

プログラマー向け QA サイトにおける質問形式に関する研究

小瀬川 博将

現代社会において、インターネットを介したコミュニケーションや情報共有は必要不可欠であり、QA サイトを利用することは一つのツールとして知られている。特に、QA サイトを用いた情報共有は、日常的に発生する困り事や疑問など様々の問題解決に有効な手段である。プログラマー向け QA サイトに Stack Overflow というサイトがある。Stack Overflow における、回答に関する方針は「回答を得ることがこのサイトの全て」、「将来その問題を検索する人のためにもなるような回答ができればベスト」と定めている。また、質問に関しては「主観性の高い質問は避ける」、「質の低い質問を避ける」と定めている。このため、Stack Overflow では、質問や回答の質を向上する不明確な質問には質問の編集またはコメントを追加できる機能が提供されている。

しかし、このような機能では質問解決までに多くの時間と労力が必要となる。また、質問の質や有用性を判定するために、利用者による投票機能も用意されている。投票の機能によってマイナスの投票がされる質問は下位となる。Stack Overflow では「良い質問の仕方を学ぶのは価値のあること」ということを目指しているが、マイナス投票が多い質問に対して、役立つ情報提供はあっても原因の提示はされていない。

本論文は、未回答質問における記述の不明確さや回答が付けられていない原因を明らかにすることを目的とする。また、利用者間にかかる時間と労力を軽減し、QA サイトをより良い問題解決の場としていくことに貢献する。このため、未回答質問における質問解析を行った。投稿される質問の多くは、動作の不具合や使い方に関する解決法を求めていることから、動作環境や不具合の症状、質問するまでに調べた状況などの情報が不可欠であると考えられる。これらの情報が十分に記述されているかを調査し、回答を得るために有効であるかを検討した。この結果、未回答質問の本文に関しては、関連する質問の URL が記載されていないことや質問文中に箇条書きや空行がないという特徴が見られた。未回答質問に対して付与されたコメントに関しては、開発環境は何かを質問文に追記させるコメントや問題に至るまでの調査の確認がされているという特徴が見られた。以上のことから、質問記事を解析して、動作環境や不具合に関する具体的な記述がなされているか、また、質問記事の書き方あるいは形式が適切であるかを判別することで、回答を得るのに有用な質問記事の書き方に対する支援が行えると期待される。

(指導教員 佐藤 哲司)

プログラマー向けQAサイトにおける
質問形式に関する研究

2016年3月

201413130

小瀬川 博将

筑波大学情報学群
知識情報・図書館学類

目次

第 1 章	序論	1
1.1	QA サイトに関する背景	1
1.2	Stack Overflow における課題	2
1.3	研究目的	3
1.4	本論文の構成	3
第 2 章	関連研究	4
2.1	QA サイトの分析に関する先行研究	4
2.2	QA サイトにおける推薦に関する先行研究	5
第 3 章	質問形式の分析手法	7
3.1	使用するデータ	7
3.2	回答件数に関する調査	8
3.3	未回答質問の特徴抽出	9
第 4 章	未回答質問となる原因の検討	13
4.1	未回答質問の特徴付け	13
4.2	未回答質問の特徴付けを行った結果	14
4.3	認証済み質問の特徴付けを行った結果	20
4.4	未回答質問と認証済み質問の比較	24
4.5	未回答質問の特徴についての考察	24
第 5 章	結論	26
	謝辞	27
	参考文献	28

第1章 序論

本章では，研究を行う上での背景として，QA サイトの具体的な例を二つ挙げそれぞれのサイトの方針を説明する．研究の背景から問題点を明確化し，研究の目的を述べる．

1.1 QA サイトに関する背景

現代社会において，インターネットを介したコミュニケーションや情報共有は必要不可欠であり，QA サイトを利用することは一つのツールとして知られている．QA サイトとは，電子掲示板上で質問の投稿や解答が利用者によって行われ，質問・回答が蓄積される Web サイトである．また，QA サイトを用いた情報共有は問題解決に有効な手段である．ここで，QA サイトの具体例として二点挙げる．また，それぞれのサイトにおける方針を説明する．一つ目は，QA サイトの代表例でもある Yahoo!知恵袋を挙げる．二つ目は，プログラマー向けに作られた QA サイト Stack Overflow を挙げる．以下に示す 1 は Yahoo!知恵袋の方針¹であり，2 は Stack Overflow の方針²である．

1. Yahoo!知恵袋の方針は，コミュニケーションを通じて利用者が参加し知恵を共有する場を提供することである．そのコミュニケーションの中で知恵を交換していく QA サイトである．このことから，一つの質問・回答を類似質問にも対応するような質問・回答に書き換え，残していくことよりも QA サイトという場を利用し，コミュニケーションをとるということを重要視している．また，一度投稿した質問・回答の編集を基本的には許しておらず，類似した質問・回答を投稿する場合は再度投稿する必要があるので，質問・回答の質の向上よりも，数多くの質問とその回答を残していくことを目指しているといえる．
2. Stack Overflow では，回答と質問に関してそれぞれ方針を示している．回答に関しては「回答を得ることがこのサイトの全て」，「将来その問題を検索する人のためにもなるような回答ができればベスト」と定めている．質問に関しては「主観性の高い質問は避ける」，「質の低い質問を避ける」と定めている．このようなことから，投稿される質問・回答の数より質を重要視している．また，どの質問に対しても回答

¹<http://chiebukuro.yahoo.co.jp/docs/wish.html>

²<http://ja.stackoverflow.com/tour>

が用意されており、利用者が閲覧したとき納得できる情報の共有・蓄積していくことを目指しているといえる。

Yahoo!知恵袋と Stack Overflow の違いとして、Stack Overflow には利用者全員が積極的に参加し、良質な質問・回答を残すために内容を編集する機能が付けられている。また、Yahoo!知恵袋における最も良い回答に付けられる「ベストアンサー」に対して、Stack Overflow では「認証済み」が付けられる。認証済みとは、その回答が完璧な回答であることを示すのではなく、基準は質問者に役に立ったかどうかであることからベストアンサーと認証済みとは、ほぼ同等の機能といえる。Stack Overflow の現状として、未解決や未回答のままの質問が残されている。特に、主観性の高い質問や質の低い質問（Stack Overflow に向かないタイプの質問）では編集が必要である。Stack Overflow に向かないタイプの質問の問題解決には質問の編集や内容を編集するようなコメントが行われている。また、Stack Overflow に向かないタイプの質問と判断された場合、その質問がクローズ（一定の基準を満たしたユーザがクローズ票を投票することで回答が受け付けられなくなる状態）される。クローズされた質問の問題解決も誰かの手によって、質問の編集や内容を編集するようなコメントが行われている。

本論文では、Yahoo!知恵袋ではなく Stack Overflow を扱う。なぜなら、Stack Overflow には良質な質問・回答を残すために内容を編集する機能が付けられているにもかかわらず、Stack Overflow に向かない質問が未解決や未回答のまま残されているという課題があることによる。Yahoo!知恵袋には、基本的に質問の編集機能がないため扱わないことにする。また、Stack Overflow を扱うがプログラマー向け QA サイトであれば、共通の課題だといえる。

1.2 Stack Overflow における課題

背景で述べたように、プログラマー向け QA サイトである Stack Overflow の現状として、未解決や未回答のまま質問が残されている。未解決や未回答のまま質問が残されている原因に関して以下の2点を上げる。

- (1) 質問内容が高度な場合
- (2) 質問者が投稿する質問内容が不明確な場合

原因 (1) を解決するためには、甲谷ら [1] が回答者に適切な質問を推薦するシステムを提案している。また、原因 (2) に対して提案システムによる推薦を行っても解決が困難である。原因 (2) は、適切な質問として推薦を行っても回答までに補足情報を必要とする。原因 (2) について、Stack Overflow では不明確な質問に対して、質問の編集またはコメントを追加できる機能が提供されている。

