体の 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 を解析して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 の高度な検索では「キーワード、言語、日時、期間」を指定して検索する。例えば、序論で述べた 2016 年 1 月 28 日に発生した GitHub の障害について検索する場合、「GitHub lang:ja since:2016-01-28_00:00:00_JST 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ツールを使って検索し、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実行する。

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であったため対象に加えない）。

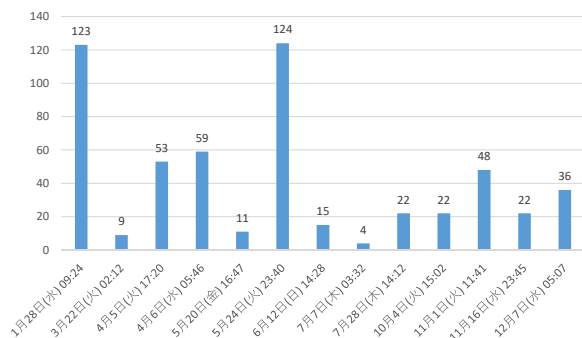


図1 サービス停止から復旧までの間隔

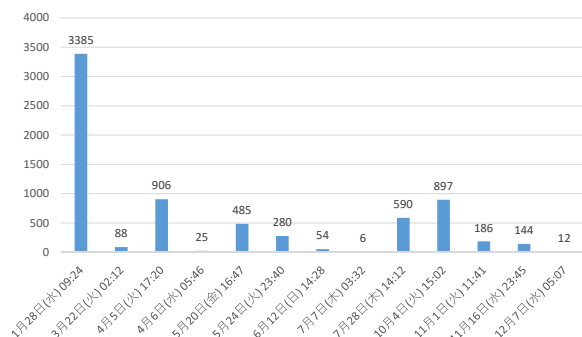


図2 サービス停止中に投稿されたツイートの数

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 回分の障害を比較してみると、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であった。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する。

平日の日中であればツイート数が多いのは勿論だが、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であ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することは頻繁に有ることではなく、一時的なものだが、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す。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.