

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。

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[1]。

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のAPIには、1週間以上前のツイートを検索して取得することができないという制限があるため、本研究のためには使えない[2]。そこで、インターネットブラウザ上のTwitterで利用できる「高度な検索」を利用する。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・言語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。本調査では「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の障害発生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。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。これを繰り返すことで、2016年1月28日に投稿された「GitHub」を含む日本語のツイートを全て表示できる。このブラウザを使ったTwitterの検索画面からツイートの本文と時間のみをWebスクレイピングするためのツールを開発し、過去のツイートを取得する。

ブラウザのTwitter検索結果からデータを収集するツールを開発するために2つのプログラムを作成する。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ページ全体をHTMLファイルで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ファイルで保存する。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、Pythonのライブラリであり、HTMLファイルを解析してスクレイピングを行うことのできる「BeautifulSoup4」を使ってデータを抽出する。

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。本研究では、そのうちの 13 回分を調査する。残りの 1 回は、同時期に Twitter にも障害が発生していたため、本研究の調査対象とはしない。

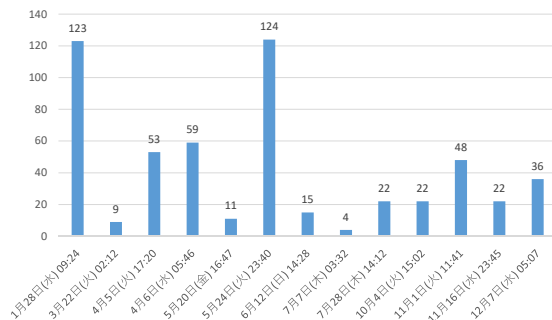


図1 サービス停止から復旧までの間隔

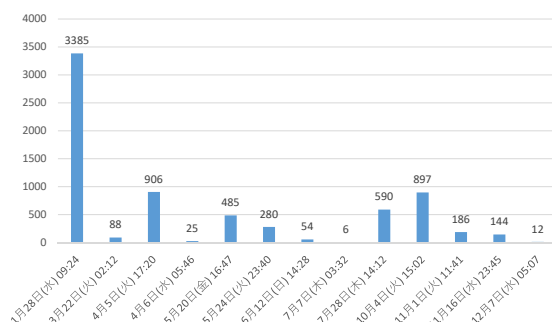


図2 サービス停止中に投稿されたツイート数

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ば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。1 月 28 日（水）が最も多い 3385 ツイートとなっているが、やはり 9 時台という出勤にあたる時間で 2 時間以上の停止をしたことが原因だと考える。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と考える。

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、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。この時間でも仕事にならないという反応が観測されている。

以上のことから、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、サービスが停止するリスクを考慮する必要があることがわかる。

また、本研究では 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、その Status Message に記されているサービス停止のアナウンスよりも、平均約 7 分早くサービス停止に関するツイートが観測されていた。これは、サービスの状態について運営元が発表している情報が必ずしも正しくはないことを示唆している。このことは、サービス停止がビジネスに重大な影響を及ぼす状況では、問題になり得る。

6 結論

Twitter のブラウザでの検索結果を保存するツールを開発し、それを用いて GitHub や Slack などのウェブサービスの停止に対する開発者の反応を調査した。その結果、日中はもちろん深夜でもサービス停止の影響は大きいこと、サービス運営元による停止時間についての発表は実際のそれとはずれていることがわかった。このようにウェブサービスの障害とその影響を調査することが、ウェブサービスを利用するソフトウェア開発のマネジメントにおいて有用な知見となることが期待される。

参考文献

- [1] 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).
- [2] 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.