しかし、このような機能では、質問解決までに多くの時間と労力が必要となる。また、投票という機能が用意されており、利用者はプラスかマイナスの票を入れることができる。投票の機能によってプラスの投票がされるような良質な質問は上位となるが、マイナスの投票がされる質問は下位となる。Stack Overflow では「良い質問の仕方を学ぶのは価値のあること」ということを目指しているが、マイナス投票が多い質問に対して、役立つ情報提供はあっても原因の提示はされていない。さらに、プラスの投票があっても回答が付与されていない質問も残されている。

1.3 研究目的

前節の問題点を解消するため未回答質問における質問解析を行う。

本論文は、未回答質問における記述の不明確さや回答が付けられていない原因を明らかにすることを目的とする。問題点で述べたように、質問解析として未回答質問における特徴付けを行い、質問として成り立つには何が書かれてればよいのかを調査する。

具体的には、Stack Overflow において投稿される質問の多くは、動作の不具合や使い方に関する解決法を求めていることから、動作環境や不具合の症状、質問するまでに調べた状況などの情報が不可欠であると考えられる。これらの情報が十分に記述されているかを調査し、回答を得るために有効であるかを検討する。回答が付けられていない原因を明らかにすることによって、利用者間にかかる時間と労力を軽減し、QA サイトをより良い問題解決の場としていくことに貢献する。

1.4 本論文の構成

以降の本論文について、第2章では研究を行う準備として関連研究について説明する。また、全体像として第3章では Stack Overflow におけるデータの収集および分析について述べる。第4章で分析により明らかになったことをまとめ、分析で得られる特徴について考察する。第5章では結論を述べ、本稿をまとめる。

第2章 関連研究

2.1 QA サイトの分析に関する先行研究

本研究の分析に関する先行研究として、片山ら [2] は、回答者が実際に回答している質問にどのような特徴があるかを回答者の履歴に着目し分析を行っている。分析手法として、質問の履歴と質問の文書をベクトル化し類似度計算を行っている。また、使用する質問には実際に回答している質問とそれ以外の質問に分け比較実験を行っている。その結果、実際に回答している質問のほうが回答していない質問よりも類似度が高くなる傾向から、過去に回答した質問(履歴にある質問)と回答する質問では、回答していない質問よりも語の重なりが多いことを明らかにしている。ユーザーの特徴を履歴データから抽出できる可能性を示している。

甲谷ら [3] は、教えて!Goo において、三つのコミュニケーションタイプを定義し教えて!Goo のデータのログを用いて、各カテゴリの分類を行っている。また、コミュニケーションは時間とともに変化することがわかっており、QA ネットワークが時間とともにどう成長するのか分析を行っている。評価実験として、各カテゴリを「知識交換」「相談」「議論」という三種類のコミュニケーションタイプに分類した上で、先行研究でわかっている Yahoo Answers における同様の分析結果との比較を行っている。その結果、一回答あたりの平均語彙数に関してふるまいが異なることを明らかにしている。

栗山ら [4] は、投稿者の質問・回答(「OKWave」の「恋愛・人生相談」カテゴリに投稿された質問・回答)の特徴について比較を行っている。比較では、タイトルと本文に分け、質問中に出現する語の抽出と特徴語の抽出を行い、TF-IDF で重み付けを行っている。扱うデータは読売新聞の「人生案内」の質問と OKWave の「恋愛・人生相談」の質問であり、それぞれのデータについて形態素解析を行っている。また、質問者がどのような悩み・質問を抱いているかを、インターネット以外の媒体を参照することで、より適切な回答が得られる可能性があるのかについて考察している。考察として、共通点と相違点をまとめている。共通点として、人間関係に関する質問、特に自分のパートナーに関する質問が多く、質問文にはパートナーや身近な人間を表す語や評価や感情を表す語がよく用いられることを明らかにしている。また相違点として、新聞の相談者は女性の割合が高く年齢層があまり偏っていないのに対し、QA サイトの質問者はほぼ男女半々で 10~20 代の若年層に偏っており、回答者はそれよりも少し上の 20~30 代が多いことがわかっている。これら

の共通点と相違点から相補的に利用することで有用な回答を広く探すことが可能であることを明らかにしている。

Liら [5] は, Research Gate(RG) という QA サイトを使用し, 質問の質の特徴付けを行っている。手法としては, SVM 分類アルゴリズム, 三予測モデルの対照分析, 特長の対照分析の分類モデルを提案している。各提案手法に対して, 回答品質のための最適化を行った結果, 挨拶や他の感情の言葉は実際の回答品質にネガティブな影響があると示している。

2.2 QA サイトにおける推薦に関する先行研究

本研究の背景や課題に関する先行研究として, 甲谷ら [1] は, これまでの QA サイトの問題点として回答者は回答したい質問を自分で見つけてくる必要があるという点を解決するために, 回答者に適切な質問を推薦する手法を提案している。QA サイトでは教えて!goo を扱っており, ユーザをノード, 回答をエッジとして, QA サイト上のコミュニケーションをグラフにモデル化したもの QA ネットワークに着目している。特に, 3 ノード部分グラフに着目している。推薦手法として, ある時点 t_0 の QA ネットワークに含まれる 3 ノードからなる構造を全て抽出し, 回答までの時間を指標に, 次の回答が発生しやすいようなユーザを発見している。提案手法の有効性の評価について, カイ二乗検定を行った結果, 上位にランクインされたユーザ間には, 実際に回答が発生する傾向があることを明らかにしている。また, 提案手法とランダム選択した場合を比較し, 回答者に適切な質問を推薦するのに有効である可能性を示している。

石川ら [6] は, QA サイトのベストアンサーを計算機が推定可能かの検証を行っている。推定実験の質問 (2~4 件の回答付きに限定し, 全 50 件) を作成し, 推定結果が人間の推定結果にどの程度接近かの調査を行っている。結果として, 「パソコン」においては判定者らの結果を上回っており, 「一般教養」「政治」はほぼ同等であり, 「恋愛相談」においては判定者らの結果を下回ったという結果を示している。

見市 [7] は, QA サイト内において質問者の質問意図と検索利用者の検索意図に差が生じるため, 求める回答にも差が生じ, 結果的に検索利用者の意図した回答が得られないという問題点を解消するために, 検索利用者の検索意図は複数存在することに着目し, 検索利用者の意図を考慮した回答検索システムを提案している。使用するデータは Yahoo!知恵袋を使用している。検索利用者における複数存在する検索意図を S 項目と定義しており, 二つの提案について比較を行っている。S 項目を用いた回答検索手法における, 一つ目の提案はベクトル間類似度を用いた回答検索手法である。二つ目の提案は回答抽出による回答検索手法である。Yahoo!知恵袋の検索結果と提案 1,2 の比較では, 一つ目の提案は全ての検索クエリにおいて検索精度の上昇を示している。二つ目の提案は三つの検索クエリにおいて検索精度の上昇を示している。このため, S 項目を考慮することでより回答検索の

際に検索利用者の検索意図を反映することが可能であることを示している。また、提案システムによる検索利用者の意図を反映した回答の推薦が可能であると示している。

第3章 質問形式の分析手法

前章では，QA サイトに関する先行研究について述べた．そこで，関連研究を踏まえ研究目的を達成するための提案を行う．本章では，研究の準備段階について述べる．Stack Overflow における未回答質問の件数を調査について述べる．また，収集するデータと未回答質問の分析手法について述べる．

