

プログラミング QA サイトにおける閲覧者の視点に着目した回答の予備調査

三野 晃季 福安 直樹 満田 成紀 鯨坂 恒夫

ソフトウェアの開発者同士が質問と回答を行い知識を共有するための QA サイトが広く利用されている。こうした QA サイトでは、利用者の大半は質問や回答をしない閲覧者である。一方で、知識の共有は質問者と回答者のやり取りによって行われることから、そのやり取りは直接的には閲覧者に向けられておらず、また回答の閲覧者にとってのユーザビリティは質問者や回答者に比べてそれほど高くないという現状がある。そこで、質問者と閲覧者の双方にとって良い回答を明らかにすることを目的とし、本研究では、QA サイトの一つである Stack Overflow により提供されるデータダンプを用いて、予備調査を行った結果を報告する。

1 はじめに

プログラミング中の開発者が自身の持つ問題を解決するために情報を収集する手段は複数ある。公式に Web で公開されるドキュメントを参照することや、書店で販売される書籍を参照するといった手段の他、Google などの検索エンジンを用い、Web 上の公式のものでない資料を参照するといったことも近年では広く行われている。

こうした Web を活用した情報収集には、ブログなどのユーザが自由に記事を投稿可能な個人サイトの参照や、Qiita などの投稿型コミュニティサイトの参照など、ユーザが自らの知見を投稿し共有する投稿型の Web ページの閲覧といった手段に加え、ユーザが自らの疑問や解決困難な事象についての質問を投稿し、その問題の解決策や提案を持ったユーザが回答を提示する、Yahoo!知恵袋や Stack Overflow, teratail などの QA サイトと呼ばれる Web ページを活用する

手段がある。こうした QA サイトは質問者が実際に直面した問題についての議論がなされるため、現実に即した事例ベースの情報が得られやすい利点があり広く利用されている。また、QA サイトでは自ら質問を投稿する他、既に過去に投稿された質問と回答を参照し自らの問題解決に利用するといった活用方法もある。こちらも一般的に行われており、一例を挙げると Stack Overflow ではサイトの訪問者の大部分がユーザ登録を行っておらず [1]、サイトの訪問者の 90%以上が検索エンジンを用いてアクセスしている [2]。

しかしこうした QA サイトはその特性上、サイトのコンテンツはあくまで質問者と回答者のやり取りに過ぎない。つまり、利用者の大半が質問や回答を行わない閲覧者でありながら、サイトのコンテンツはその閲覧者に向けられたものではなく、閲覧者は自分自身に向けられていない回答から自分にとって必要な回答を探す必要がある。多くの閲覧者が日々参照し情報源とする大規模な QA サイトではこの問題は決して小さなものではない。

そこで本研究では情報技術、特にプログラミングに関する技術を中心とした開発者向けの QA サイトである Stack Overflow を対象に、最終的な目標を以下の二点に設定した。

1. 閲覧者の視点を考慮し、閲覧者にとっても良い

a preliminary survey of programming QA site answers focusing on perspectives of visitors

Koki Mino, 和歌山大学大学院システム工学研究科, Dept. of Systems Eng., Wakayama University.

Naoki Fukuyasu, Naruki Mitsuda, Tsuneo Ajisaka, 和歌山大学システム工学部, Fact. of Systems Eng., Wakayama University.

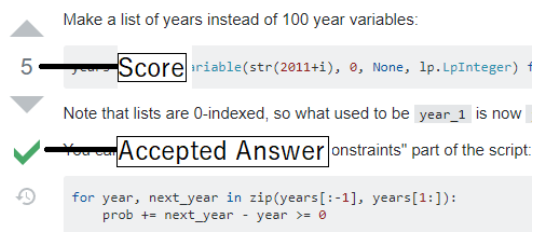


図 1 Stack Overflow における評価 [4]

