

Stack Overflowの回答更新 タイミングの推薦モデルの構築

オープンソースソフトウェア工学研究室

学籍番号 60230010

氏名 井町 瑠貴

背景 | Stack Overflow における回答の陳腐化



stackoverflow

人気のある技術系CQAサイト
膨大な量の知識を蓄積しており、
知識の共有や学習に用いられる

| 投稿された回答が、ソフトウェアの急速な進化により、役に立たなくなる場合がある



回答の陳腐化という

背景 | Stack Overflow における回答の更新

- | 陳腐化した回答は、ユーザにとって不利益なものであるため、更新を行う必要がある
- | しかし、回答者はいくつも回答している場合があり、自分の過去の回答がいつ陳腐化しているのかわからない可能性がある
- | 回答の更新タイミングを推薦することで回答の更新を促す

学部研究によって明らかになった課題

研究課題 1

- 投票数と更新数の変化に関するか？（別々の現象なのか？）

研究課題 2

- 更新時のテキスト内容と増加量と陳腐化に関するか？
- 更新時のテキスト内容と増加量を用いて陳腐化を検出できるか？

研究課題 3

- 投票数・更新数（＋更新内容・量）以外で陳腐化の検出に有用な特徴量はあるか？
 - 回答への参照数，コメント数，回答者のReputation
 - カテゴリ（≡タグ）毎の陳腐化の傾向の違い
（例：JavaScript関連の回答は陳腐化が早い）

研究課題 4

- 研究課題 1～3 で得られる知見を統合し，更新タイミングを推薦するためのモデルを構築できるか？

研究の目的とアプローチ

目的

- Stack Overflowの回答更新タイミングを推薦するモデルを構築すること

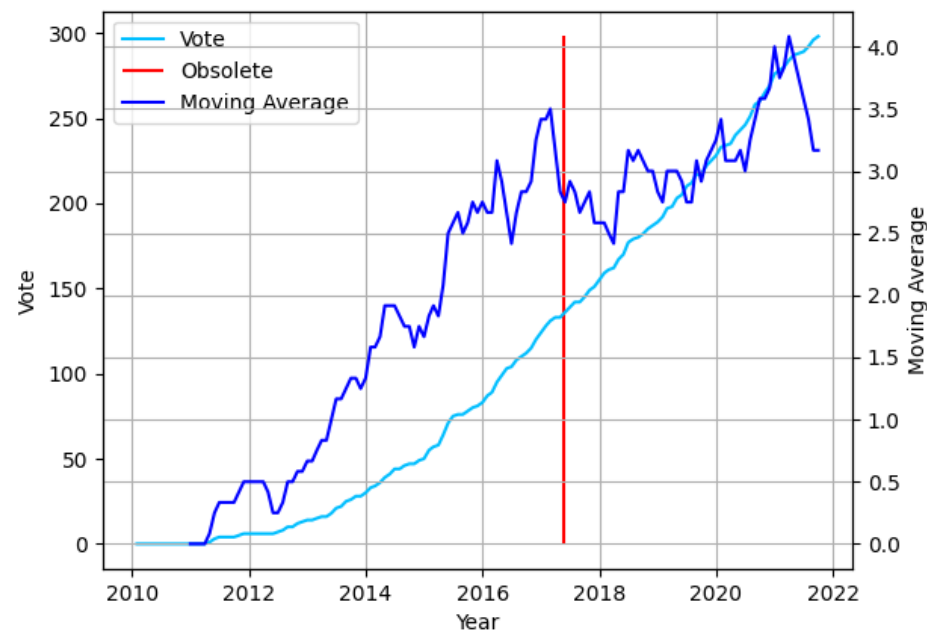
アプローチ

- RQ 1 : 投票数と更新数の変化に関係があるか？
- RQ 2 : 更新時のテキストと陳腐化に関係があるか？
- RQ 3 : 投票数・更新数（＋更新内容・量）以外で陳腐化の検出に有用な特徴量はあるか？
- RQ 4 : 回答更新タイミングを推薦するモデルを構築できるか？