3.1 使用するデータ

本論文では Stack Overflow が提供するアーカイブデータの分析手法を提案する．使用するアーカイブデータは 2015 年 8 月 18 日に更新されたデータ¹である．Stack Overflow では Apple や Codereview, Unix, Tex など様々な専門性に特化したサイトを複数運営している．本論文では Stack Overflow を構成する個々のサイトをカテゴリと称する．アーカイブデータはカテゴリ分けされている．表 3.1 に，質問におけるそれぞれのカテゴリが持つ情報について，表 3.2 に，回答におけるそれぞれのカテゴリが持つ情報について示す．また，使用するデータではアーカイブデータの apple カテゴリとする．アーカイブデータ以外のデータとして Stack Overflow の実際のサイトからコメント文を取得したので，表 3.1 の下部に追加した．追加した理由として，コメント文は質問の補足情報や意見を受け付けるために作られた機能なのでコメント文を分析することで，未回答質問においてどのような情報が不足しているのかが明らかになると考えたことによる．表 3.1, 3.2 の中にある「*」のマークが付けられた項目はタグ名が同じになっており，また共通する項目を意味する．

本論文は，表 3.1 の投稿 Id, 投稿タイプ, 認証済みとなる Id, 投稿日付, 質問文と表 3.2 の投稿 Id, 投稿タイプ, 回答の親 Id, 投稿日付を扱う．また，質問分析として扱う質問は，投稿タイプが 1 であるものかつ回答の親 Id が質問における投稿 Id に存在しないものとする．さらに，削除された質問に関しては扱わないこととする．認証済みされた質問では，表 3.1 に示す投稿タイプが 1 であるものかつ認証済み回答 id が付いたものとする．また，未回答質問と認証済みされた質問との比較を行う．

¹<https://archive.org/download/stackexchange>

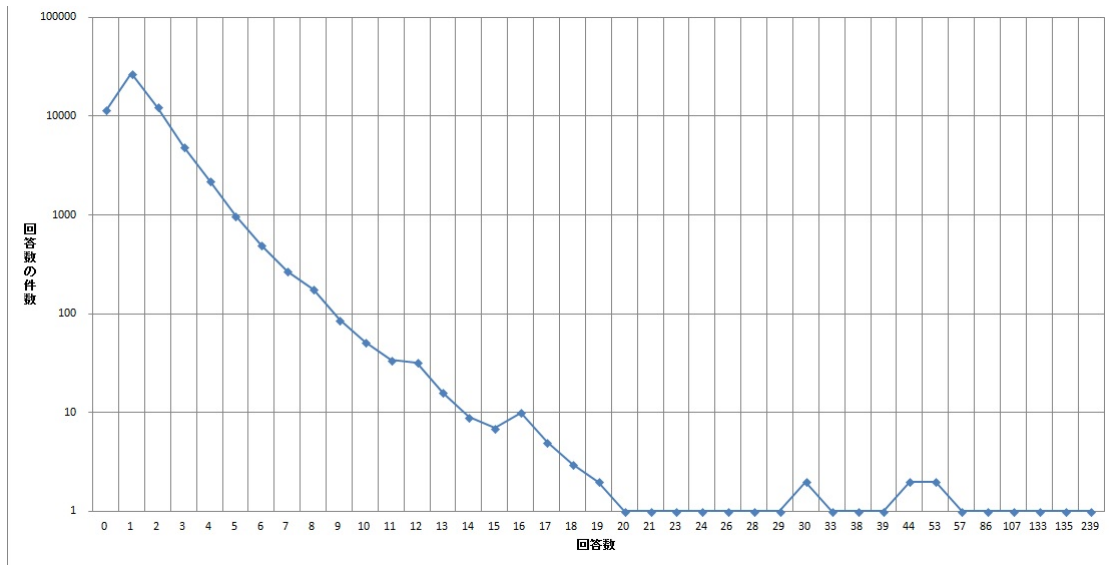


図 3.1: apple カテゴリにおける回答のデータ分布

3.2 回答件数に関する調査

前節より、使用するデータを決めたので、研究を始めるにあたり、Stack Overflowにおいて未回答の質問がどのくらいの件数なのか調査する。調査を行う理由としては、未回答の質問の件数を調査することで、研究の有用性がより明確となることによる。apple カテゴリにおける全質問数、未回答質問数、ベストアンサー付きの質問数を調査することで、全体の質問数に対して研究対象となるデータ数がどのくらい占めているのかわかる。表 3.3 に、apple カテゴリの質問や回答の件数を示す。

未回答の質問が全体のどのくらい占めているのか調査した。この調査を踏まえた上で、質問内容が不明確な質問にはどのような特徴が見られるのかを調査する。

表 3.3 から、未回答の質問数は全体の質問の 19.2% であることがわかる。また、ベストアンサー付き質問数は全体の質問の 39.0% である。ベストアンサー付き質問数の割合については、Stack Overflow からベストアンサー付き質問・回答を質問仕方事例集として提示することができる可能性を示唆している。未回答の質問数の割合については、質問全体の約 20% が回答が付けられていないので、少なくとも 20% 中に質問内容が不明確な質問も含まれている可能性がある。

apple カテゴリにおける回答のデータ分布（回答が 0 件も含む）について図 3.1 に示す。図 3.1 では回答が 1 件付与されている質問が最も多く、次いで 0 件と 2 件付与されている質問が多い。2 件以降は回答数が増えるたびに減少していく傾向にある。図 3.1 示した 0 件が未回答質問となる。より良い質問・回答を閲覧するサイトにしていくためには、未回答質問として残されている原因を明らかにする必要がある。

未回答質問と未回答質問に付与されたコメントの分析を行う。ここで未回答質問とは、回答者となりうるユーザから回答が付与されていない状態の質問である。また、コメントや+の評価投票やお気に入りが付与されている場合でも回答が付与されていなければ未回答質問として扱う。未回答質問は Stack Overflow のアーカイブデータから抽出を行うので、閲覧数が0でない質問を対象としている（投稿されたばかりの新規質問は未回答質問として扱わない）。

3.3 未回答質問の特徴抽出

未回答質問の特徴付けを行った。apple カテゴリのデータからランダムに 100 件を抽出する。ランダム抽出では、投稿日時が古いものと新しいものからそれぞれ 50 件ずつ抽出する。未回答質問において着目した点を以下に箇条書きで列挙する。

- 質問文 (Body)
- タグ (Tags)
- コメント文 (comment-text)
- 最終編集者 Id (LastEditorUserId)
- 投票スコア (Score)
- お気に入り数 (FavoriteCount)

未回答質問となる原因を明らかにするために特徴付けを行う。また、箇条書きで示した点に注目し 100 件の未回答質問に特徴付けを行う。未回答質問の特徴として、質問文に環境、症状、状況についての内容が含まれるという仮説を立てる。また、環境、症状、状況の全ての内容が少しずつ含まれている質問より環境、症状、状況のうちどれか一つだけが詳しく書かれている内容の方が未回答質問になるという仮説を立てる。環境、症状、状況について、どのようなことが述べられているとそれぞれに判断されるのかの基準を以下に箇条書きで示す。

- 環境について

開発環境や OS やソフト名が記載されていないと回答が困難であると考え。また、記載されていない場合プログラマー向け QA サイトを使用する際の周辺知識が必要であると考え。

- 症状について

エラーメッセージや問題が起こる前に何をしたかやどのようなコードを書いているのかが記載されていないと回答が困難であると考え。また、コードの変数名や使用されるデータをそのまま使うことについて、Stack Overflow ではサンプルを用意するように定めている。プログラミングの構造についての理解が必要であると考え。

- 状況について

問題解決のために何を調べたのかやどこまで問題を把握しているのかや解決してほしいポイントが明確であるかが記載されていないと回答が困難であると考え。また、関連する質問を調べるための方法やどんな質問が関連しているかプログラマー向け QA サイト以外で情報収集する能力が必要であると考え。

本章では、使用するデータについて述べた。また、そのデータの予備調査の結果を示した。予備調査を行うことで、研究の有用性を明確にし、研究目的を達成するための未回答質問における分析手法について述べた。次章では、分析手法の結果および考察について述べる。

