

Twitter 発言の分析による Web サービス障害の影響調査

プロジェクトマネジメントコース 矢吹研究室 1442012 岩瀬翔

1. 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。GitHub 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 上に投稿された [2]。

2. 目的

GitHub 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。また、調査結果から Web サービスの障害発生に対するリスク対策案を検討する。

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、仕様によって過去のデータが取りきれないので以下の手順で行う。

1. 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」を参照し、2016 年で主要なサービスが停止、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 で検索し、プログラムを使って時間と本文のみを抽出する。これを各障害の発生時間ごとに繰り返す [3]。
2. サービスの停止から復旧までの 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。調査結果をもとに、リス

ク対策案を検討する。

4. 想定される成果物

GitHub におけるサービスの停止から復旧までに投稿されたツイート数及び、サービスの停止から復旧までの時間間隔でデータを集める。これらのデータから考察することにより、障害発生に対するリスク対策案を検討する。

5. 進捗状況

GitHub Status を参照に調べたところ、2016 年の GitHub における障害発生回数は 14 回であった（ただし、10 月 21 日の障害は同時に Twitter も障害が発生していたため対象に加えない）。それらの障害についてはデータの収集を完了した。

集めたデータから考察すると、サービスの停止した時間帯や時間間隔が様々であることがわかった。そして、停止から復旧までの間隔が同じくらいでも、時間帯によってツイート数に違いがあることがわかった。

6. 今後の計画

現段階では 2016 年のデータしか集めていないため、更に過去のツイートや 2017 年のデータを集めることを検討する。

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.