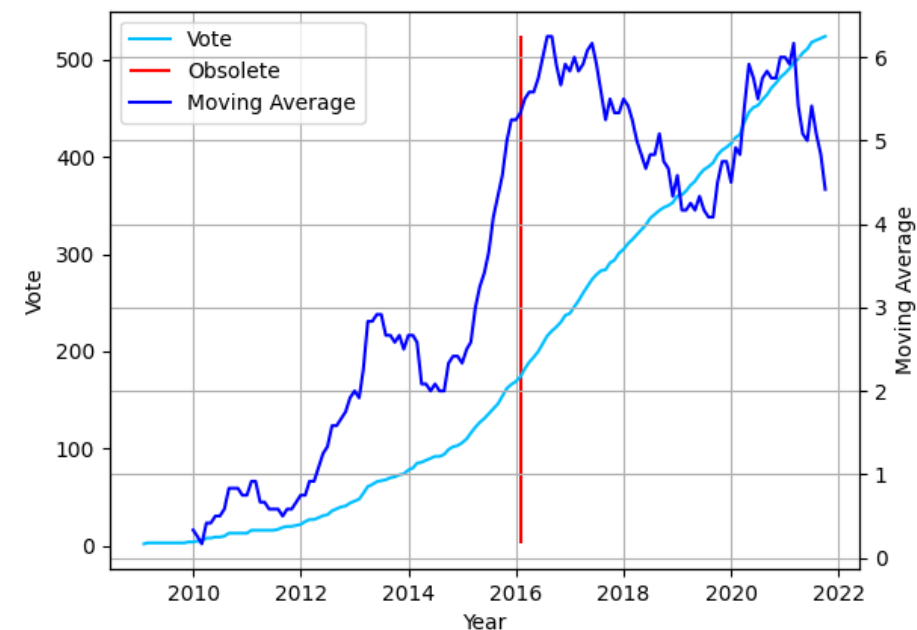
研究課題 1

投票数と更新数それぞれ単体では、
陳腐化した回答を十分には見い出せない

投票数



推測できる例

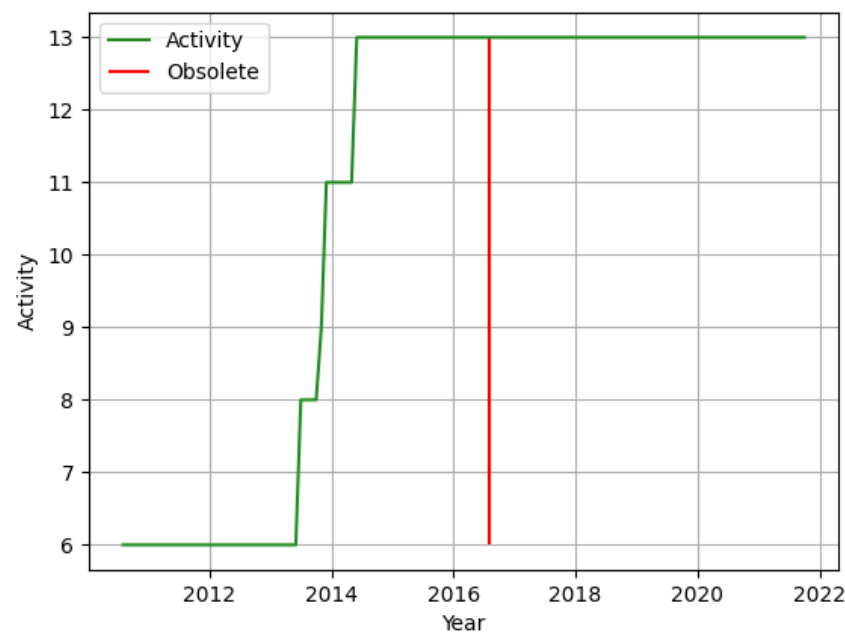


推測できない例

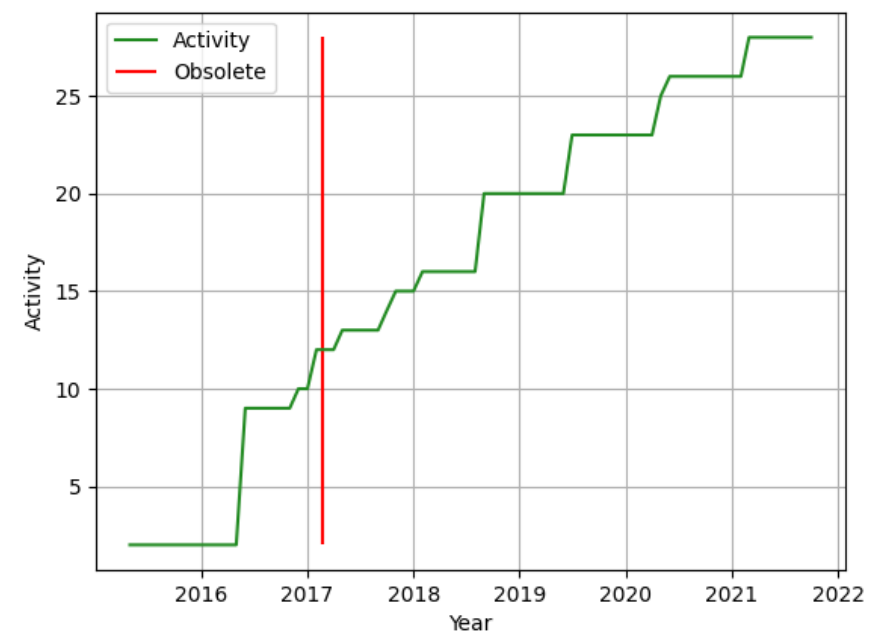
研究課題 1

投票数と更新数それぞれ単体では、
陳腐化した回答を十分には見い出せない

更新数