表 3.1: 質問における収集データ

ラベル	タグ名	説明
*投稿 Id	Id	投稿順に付与される Id を表す
*投稿タイプ	PostTypeId	質問か回答かを判断するタイプを表す. タイプが1の場合は質問を表し, タイプが2の場合は回答を表す
認証済み回答 Id	AcceptedAnswerId	認証済みとされた回答の Id が付けられる. また投稿タイプが1のものに付けられる
*投稿日時	CreationDate	投稿された年月日および時間を表す
*投票スコア	Score	それぞれのユーザ自身の判断で+か-の票を入れることができ, 数値が高いほど上位に, 低いほど下位となる
*閲覧数	ViewCount	質問にアクセスされた回数 (閲覧数) を表す
質問文	Body	質問の本文を表す
質問者 Id	OwnerUserId	質問者に付与される Id を表す
最終編集者 Id	LastEditorUserId	最後に質問を編集したユーザの Id を表す
最終編集の日時	LastEditDate	最後に質問を編集した年月日および時間を表す
*最終動作日時	LastActivityDate	最後に動作があった年月日および時間を表す
質問タイトル	Title	質問のタイトルを表す
タグ	Tags	質問に関連する単語を最大五つ付けられる. また関連する単語については Stack Overflow が定めている単語となる
回答数	AnswerCount	回答の数を表す
*コメント数	CommentCount	コメントの数を表す
お気に入り数	FavoriteCount	お気に入りの数を表す
コメント文	comment-text(実サイトから取得)	質問者とその他のユーザによる質問に対してのコメントの文

表 3.2: 回答における収集データ

ラベル	タグ名	説明
*投稿 Id	Id	投稿順に付与される Id を表す
*投稿タイプ	PostTypeId	質問か回答かを判断するタイプを表す. タイプが1の場合は質問を表し, タイプが2の場合は回答を表す
回答の親 Id	ParentId	どの質問に対する回答かを表す Id が付けられる. また投稿タイプが2のものに付けられる
*投稿日時	CreationDate	投稿された年月日および時間を表す
*投票スコア	Score	それぞれのユーザ自身の判断で+か-の票を入れることができ, 数値が高いほど上位に, 低いほど下位となる
*閲覧数	ViewCount	質問にアクセスされた回数 (閲覧数) を表す
回答文	Body	回答の本文を表す
回答者 Id	OwnerUserId	回答者に付与される Id を表す
*最終動作日時	LastActivityDate	最後に動作があった年月日および時間を表す
コメント数	CommentCount	コメントの数を表す

表 3.3: 2015 年 8 月 18 日における Stack Overflow の質問件数

データ種別	データ数
質問総数	60,327
未回答の質問数	11,553
ベストアンサー付きの質問数	23,549

第4章 未回答質問となる原因の検討

本章では、未回答質問に関して、質問の特徴付けを行う。また、未回答なる原因を明らかにする。未回答質問における特徴付け、原因について考察を行う。

4.1 未回答質問の特徴付け

Stack Overflow が提供するアーカイブデータにおける、apple カテゴリの未回答質問 100 件に対して特徴付けを行った。結果を表 4.1, 4.2, 4.3 に示す。具体的にどのような特徴に着目したか以下に箇条書きで示す。列挙する項目として、未回答質問の本文、その質問に付与されたコメント、その質問に対する評価 (投票スコア, お気に入り数, 最終編集者 Id, タグ) の三点を挙げる。

- 質問文

文と文をつなぐときの単語が本文中に出現の有無およびその単語頻度に着目した。質問文の内容を読み、質問に付与されているタグと一致する単語はどのくらいあるのかに着目した。また、? マークや空行の数や他者からの編集がされたのかやタグについて着目した。

- コメント文

何人のユーザからコメントがされているかの件数に着目した。コメント内容に書き加える指示がされているのかに着目した。また、別の方法を試しているかどうかの確認がされているのかに着目した。最終投稿されているコメントに関して、そのコメントが質問者で? マークで終わっているか着目した。最終投稿されているコメントに着目した理由としては、質問に加えてコメントでも問題が発生し、その問題も解決されていない可能性があることによる。

- 未回答質問に対する評価 (投票スコア, お気に入り数, 最終編集者 Id, タグ)

- 投票数投票数とは、ある一定の評価を持ったユーザが行うことができる機能であり、+ の投票をされた質問は上位になる。上位になれば質問が注目されやすくなるので着目した。

- お気に入りお気に入りとは，Stack Overflow に登録している全てのユーザが行える機能であり，質問をお気に入りにすると自分のお気に入りにその質問が追加される．お気に入りに追加される質問はお気に入りに追加したユーザはいつでも閲覧することができるので着目した．
- 編集の有無編集の有無について，Stack Overflow では投稿した質問や回答やコメントをすべての利用者が編集を行うことができる．編集の有無に関しては，未回答質問にはどのくらい質問者以外のユーザから編集されるか着目した．
- タグタグとは，Stack Overflow で決められている単語群のことであり，質問を投稿する際質問がどういった分野なのかを表すために，質問に関連する単語を単語群から 1～5 個選出し，質問にタグ付けする．タグ付けに関しては，タグとして使った単語がどのくらい質問文に含まれているかに着目した．質問文にタグとして選出した単語が含まれているかに着目した理由は，質問を表す単語がタグの単語なので，タグ付けした単語が質問に含まれていないと質問内容がタグに関連していない可能性があることによる．タグに関連していないと回答者となりうるユーザが回答できない場合がある．

4.2 未回答質問の特徴付けを行った結果

表 4.1, 4.2, 4.3 に未回答質問の特徴付けを行った結果を示す．

表 4.1: 2015 年 8 月 18 日における Stack Overflow の未回答質問文の分析 (1)

	OS	AP	URL	IM	NS	NW	NL	NB	COMMENT	NQ	VT	FQ	WORD	T1	T2	T3	T4	T5	CA	AE
1	10	0	01	0	5	124	8	4	02003	0	0	0	1001000	0	0	0				0
2	00	0	00	0	4	47	1	0	10002	1	1	0	0000000	1	1					1
3	00	0	00	0	4	72	1	0	12003	1	0	0	0000000	2	2	2	2	2		1
4	11	0	00	0	6	73	5	2	10002	2	2	0	0100000	1	2	2	2			1
5	10	0	00	0	15	234	13	6	01003	2	1	0	0011002	2	2	2	2	4		0
6	00	0	00	0	6	100	1	0	00000	2	1	0	1000000	2						0
7	--	--	--						-----				-----						1	1
8	00	0	00	0	1	12	1	0	00000	1	1	0	0000000	0	1					1
9	00	0	00	0	7	67	5	2	12002	1	0	0	1000000	1	2					0
10	00	0	00	0	4	62	1	0	22003	0	0	0	1000000	0	1	0				0
11	00	1	00	0	4	71	8	5	11102	1	1	1	0000000	2	2					0
12	00	0	00	0	5	106	4	3	00000	1	2	0	1102000	1	0					0
13	10	0	01	0	5	56	6	3	02003	0	1	2	3000000	3	0	0	0			0
14	10	0	01	0	5	87	2	0	22012	1	2	0	0100000	1	2	2				0
15	--	--	--						-----				-----						1	0
16	--	--	--						-----				-----							
17	11	1	01	1	7	84	15	5	00000	1	0	0	0000020	1	5	0	0			0
18	00	0	00	0	8	148	5	2	11102	0	1	0	3001000	3	8	0				1
19	11	1	01	0	2	10	5	2	02103	0	1	0	0000000	0						0
20	00	1	00	0	16	387	14	6	01003	4	1	0	3125200	1	2	10	0			0
21	--	--	--						-----				-----							
22	00	0	00	0	2	33	3	1	11002	0	0	0	0000000	0	0					1
23	--	--	--						-----				-----						2.BA	0
24	--	--	--						-----				-----							
25	11	1	11	1	8	151	13	5	00000	3	1	0	0000011	1	0	2				1
26	--	--	--						-----				-----							
27	11	0	00	0	5	105	6	2	21012	0	0	0	0000200	0	0	4	2			0
28	--	--	--						-----				-----							
29	00	0	00	0	4	51	1	0	02003	1	0	0	1000000	0						0
30	--	--	--						-----				-----							
31	11	0	00	0	5	92	5	2	00000	1	2	0	1000000	1	1					
32	--	--	--						-----				-----							
33	--	--	--						-----				-----							
34	--	--	--						-----				-----							
35	00	1	00	0	4	66	9	4	00000	0	1	0	0100000	0	1	4				1