Stack Overflow の回答の特徴を明らかにする

2. その結果に基づき回答のガイドラインを作成する

本稿では、そのために行った予備調査の結果について報告する。

2 Stack Overflow

2.1 Stack Overflow について

Stack Overflow は Stack Exchange, Inc. によって 2008 年に開設された情報技術、特にプログラミングに関する技術を中心とした開発者に向けた QA サイトである。開設から今日に至るまでに 1900 万件以上の質問が投稿され、月間訪問者数は 1 億 2000 万人を超える極めて規模の大きなコミュニティである [3]。Stack Overflow 上では、ユーザによって質問は Score を用いて、回答は Score と Accepted Answer の 2 つを用いて評価される。図 1 は回答の一例である。質問と回答どちらにも存在する Score は閲覧者による評価であり、回答のみがもつ Accepted Answer は、質問者が最も良いと認めた 1 つの回答につけることができるものである。また Web ページ上で表示される質問と回答を含んだスレッドでは、質問に続いて Accepted Answer が表示され、その下にその他の回答が Score 順に続くといった構造となっている。

2.2 Stack Overflow を対象とした関連研究

ここでは、Stack Overflow の質問や回答の品質とその特徴を主題とした研究において、どのような基準で質問や回答を分類したかについて述べる。

2.2.1 質問の分類

Duijn らは、Stack Overflow の質問に含まれるコードの特徴から低品質な質問を識別することを試みた [5]。この研究において、質問は以下の 2 つの指標で分類されている。

- 未編集かつ Score が 0 未満である質問を低品質、未編集かつ Score が 0 以上である質問を高品質とみなす。
- 未編集であるものを低品質、編集済みであるものを高品質とみなす。

2.2.2 回答の分類

Nasehi らは、プログラミング学習を支援する効果的なコードサンプルの特徴を発見することを目的とし、Stack Overflow 上の評価の高い回答を分析しコードの特徴を明らかにすることを試みた [6]。この研究において、良い回答を以下のいずれかに当てはまるものと定義している。

- Accepted Answer であるもの
- Accepted Answer でないもののうち正規化された Score が 0.4 以上であるもの
- Score が 10 以上であるもの

また Masson らは、Stack Overflow 上の多様な回答をカテゴライズすることを目的とし、Stack Overflow 上のデータからは観測しづらい閲覧者に目を向けた分析を行った [7]。この研究では、閲覧者が好む回答の特徴を得るために調査対象とする回答を、Accepted Answer であるが Score が最も高い回答が他にある回答と Score が最も高いが Accepted Answer でないもの、つまり質問者と閲覧者による評価が一致しないもののみとした。

2.2.3 関連研究の比較

2.2.1 節では質問を、2.2.2 節では回答を対象とした調査を行っているものを取り上げた。これらの Stack Overflow を対象とした調査では、文献 [7] 以外に、閲覧者の視点を考慮し Accepted Answer と Score を得た回答を区別し分析した研究はなく、また文献 [7] においても Accepted Answer であるが Score が最も高い回答が他にある回答と Score が最も高いが Accepted Answer でないものの比較という限定的なものにとどまっている。そこで本研究では閲覧者の

視点を考慮しつつ、対象とする回答の範囲をより拡張した分析を行う。

3 閲覧者に着目した回答の調査

3.1 調査対象

本研究は Stack Exchange, Inc. によって提供される Stack Overflow のデータダンプ [9] を利用し分析を行った。2019 年に Stack Overflow に投稿された 1,421,073 件の質問と、それらに対する 2,051,044 件の回答を調査対象とした。また、前述の関連研究との比較から、調査対象とした回答を以下のように分類した。

- Accepted and Scored
 - Accepted Answer であり、Score が最も高い回答
- Accepted
 - Accepted Answer であるが、Score は最も高くない回答
- Scored
 - Accepted Answer ではないが、Score が最も高い回答

Accepted Answer は質問者による評価であり Score は閲覧者による評価であるため、それぞれ質問者と閲覧者の両者に評価された回答、質問者にのみ評価された回答、閲覧者にのみ評価された回答であると考えられる。また、これらとの比較の指標とするため、上記の 3 つ全ての分類と、分類に当てはまらないその他のものも含めた回答を分析した ALL という区分も調査対象とした。

3.2 調査方法

今回この研究の予備調査として、文献 [6] 及び文献 [7] において取り上げられているコードの含有率と、文献 [7] において取り上げられている質問と回答のコードの類似度について調査を行った。

3.2.1 コードの含有率

コードの含有率を調査するにあたって、コードの定義を図 2 のような、Stack Overflow のデータダンプ上では `<code>~</code>` で括られ表示の上では灰色のハイライトが挿入されているものとした。また図