推測できる例

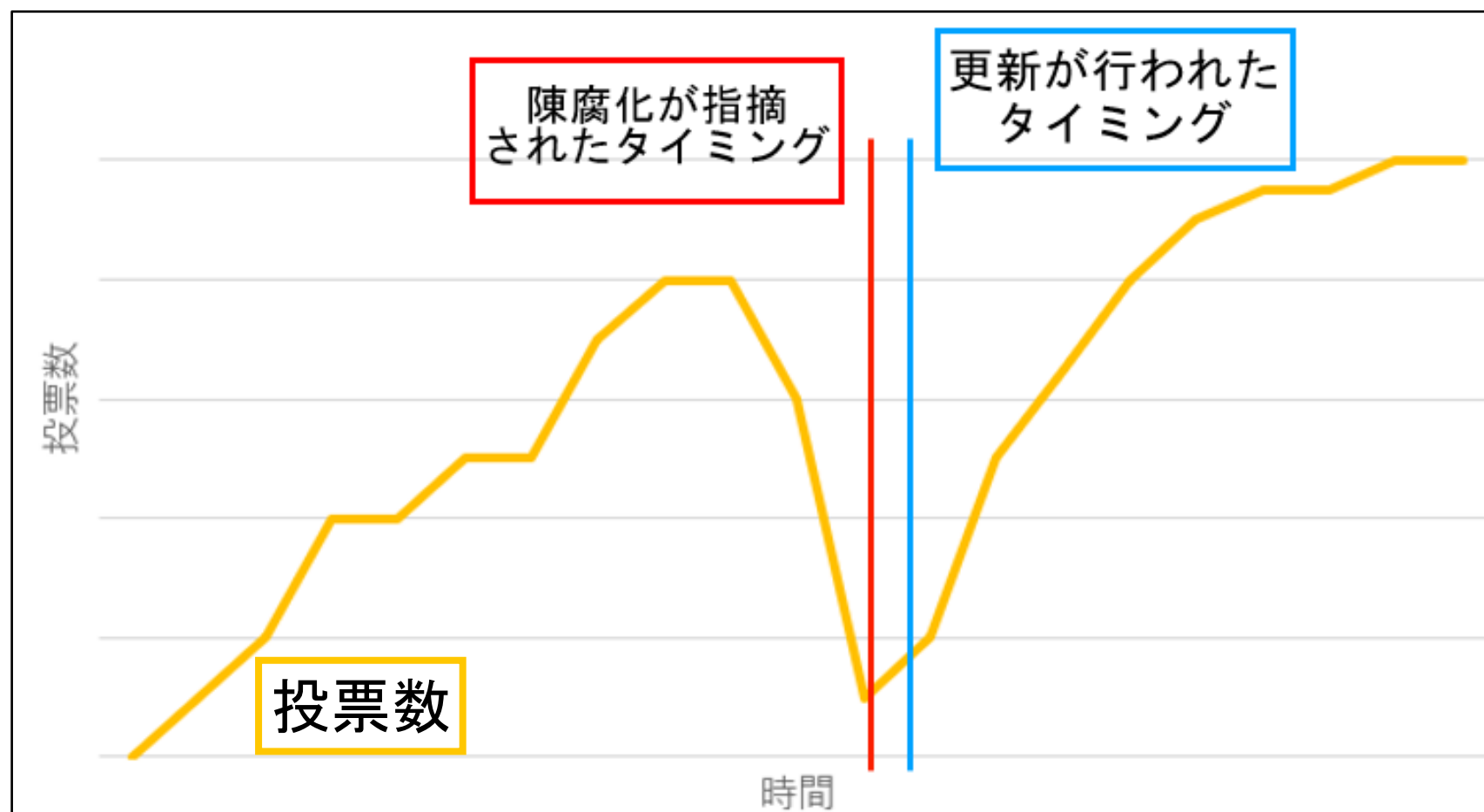


推測できない例

研究課題 1

どのように改善すべきか

- 投票数と更新数の関係を分析すれば陳腐化が発生しているタイミングをより正確に絞り込めるのでは？

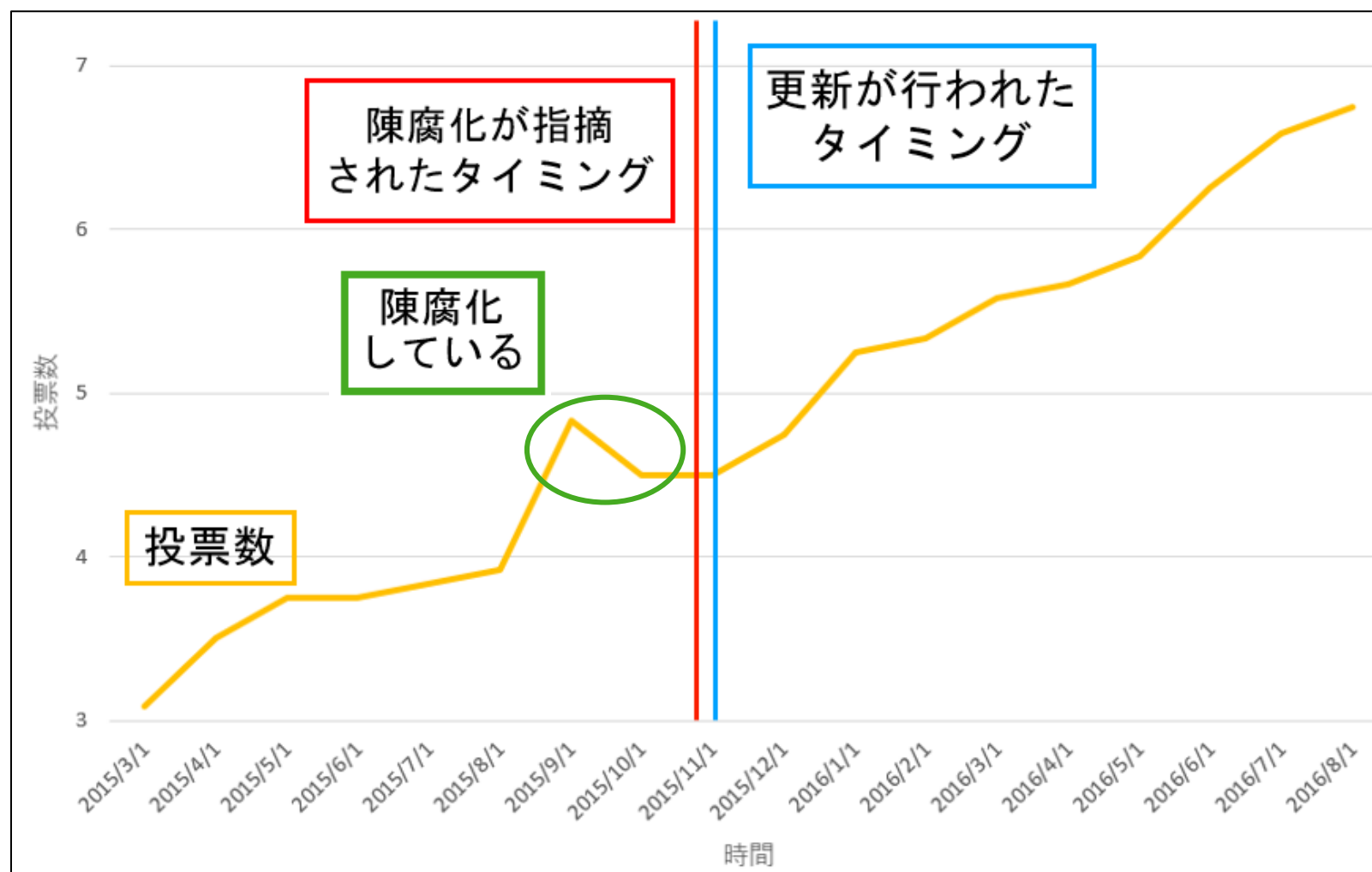


仮説のグラフ

調査結果の例（研究課題 1）

#13007325 : IPアドレスの取得方法^[4] (android)

- 2015/11/2 に指摘され, 2015/11/2に更新が行われている



[4] <https://stackoverflow.com/questions/6064510/how-to-get-ip-address-of-the-device-from-code/13007325#13007325>

今後の課題（研究課題 1）

| 分析の自動化

- 閾値の設定
 - 更新後の投票数の増加量を設定
 - 投票数の増加量が上位何番目以内かを設定
 - より効果的な設定方法がないか検討が必要
- 例外処理
 - 必ずしも、更新後投票数が増加するとは限らない
 - どのような処理をすべきか検討が必要

今後の予定

研究課題 1 (M1: 1Q2Q)

- 「投票数と更新数」と陳腐化の関係分析

研究課題 2 (M1: 3Q4Q)

- 「更新時のテキスト内容と増加量」と陳腐化の関係分析

研究課題 3 (M2: 1Q2Q)

- 陳腐化の検出に有用なその他の特徴量の分析
- カテゴリ (≡ タグ) 毎の陳腐化の傾向の違いの分析

研究課題 4 (M2: 3Q4Q)