表 4.2: 2015 年 8 月 18 日における Stack Overflow の未回答質問文の分析 (2)

	OS	AP	URL	IM	NS	NW	NL	NB	COMMENT	NQ	VT	FQ	WORD	T1	T2	T3	T4	T5	CA	AE
36	00	1	00	0	4	31	1	0	23000	0	2	0	0000000	0	0					1
37	10	0	00	0	3	83	1	0	23003	1	0	0	1001000	0						0
38	00	1	00	0	6	83	1	0	11000	1	0	0	1001000	0	0	0				0
39	11	0	00	0	7	129	5	2	00000	1	1	0	3001020	3	5	1	6	3		0
40	--	--	--	--	--	--	--	--	-----	--	--	--	-----	--	--	--	--	--	3	1
41	01	0	00	0	2	31	1	0	00000	0	1	0	0000000	0	1					1
42	00	0	00	0	3	69	1	0	00000	0	1	0	1000000	1						0
43	00	0	00	0	5	64	3	1	11101	1	0	0	2001000	3	1	1				0
44	--	--	--	--	--	--	--	--	-----	--	--	--	-----	--	--	--	--	--	--	--
45	--	--	--	--	--	--	--	--	-----	--	--	--	-----	--	--	--	--	--	--	--
46	--	--	--	--	--	--	--	--	-----	--	--	--	-----	--	--	--	--	--	--	--
47	10	0	00	0	18	242	14	6	12012	1	0	0	2002000	4	5	0				0
48	--	--	--	--	--	--	--	--	-----	--	--	--	-----	--	--	--	--	--	--	--
49	--	--	--	--	--	--	--	--	-----	--	--	--	-----	--	--	--	--	--	--	--
50	11	0	00	0	8	131	7	3	13003	1	0	0	2000000	1	2	1	1			0
51	11	1	01	1	29	400	27	10	22013	6	3	0	6002352	1	5	7				0
52	00	0	01	0	7	145	9	4	22012	2	1	0	0001000	3	0	3				0
53	00	0	00	0	3	60	2	0	00000	1	0	0	0001000	0	0			0		0
54	01	1	00	1	9	105	12	6	32111	1	0	0	2001000	0	28	0	2			0
55	11	0	00	0	11	156	5	2	12113	1	0	0	4001020	3	0					0
56	00	0	00	0	4	58	2	0	01103	0	0	0	0000001	1						0
57	11	0	00	0	5	81	1	0	00000	0	0	0	0000000	1	0	2	2	2		1
58	00	1	01	0	3	70	5	2	00000	0	0	0	0000000	0	0	0	0			0
59	01	1	10	0	16	244	21	10	00000	1	0	0	3001100	0	10	0	2			0
60	00	1	00	0	10	78	9	4	12102	2	0	0	0000000	1	0					0
61	11	0	00	0	15	285	21	10	00000	0	1	0	5011100	1	1	0	0			0
62	--	--	--	--	--	--	--	--	-----	--	--	--	-----	--	--	--	--	--	1	0
63	10	0	00	0	2	68	3	1	01103	1	0	0	0002100	3	0	1				1
64	10	1	00	0	6	117	6	3	00000	0	0	0	2002000	1	1	5	1	2		0
65	10	0	00	0	2	37	1	0	00000	0	0	0	0100000	0	2	0				0
66	10	0	00	0	4	40	1	0	00000	2	0	0	1000000	0						0
67	00	0	00	0	5	52	4	1	14102	1	1	0	2100000	0	1					0
68	--	--	--	--	--	--	--	--	-----	--	--	--	-----	--	--	--	--	--	5-BA.0	0
69	10	0	00	1	12	126	16	4	00000	1	1	1	0001000	1	0	1	0	1		0
70	--	--	--	--	--	--	--	--	-----	--	--	--	-----	--	--	--	--	--	2	0

表 4.3: 2015 年 8 月 18 日における Stack Overflow の未回答質問文の分析 (3)

OS	AP	URL	IM	NS	NW	NL	NB	COMMENT	NQ	VT	FQ	WORD	T1	T2	T3	T4	T5	CA	AE	
71	10	0	00	1	5	131	10	5	01013	1	3	1	0002000	8	3	2	2	2	0	0
72	01	1	00	1	10	309	27	13	23103	1	0	0	2003000	2	0	4	0			0
73	11	0	00	0	8	228	7	3	00000	1	1	0	2115000	1	5					0
74	11	0	00	0	8	57	10	4	00000	1	1	0	0000000	1	0	1				0
75	00	0	00	1	4	69	4	0	44113	1	0	0	0000000	1	0					0
76	10	0	10	0	5	132	5	2	11002	1	0	0	1001000	3	1	1	2			1
77	11	0	00	0	2	39	3	1	11002	2	0	0	0000000	1				0		1
78	00	0	00	1	7	113	11	4	00000	1	1	1	0000000	0	0					0
79	00	0	11	0	3	109	3	1	11002	1	1	1	2000000	0	0	0				1
80	00	0	00	0	4	73	6	2	00000	1	1	1	0000000	0						0
81	00	0	00	0	10	110	10	3	00000	0	2	0	0000000	2	3	2	1			0
82	00	0	01	0	4	44	3	1	00000	1	1	0	1000000	0	1	1	1			1
83	00	0	00	0	3	78	5	2	00000	1	1	0	0010000	1	2	6				1
84	00	0	00	0	6	101	5	2	00000	1	2	0	0010000	2	3	1	3			0
85	00	0	00	0	5	92	1	0	00000	0	1	0	1001000	0	0	1				0
86	10	0	00	0	5	71	1	0	00000	1	2	0	0100000	1	1					0
87	00	1	01	0	15	274	17	8	00000	5	2	0	1001000	1	3	0				0
88	--	--	--	--	--	--	--	--	-----				-----					4_BA.0		1
89	00	0	00	0	3	88	1	0	01003	1	1	0	0000000	0						0
90	00	0	11	0	3	47	4	1	11002	1	0	0	1000000	0	2					0
91	00	0	00	0	2	36	1	0	10002	1	1	0	0000000	0	0	1				0
92	10	0	00	0	6	83	5	2	11012	2	0	0	0000000	0	0	0				0
93	00	0	00	0	3	46	1	0	21113	1	0	0	0000000	0	1	0				0
94	00	0	02	1	13	139	13	5	22112	2	1	0	0000000	2	0	3	4	3	0	0
95	00	0	00	0	7	64	1	0	00000	2	1	0	0001000	0	0					0
96	00	0	00	0	7	45	6	2	01003	2	1	0	0000000	0	1	0	1			0
97	00	0	00	0	7	161	5	2	12102	2	0	0	1000100	2						0
98	11	0	01	0	6	89	12	6	00000	1	1	0	0000000	0	0	2				0
99	10	0	00	0	2	43	1	0	12103	2	1	0	0000000	0	0			0		1
100	00	0	00	0	7	177	11	5	13013	2	0	0	2110101	1	4	0				0