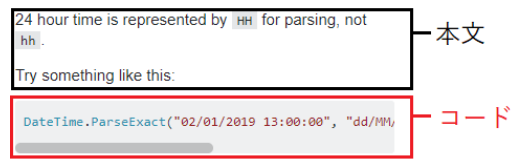


図 2 文章とコードの区分 [8]

中の HH や hh など同様のハイライトがあり、データダンプ上では同様に `<code>~</code>` で括られおりコードであるとみなすこともできるが今回は対象外とした。これらは、Stack Overflow のページ上の回答入力欄では、前述の定義したコードとは入力方法が異なり、コードであると明示して入力するものもないからである。

このような条件で抽出した本文とコードから、全体のうちのコードが占める割合を文字数で比較し求めた。

3.2.2 質問と回答のコードの類似度

質問と回答のコードの類似度の調査では、文献 [7] に従い Jaccard 係数を用いて比較を行うこととした。Jaccard 係数は比較するそれぞれの文章から重複を取り除いた語群を作成し、その積集合を和集合で除算することで求められる。つまり Jaccard 係数は純粋に単語群を比較するのみであるため、語順などの要素は無視されることとなる。またコードを単語に分割し語群を作成するにあたり、多くのコードが共通して持ち、それ単体では意味を成さない例が大半であると考えられる記号類を全て除去した。

3.3 調査結果

3.3.1 コードの含有率

まずコードを含む割合を求める前に、コードを含む回答の割合について調査した。その結果が図 3 である。全体 (ALL) ではコードを含む割合は 71.3%であったのに対して、Accepted and Scored では 83.0%と、10 ポイント以上の差があるという結果を得た。また Accepted と Scored に関しては、やや Accepted がコードを含むものが多いものの、それぞれ約 80.5%と 79.0%と約 1.5 ポイント程度の小さい差となった。

次に、実際にコードが占める割合について調査し

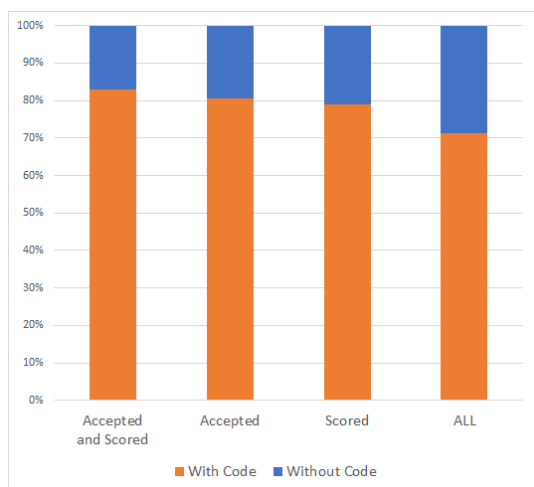


図3 コードを含む回答の割合

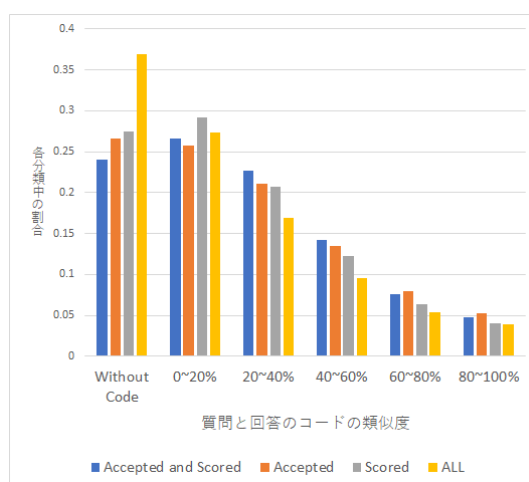


図5 回答のコードの割合

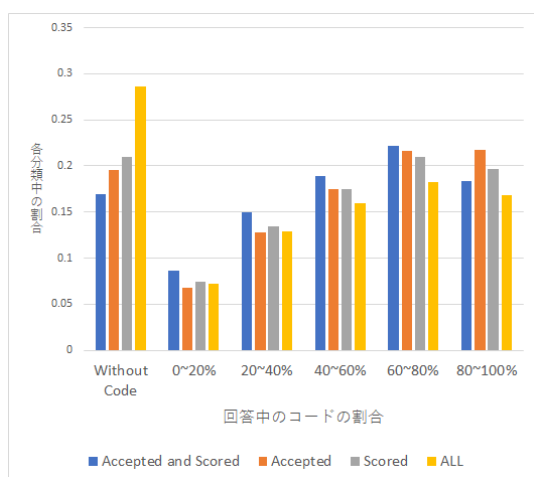


図4 回答のコードの割合

た。その結果が図4である。Accepted and Scored, つまり質問者と閲覧者の両方に評価される回答は、コードが極端に多いものも好まれないことが伺える。また、Accepted と Scored を比較すると、グラフの形状は類似しているものの Scored はコードが少ないものが、Accepted はコードが多いものが多い傾向にあることがわかる。