- 回答の更新タイミングを推薦するためのモデルの構築

大学院生MTGを踏まえて

背景 | Stack Overflow における回答の陳腐化



stackoverflow

人気のある技術系CQAサイト
膨大な量の知識を蓄積しており、
知識の共有や学習に用いられる

| 投稿された回答が、ソフトウェアの急速な進化により、役に立たなくなる場合がある



回答の陳腐化という

背景

| Stack Overflow になんてほしいのか

- ユーザが, 「(陳腐化している) 回答の手法を試してうまくいかなかった」 になんてほしくない



| Stack Overflow とユーザの間に入る

- 陳腐化していることをユーザに知らせる
- 提案するツールで検索を行うと, 各回答が陳腐化している可能性も表示できるイメージ

陳腐化した回答

| すべての回答をまとめて考えるのは不可能

- 回答を分けて考える必要がある

| どのように分けて考えるか

- 本質的な問題の所在
（例：OS, ライブラリなど）
- 同じようなタイミングで更新が行われている
（≡同じバージョンを使っている回答）

今後の予定

| 現在行っていること

- Stack Overflowでの目視
 - 付与されているタグと本質的な原因に違いがあるか
 - 本質的な原因の種類や割合

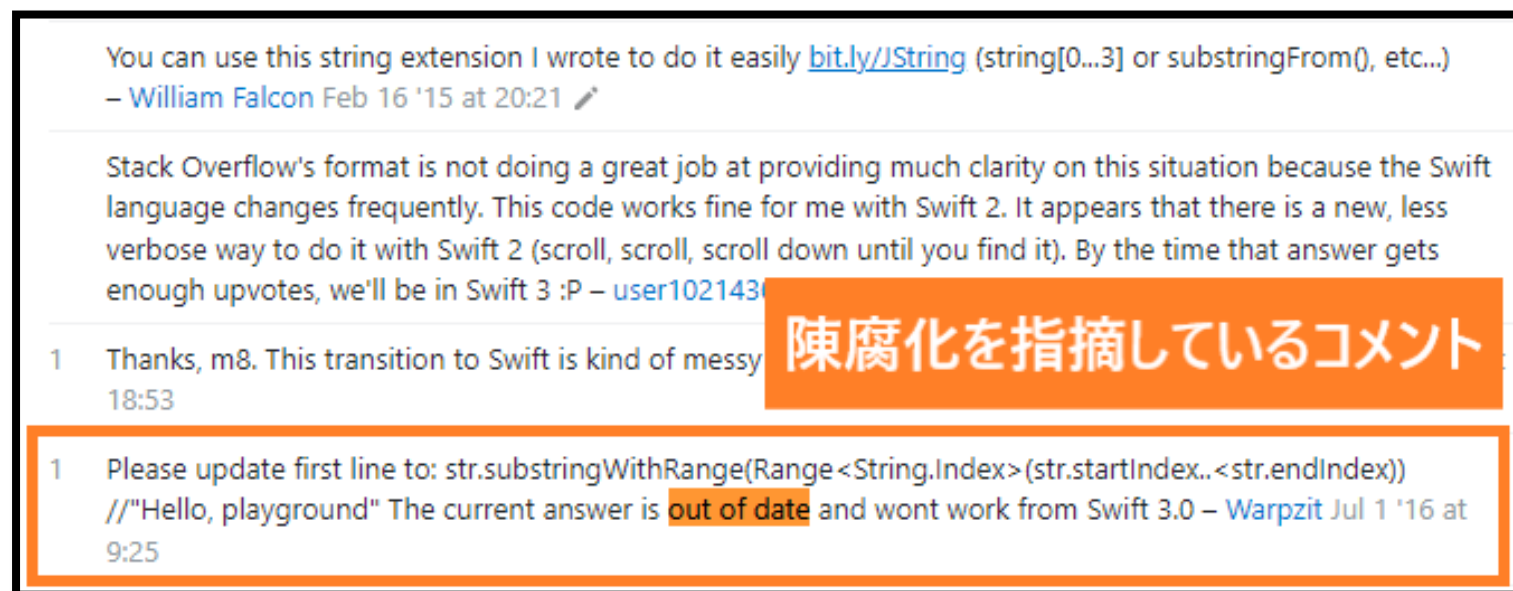
補足資料

補足. Stack Overflow | コメント機能

コメント機能

- 質問や回答にコメントすることができる
- ユーザーが回答の内容が陳腐化していることをコメントで指摘することがある

例) 回答が時代遅れであり機能しないため、
回答を更新するようにコメントで提案している^[5]

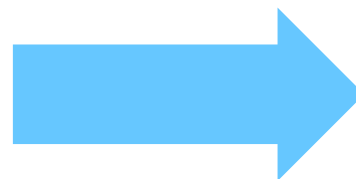


[5] https://stackoverflow.com/questions/24044851/how-do-you-use-stringsubstringwithrange-or-how-do-ranges-work-in-swift/24045156#comment63713320_24045156