表 4.1, 4.2, 4.3 における各項目についての説明を以下に、箇条書きでまとめる。

- OS

質問文中に OS 名が記載されているかとバージョンが記載されているかを表す。1 が記載ありを表し、0 が記載なしを表す。具体的には、表内に「1 0」となっている場合、OS 名の記載はあり、バージョンの記載はないことを意味する。

- AP

質問文中におけるコードの有無を表す。コードのが添付されている場合 1, 添付されていない場合 0 とする。またコードの添付には、実行コマンドや log や画像を含む。

- URL

質問文中に関連する質問の URL が記載されているかと関連 URL が記載されているかを表す。1 が記載ありを表し、0 が記載なしを表す。具体的には、表内に「0 1」となっている場合、関連する質問 URL の記載はなく、関連する URL の記載はあることを意味する。

- IM

箇条書きの有無を表す。1 が箇条書きがありを表し、0 が箇条書きなしを表す。

- NS

質問文がいくつの文で構成されているかを表す。また、表内の数字が文の数を意味する。

- NW

質問文がいくつの単語で構成されているかを表す。また、表内の数字が単語の数を意味する。

- NL

質問文がいくつの改行で構成されているかを表す。また、表内の数字が改行の数を意味する。

- NB

質問文がいくつの空行で構成されているかを表す。また、表内の数字が空行の数を意味する。

- COMMENT

質問文に対するコメントに関する項目である。コメントに関して、質問者のコメント数、他のユーザのコメント数、他のユーザのコメントによる追記の指示の有無、他のユーザのコメントで別の方法を試したかの確認の有無、最後のコメントが質問者でかつ？またはピリオドで終わるまたは他のユーザで終わるかを表す。具体的には、表内に「02003」となっている場合、質問者によるコメント数が0であり、他のユーザのコメント数が2であり、他のユーザのコメントによる追記の指示がなく、他のユーザのコメントで別の方法を試したかの確認がなく、最後のコメントが他のユーザで終わっていることを意味する。また、最後のコメントが質問者でかつ？で終わっている場合は1であり、質問者でかつピリオドで終わっている場合は2である。

- NQ

質問文がいくつの？マークが出現したかを表す。また、表内の数字が？マークの数を意味する。

- VT

質問に対する投票数を表す。また、表内の数字が投票された結果の数値であり、プラスは省略しており、マイナスには「-」を付ける。

- FQ

質問に対するお気に入り数を表す。また、表内の数字が他のユーザに登録されたお気に入りの数を意味する。

- WORD

質問文ある単語に関する項目である。質問において、どういう意味を表す文なのかを判断する単語として、But, However, Though, So, Because, Error, Bug がどのくらい出現したかを表す。また、表内の数字がそれぞれの単語の数を意味する。具体的には、表内に「1001000」となっている場合、But の数が1であり、However の数が0であり、Though の数が0であり、So の数が1であり、Because の数が0であり、Error の数が0であり、Bug の数が0であることを意味する。

- T1~T5

T1~T5 は質問文に対するタグに関する項目である。質問者が付与したタグが質問文にどのくらい出現したかを表す。また、表内の数字が質問文に出現したタグの数を意味する。T1~T5 間における空のセルについては、タグが付与されていないことを意味する。

- CA

アーカイブデータでは未回答質問として抽出したが、時間の経過によりその質問が何らかの変化があったかに関する項目である。実際に Stack Overflow で質問を確認した際に、その質問がクローズされているまたは回答が付与されていたかを表す。表内における空のセルは変化がなく未回答質問であることを意味し、0 以外の数字は回答の数を意味しており、0 はクローズを意味する。また、表内の「BA」は承認済みされたことを意味する。

- AE

質問文に対して、他のユーザが質問を編集したかを表す。1 が他のユーザに編集されたことを表し、0 が編集されていないことを表す。

4.3 認証済み質問の特徴付けを行った結果

表 4.4, 4.5, 4.6 に認証済み質問の特徴付けを行った結果を示す。

表 4.4: 2015 年 8 月 18 日における Stack Overflow の認証済み質問文の分析 (1)

	OS	AP	URL	IM	NS	NW	NL	NB	COMMENT	NQ	VT	FQ	WORD	T1	T2	T3	T4	T5	CA	AE
1	00	0	00	0	2	28	3	1	01003	0	8	3	0000000	0	0				3	1
2	11	0	00	0	3	45	1	0	00000	3	19	5	0000000	1	1				3	1
3	10	1	01	1	11	149	13	6	12002	4	12	1	1001000	1	0	5			6	1
4	11	0	00	0	2	32	2	0	01003	1	7	3	0000000	1	1	0			4	0
5	00	0	00	0	7	76	7	3	02003	2	10	3	3001000	1	3				4	0
6	10	0	00	0	3	120	3	1	00000	2	8	4	0000000	1	0	1			4	0
7	00	0	00	0	2	36	1	0	02003	1	32	5	0000000	0	0	0			5	1
8	11	0	00	0	5	60	3	1	00000	1	1	0	0002000	1	1	2			1	0
9	11	0	00	0	15	105	28	11	12103	3	0	0	0100000	4					2	0
10	10	0	00	0	4	101	5	2	11002	1	3	1	0010000	1	0	4			2	0
11	10	0	00	0	6	99	5	2	00000	1	0	0	0000000	2	3	2			2	0
12	00	0	00	0	3	72	3	1	01003	1	0	0	0000000	0	3	2	0		1	1
13	10	0	00	0	4	74	7	3	01013	0	1	0	1000000	2	0				1	1
14	00	1	00	0	5	103	11	2	32001	3	3	1	0002000	6					1	0
15	11	0	00	0	2	4	3	1	01003	1	1	0	0000000	1	0	0	1		4	0
16	10	0	00	0	2	62	3	1	02103	1	4	0	1000000	2	1	0	0		4	1
17	11	1	00	0	11	225	7	3	12003	2	0	1	2004200	9	1				2	0
18	11	0	00	0	6	115	8	2	12003	1	1	1	1000100	2	4				1	1
19	11	0	00	0	2	51	3	1	10002	1	64	9	0000000	1	1				4	1
20	00	0	00	0	2	27	1	0	01003	2	2	1	0000000	0	0	1			1	0
21	00	1	10	0	5	85	10	5	00000	0	2	1	0102000	3	3	2			2	0
22	11	0	00	1	18	339	14	5	13003	2	4	1	1122100	4	0	0			4	0
23	11	0	00	0	6	52	5	2	00000	1	0	0	1000000	1	2				1	1
24	00	0	00	0	2	51	3	1	00000	1	1	0	0000000	0					1	0
25	10	0	00	0	3	73	5	2	12002	1	0	0	0010100	1	1	1			2	1
26	00	0	00	0	3	73	3	1	01000	0	3	0	1001000	2	1	0			2	0
27	10	1	00	0	7	96	10	3	10002	0	2	0	0000000	0	0	1	7		5	1
28	10	0	00	0	3	61	5	2	00000	1	6	1	0000000	1	1	6			3	0
29	11	0	00	0	5	89	5	2	00000	1	2	0	0001000	0	0				1	0
30	10	0	00	0	3	51	3	1	00000	1	2	1	0100000	1	1				3	0
31	10	0	00	0	15	307	19	9	00000	1	7	1	5106000	2	0	3			1	0
32	11	0	00	0	2	52	3	1	00000	2	2	0	1000000	0	1	2	2		1	0
33	11	1	00	0	3	49	4	1	12003	1	12	2	0000100	2	1	0			3	1
34	11	0	00	0	1	15	1	0	00000	1	5	0	0000000	1	1	1			1	0
35	10	0	00	0	7	101	7	3	01003	2	6	2	1001000	3	0				4	0

表 4.5: 2015 年 8 月 18 日における Stack Overflow の認証済み質問文の分析 (2)