3.3.2 質問と回答のコードの類似度

それぞれの分類において Jaccard 係数による分析を行い図5のような結果を得た。図における With-

out Code は質問と回答いずれかにコードが含まれておらず、Jaccard 係数では比較不可能なものである。Accepted と Scored の関係に注目すると、コードの割合と同様にグラフの形状は類似しているものの、コードの類似度が低いものは Scored, つまり閲覧者によって評価される傾向があり、類似度が高いものは Accepted, つまり質問者によって評価される傾向があることがわかる。

4 考察

今回得られた結果の中でも、特に Accepted と Scored, つまり質問者もしくは閲覧者のどちらかのみが評価した回答の結果に注目しそれぞれを比較すると、

- 質問者が評価する傾向にあるもの (Accepted)
 - コードを多く含む
 - 質問と回答のコードが類似している
- 閲覧者が評価する傾向にあるもの (Scored)
 - コードが比較的少ない
 - 質問と回答のコードが類似していない

といった傾向があることがわかる。コードの含有率について着目すると、コードとそれ以外の本文は多くの場合回答と解説・補足のような関係にあり、質問者は単に自分の提示した質問に対する答えや提示したコードの修正案のみを必要としており、一方、質問者

とは異なる環境での疑問の解決策を求める閲覧者は現象そのものやその解決策に関する解説などより具体的に説明的な内容の回答を求めているのではないかと推測することができる。また、同様に質問と回答に含まれるコードの類似度について着目した場合も、質問者は質問者自身が提示したコードが持つ問題が解決できればそれでよく、一方の閲覧者はより一般化された問題に対する解説を求めているのではないかと推測することができる。これらの結果から、現時点で得られたデータからは質問者は答えを求め、閲覧者は知識を求めているのではないかと考えられる。

5 おわりに

本稿は最終的な目標を達するための予備調査という位置づけにある。現時点の調査によって判明した結果は、目標を達するために十分なものであるとは言えずより調査が必要である。また関連研究での分析では、コードの分割数や太字などの装飾のフォーマットの利用の程度など今回調査ができていない観点での調査も多く残されているため、まずはその調査を行うことを当面の目標としたい。

参考文献

[1] Mamykina, L., Manoim, B., Mittal, M., Hripacsak, G., Hartmann, B.: Design lessons from the

fastest q&a site in the west *Proc. CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 2857–2866.

[2] Jeff Atwood, Stack Overflow 2010 Analytics - Stack Overflow Blog, <https://stackoverflow.blog/2010/12/15/stack-overflow-2010-analytics>

[3] About - Stack Overflow, <https://stackoverflow.com/company>

[4] user2357112 supports Monica, linear programming - How to simplify repetitive Python PuLP syntax? - Stack Overflow, <https://stackoverflow.com/questions/17754363>

[5] Duijn, M., Kucera, A., Bacchelli, A.: Quality Questions Need Quality Code: Classifying Code Fragments on Stack Overflow, *Proc. 2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*, IEEE, 2015, pp. 410–413.

[6] Nasehi, S., Sillito, J., Maurer, F., Burns, C.: What Makes a Good Code Example? A Study of Programming Q&A in StackOverflow, *Proc. CS846 Final Project, Fall 2019, Waterloo, ON, Canada*, 2019.

[7] Masson, D., Melo, G.: What Answer for Which Audience? Identifying Categories of Answers on Stack Overflow, *Proc. 2012 28th IEEE International Conference on Software Maintenance (ICSM)*, IEEE, 2012, pp. 25–34.

[8] Jonathon Chase, `c# - string not recognized as valid datetime with same format` - Stack Overflow, <https://stackoverflow.com/questions/54012706>

[9] Stack Exchange Data Dump : Stack Exchange, Inc. : Free Download, Borrow, and Streaming : Internet Archive, <https://archive.org/details/slashstackexchange>