補足. Stack Overflow | 投票機能

投票機能

役に立つと思われる
投稿に投票



スコアで並び変えられ
閲覧されやすくなる

スコア = プラス投票数 - マイナス投票数

```
var str = "hello, playground"  
str.substringWithRange(Range<String.Index>(start: str.startIndex, end: str.startIndex.advancedBy(1)))
```

To change the start and end index, use `advancedBy(n)`.

```
var str = "Hello, playground"
```

投票機能の例^[6]

[6] <https://stackoverflow.com/questions/24044851/how-do-you-use-stringsubstringwithrange-or-how-do-ranges-work-in-swift/24045156#24045156>

補足. Stack Overflow | 投稿後の更新

- | 投稿後も、ユーザは投稿の更新を行うことができる
- | 行われた更新はリビジョン履歴として記録されている



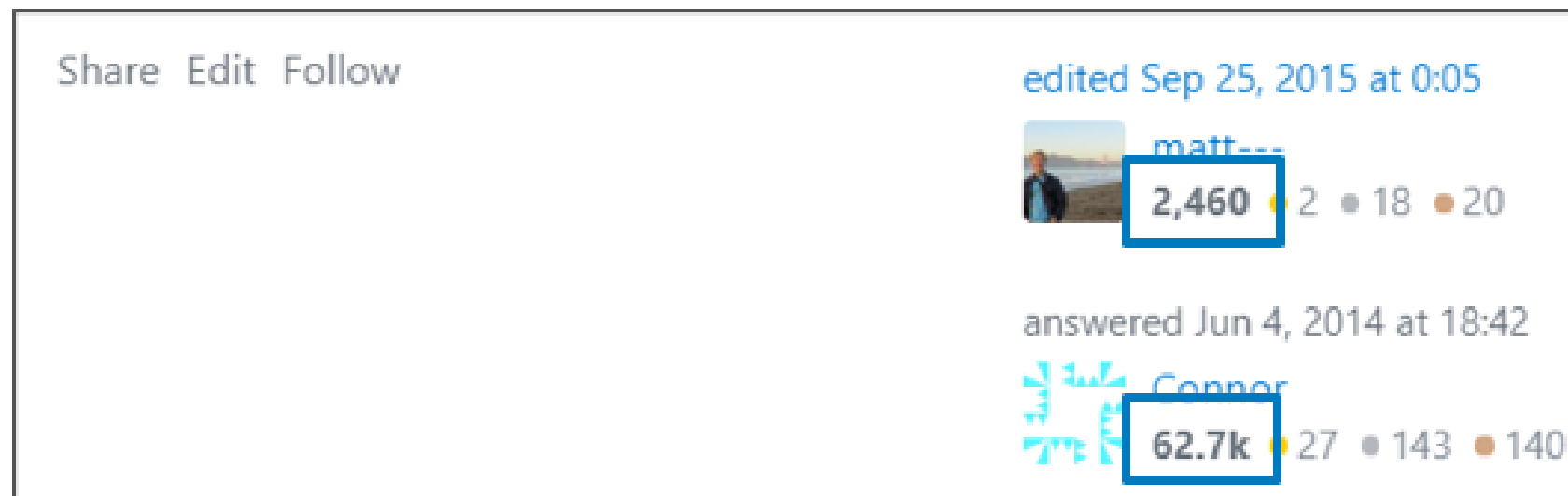
リビジョン履歴の例^[7]

[7] <https://stackoverflow.com/posts/24045156/revisions>

補足. Stack Overflow | Reputation

Reputation (信用度)

- コミュニティがユーザをどれだけ信頼しているかを大まかに測定したもの
- 良い質問や、良い回答を投稿することで獲得することができ、信用度に応じて、タグの作成などの権限が与えられる



Reputationの例^[8]

[8] <https://stackoverflow.com/questions/24044851/how-do-you-use-stringsubstringwithrange-or-how-do-ranges-work-in-swift/24045156#24045156>