	OS	AP	URL	IM	NS	NW	NL	NB	COMMENT	NQ	VT	FQ	WORD	T1	T2	T3	T4	T5	CA	AE
36	11	0	10	1	14	211	14	3	11002	1	2	0	4002000	8	0	1	0		3	1
37	00	0	00	0	2	34	1	0	00000	1	2	0	0000000	2	2	1			1	1
38	00	0	00	0	4	111	5	2	01003	1	3	0	2001000	1	0	0			3	1
39	11	0	00	0	9	154	10	4	01003	4	2	0	1001100	1	1	6	4		2	0
40	00	0	00	0	11	185	7	3	35002	2	11	2	0002200	0	3	2			2	1
41	00	0	00	0	3	49	5	2	00000	1	1	0	0000000	0	0	2			4	0
42	00	0	00	0	4	49	5	2	00000	2	0	1	0000000	0	0	0	0		2	1
43	10	0	00	0	8	58	7	3	00000	2	8	3	2000010	2					4	0
44	11	0	00	0	5	65	9	4	12003	1	3	0	0001000	0	1			2	0	
45	00	0	00	0	6	56	5	2	11002	1	1	0	2000000	1				1	0	
46	10	0	01	0	2	32	1	0	01103	0	6	0	0001000	0	0	0			5	1
47	10	1	00	0	4	80	5	2	00000	3	0	0	1000000	1	2			1	0	
48	11	0	00	0	4	74	5	2	00000	1	3	2	1000000	1	1			3	1	
49	00	0	00	0	1	20	1	0	11002	1	4	2	0000000	1	1	0	1		3	1
50	00	0	00	1	6	101	6	2	01003	1	2	2	0000000	0	0	0		1	0	
51	10	0	01	0	13	172	5	2	22002	1	1	0	2000000	2	7			1	0	
52	00	0	00	0	2	66	1	0	01103	1	3	2	0000000	1	1	1		0.2	1	
53	00	0	00	0	5	84	5	2	11002	1	1	0	1000000	0	0	6	0	4	1	1
54	00	0	00	0	6	83	5	2	11102	1	3	0	1000000	1	1	0		1	1	
55	10	1	00	0	3	59	3	1	00000	2	2	1	1000000	1	1			2	1	
56	00	1	00	0	21	323	55	16	44002	0	2	0	3000000	0	1			2	1	
57	00	0	00	0	10	177	9	3	00000	1	1	0	2101000	1	1	0	4		1	1
58	11	1	00	0	4	88	11	3	01003	2	0	0	0000000	1	0	0		1	0	
59	10	0	00	0	3	36	5	2	00000	1	2	0	0000000	1	1			1	0	
60	10	0	00	1	4	91	4	1	00000	2	0	0	1000000	4	2	2		1	1	
61	10	1	00	0	7	105	9	4	21001	2	1	0	0000000	1	1	4	2		1	1
62	00	0	00	0	3	34	1	0	00000	1	4	1	1000000	0				2	1	
63	10	1	00	0	8	113	8	3	22003	1	2	0	1000000	0	1			3	0	
64	00	1	00	0	2	65	11	2	01003	1	3	0	0000000	1				2	1	
65	11	0	00	0	3	82	5	2	02103	1	3	0	0000000	2	2	2		1	0	
66	10	0	01	0	1	36	1	0	11002	0	0	0	0000000	1	1	0	1		1	1
67	00	1	01	0	8	144	12	3	00000	0	2	0	1000000	0	0	4	0		1	1
68	10	0	01	0	6	102	5	2	13101	2	2	0	0001000	1				2	1	
69	10	1	00	0	4	62	10	3	23103	1	1	0	1000000	0	2			1	0	
70	00	1	00	0	4	62	7	3	10002	1	4	0	1000000	0				2	0	

表 4.6: 2015 年 8 月 18 日における Stack Overflow の認証済み質問文の分析 (3)

	OS	AP	URL	IM	NS	NW	NL	NB	COMMENT	NQ	VT	FQ	WORD	T1	T2	T3	T4	T5	CA	AE
71	00	1	00	0	4	145	21	6	33002	1	2	0	1000000	1	0				1	1
72	00	0	00	0	4	56	3	1	00000	1	1	0	0000000	2	2	0			2	1
73	11	1	00	0	4	127	21	5	00000	1	1	1	1001150	5	5				1	0
74	10	0	00	0	5	101	3	1	00000	0	1	0	1102000	1	2	1	2		1	0
75	10	1	00	0	6	119	13	4	00000	4	0	0	1000000	0	0	0			1	1
76	11	0	00	0	4	65	1	0	11001	1	0	0	1000000	1	3	3			1	1
77	10	1	00	0	8	265	19	8	22003	0	3	0	3003100	0	1	10	2		2	1
78	10	0	00	0	4	98	5	2	00000	1	1	0	0000000	1	2				1	1
79	10	0	00	0	3	46	1	0	00000	1	0	0	0000000	1	1	0			1	0
80	00	1	01	0	18	286	35	13	01003	5	3	1	0102000	0	2				1	0
81	11	1	10	0	11	213	15	7	00000	0	-1	0	0001100	3	1	1	0	3	1	1
82	10	0	20	0	10	108	9	4	11002	0	3	0	2102030	1	2	3			3	1
83	00	0	00	0	6	102	7	3	01103	1	2	1	1000000	4					1	1
84	00	1	00	0	15	133	17	7	00000	1	0	0	2000000	0	3	0	5	6	1	1
85	00	1	00	0	7	135	24	9	03003	1	2	1	2000000	0	0	0	1	4	1	0
86	11	1	00	1	23	380	94	16	11002	1	2	0	0001102	2	1	0	7	2	2	1
87	00	1	00	0	12	218	11	6	02013	1	1	0	2202100	8					1	0
88	11	0	00	0	5	86	1	0	23013	1	0	0	0000000	4					1	1
89	10	0	00	0	2	27	3	1	00000	1	1	0	0000000	0	2	0			1	0
90	00	0	00	0	3	29	3	0	00000	1	0	0	0000000	1	0				1	0
91	00	0	00	0	2	34	3	1	00000	1	2	1	0000000	2	2				1	0
92	11	0	00	0	2	25	1	0	11002	1	-1	0	0000000	1	2				0_1	1
93	00	0	00	0	4	45	3	1	02013	1	2	1	0000000	1	3				1	0
94	00	1	11	1	7	243	31	4	27103	1	3	0	0001000	0	4				1	0
95	11	1	00	0	8	96	9	4	01003	2	16	0	0000000	0	0	1	0	0	1	1
96	11	1	02	0	13	198	24	10	10002	0	0	0	0100100	1	3	3			1	1
97	10	1	00	0	4	1038	329	34	67003	0	1	1	0000000	1	1	1			3	1
98	00	1	00	0	4	84	6	2	01003	1	0	0	2001000	1	0	1	3		1	0
99	10	0	00	0	2	41	1	0	00000	1	1	0	1000000	0	2				2	0
100	10	0	01	0	4	75	6	2	00000	1	0	0	0100000	3	1	1			1	0

表 4.7: 未回答質問と認証済み質問に含まれる各特徴数

特徴	未回答質問	認証済み質問
OS 名	32	62
Ver. 名	20	28
コード等	16	31
関連質問 URL	5	6
関連 URL	16	11
箇条書き	10	7
書き加えコメント	16	10
別の方法の確認	13	4
他人による質問編集	21	50

4.4 未回答質問と認証済み質問の比較

表 4.7 は、未回答質問と認証済み質問の特徴付けした結果、各特徴がどのくらい含まれていたかを示す。

4.5 未回答質問の特徴についての考察

未回答質問文と未回答質問に付与されたコメントと未回答質問に対して付けられた評価の考察を行う。また、100 件の未回答質問をランダムに抽出したがその際、投稿された日時が古い 50 件と新しい 50 件を合わせて 100 件としているので、日時が古い 50 件を前半の 50 件とし、日時が新しい 50 件を後半の 50 件とする。

未回答質問に関しては、プログラムのコードが付与されていない、関連する質問の URL が記載されていない、質問文中に箇条書きや空行が含まれていないという特徴が見られた。

関連する URL については、関連する URL が記載されていない場合、質問者は質問に関連した情報が調べられていないと考えられる。または調べることができないと考えられる。

Stack Overflow おける質問で既に投稿された質問を関連 URL として記載したものと関連する web ページの URL が記載されたものの二つに分けられる。Stack Overflow おける質問で既に投稿された質問の関連 URL の前半の 50 件では 1 件であり、後半の 50 件で 4 件である。前半と後半で比較すると後半の方が多いが、100 件の未回答質問全体としては 0.05 % である。このことから、未回答質問には関連する質問の URL が付与されていない傾向があると考えられる。

関連する web ページの URL の前半の 50 件では 6 件であり、後半の 50 件で 9 件である。前半と後半で比較すると後半の方が多いが、100 件の未回答質問全体としては 0.15 % である。このことから、未回答質問には関連する質問の URL が付与されていない傾向がある

と考えられる。また、Stack Overflow おける質問で既に投稿された質問の関連 URL と関連する web ページの URL では関連する web ページの URL の方が記載されやすいとわかる。これは、Stack Overflow では質問と質問を関連付ける作業に対して、未回答となる質問者には行われていない傾向があると考えられる。

未回答質問に付与されたコメントに関しては、回答者になりうるユーザから開発環境はどのようになっているか、OS は何か、バージョンは何かを質問文に追記させる指示や質問解決に近づく作業は既に行っているかの確認がされていることがわかった。このことから、未回答質問には関連する情報が不足すると回答することが困難だと考えられる。また、表 4.1 に示す OS, AP, COMMENT, AE からわかるように動作環境や不具合の症状、質問するまでに調べた状況の情報が不可欠である。

アーカイブデータにおける未回答質問を Stack Overflow のウェブページでの確認を行った。アーカイブデータでは未回答の質問が実際に公開されている Stack Overflow のウェブページでは回答付きになっている質問が 100 件中 8 件あった。このことについて時間の経過に起因するものだと考えられる。前半と後半で分けて確認したとき、前半で 4 件、後半で 4 件であった。古い質問と新しい質問では時間による差はないと考えられる。よって、記述の修正が必要だと考えられる。また、アーカイブデータにおける未回答質問が Stack Overflow のウェブページでは見つからない質問もあった。前半では 15 件、後半では見られなかった。このことから、古い未回答質問は削除または質問 id が変更されたと考えられる。

接続語について、全体的に but や so が多く使われていることがわかった。また、because や though はあまり使われていないことがわかった。質問する上で but や so は使いやすい単語だと考えられる。

他人による質問編集では、未回答質問より認証済みの質問の方が多かった。認証済みの質問は必要な情報があるため、ユーザが積極的に参加したと考えられる。

本章では、100 件の未回答質問・認証済み質問に対して、質問の本文とその質問に付与されたコメントとその質問に対する評価の三点に着目し、考察を行った。次章では本論文の結論を述べる。

第5章 結論

本論文は、プログラマー向け QA サイトである Stack Overflow における未回答質問の特徴付けを行い、質問として成り立つには何が書かれていれば良いかを調査し検討した。また、未回答質問における記述の不明確さや回答が付けられない原因を明確にした。

未回答質問に対して、特徴付けを行ってわかったこととして、未回答質問における古い質問と新しい質問では時間による差はないと考えられる。よって、記述の修正が必要だと考えられる。また、アーカイブデータを Stack Overflow のウェブページで確認したときウェブページで閲覧できない質問の数は古い質問の方が新しい質問より多いということがわかった。このことから、古い未回答質問は削除または質問 id が変更されたと考えられる。

質問として成り立つには何が書かれていれば良いかを調査したことによって、未回答質問における不明確さは関連する情報が不足していることによるものだとわかった。具体的には、未回答質問には関連する情報が不足することにより、動作環境や不具合の症状、質問するまでに調べた状況の情報が不可欠であるということがわかった。また、これらの情報が回答を得るために有効であるかを検討した。

最後に、未回答質問における記述の不明確さや回答が付かない原因が明らかになったので、今後プログラマー向け QA サイトに質問を投稿するユーザや質問を投稿したが回答が付かないユーザに対して、どのように不明確なのかの確認や具体的な原因の提示を行うことができる。そこで、今後の研究として未回答質問における内容の不明確さや回答が付与されていない原因に基づき、質問解決支援を行う。質問解決支援とは、質問内容が不明確であっても Stack Overflow からベストアンサー付き質問・回答を質問仕方事例集として提示することである。

謝辞

本研究の一部は、JSPS 科研費 25280110 の助成を受けたものです。

佐藤哲司教授（筑波大学大学院図書館情報メディア研究科）は、研究の細部にわたり懇切丁寧な指導をしてくださいました。心より感謝いたします。佐藤哲司研究室のOB・OGの方々から支援物資により、とても快適な研究室で研究を行うことができ、卒業論文の執筆を無事終えることができました。OB・OGの方々に心より感謝いたします。常に適切なアドバイスを行ってくださり、共に助け合いながら頑張ってきた同期の上谷芽衣さん、川島崇秀さん、久保直人さん本当にありがとうございました。また、佐藤研究室の先輩である山本修平さん、伏見卓恭さん、玉田雄基さん、清野悠希さん、柘植大さんには、研究室内外でお世話になりました。佐藤研究室で学べたこと心より感謝しています。

筑波大学に編入学する以前研究活動として、近畿大学工業高等専門学校で研究や学術論文の執筆、学会発表等のご指導をしてくださった宇田隆幸先生に深く感謝しているとともに、現在の卒業研究を終えることができたことも学術面や精神面のご指導による支えがあったからこそ成り立ったことだと認識しています。また、健康な身体を維持するために相談およびご心配してくださった柔道部の倉本逸郎部長と辻本修総監督と三好紀孝監督に心より感謝しているとともにこれからの活動も健康を維持し取り組むことができると認識しています。人脈や研究や立地などにおいて素晴らしい環境である筑波大学に通わせてくれた両親に心より感謝いたします。

参考文献

- [1] 甲谷優, 川島晴美, 藤村考. QA コミュニティの成長パターンに基づく回答者への質問推薦. 日本データベース学会論文誌, Vol. 8, No. 1, pp. 89–94, jun 2009.
- [2] 片山亮, 川村秀憲, 鈴木恵二. QA サイトにおける質問推薦へ向けた履歴データの分析 (社会システムと情報技術研究ウィーク). 電子情報通信学会技術研究報告. AI, 人工知能と知識処理, Vol. 109, No. 439, pp. 11–16, feb 2010.
- [3] 甲谷優, 川島晴美, 藤村考. QA サイトにおける質問応答グラフの成長パターン分析. 日本データベース学会論文誌, Vol. 7, No. 3, pp. 61–66, dec 2008.
- [4] 栗山和子. 新聞と Q&A サイトにおける人生相談の分析. 電子情報通信学会技術研究報告. 研究報告デジタルドキュメント (DD) , Vol. 2015, No. 8, pp. 1–8, mar 2015.
- [5] Lei Li, Daqing He, Wei Jeng, Spencer Goodwin, and Chengzhi Zhang. Answer quality characteristics and prediction on an academic Q&A site: A case study on researchgate. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pp. 1453–1458, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.
- [6] 石川大介, 栗山和子, 酒井哲也, 関洋平, 神門典子. Q&A サイトにおけるベストアンサー推定の分析とその機械学習への応用. 情報知識学会誌, Vol. 20, No. 2, pp. 73–85, may 2010.
- [7] 見市高一, 川越恭二. QA サイトにおける s 項目による回答検索システム. Technical Report 25, 立命館大学大学院, 立命館大学情報理工学部, nov 2010.