

Stack Overflow における「知識の総和の増加」への志向と達成

田島逸郎 (慶應義塾大学大学院)

niryuu@keio.jp

I 背景と目的

個人メディアやソーシャル・メディアの隆盛は、人々がアクセス可能な知識を増大させるのみならず、あまりに多量の情報源が「情報爆発」の問題を引き起こしたり、ある主題に対する情報が重複しているなど、知識を求める人々にとって様々な問題を引き起こしてきた。このような Web における情報源の特性はしばしば適切に整備・組織化された知識体系と比較されている。

しかし、Web は知識の発信手段だけでなく、それを生み出す共同体にとってよりよくアクセスできるように、参加者が知識のストックを組織することを可能にするようなプラットフォームともなる。例えば Wikipedia は、「百科事典を作る」というプロジェクトの元に、あるトピックに関して知識を持った幅広い人々の参加を認めている。しかし、参加者は異なった思想信条や専門性を持ち、偏った方向性のある記事が書かれたり論争を引き起こしてしまう可能性がある。Wikipedia ではこれを避けるため「中立的な観点」といった文化を共有し、百科事典として知識を提供できるようにしている¹⁾。

ユーザーによる知識ストックの管理は、未だ一部のサービスや領域に限られているものの、今後重要になると考えられる。しかしその実際については、十分に研究がなされていない。本研究ではその端緒として、ユーザーによる知識ストックの管理の研究方針を実際のサイトと先行研究から検討する。

A 研究対象:Stack Overflow

本研究で扱うのは、プログラマーのための質問回答サイト Stack Overflow(以下 SO)²⁾である。SO はプログラミングに関する実際の問題を誰もが質問し回答できるサイトであるが、モデレーションや質問回答の編集など知識を管理する機能を備えており、ユーザーの参加に基づいて有用な知識を蓄積していくように共同体を形成する明確な意図のもとに設立、運営されている。実際に多くのプログラマーにとって有用な情報源となっている。以上のことから、SO は本研究の目的に適した対象であるといえる。

本研究では、まず SO の独特の知識や共同体に

関する考え方や方針を明らかにする。その上で、現在質問回答サイトがどのような視座のもとに研究されているかをまとめ、そこで行われている活動が既存の質問回答サイト研究の枠組では把握しきれないことを示す。最後に、SO のようなユーザーによる知識のストックの管理を行うサービスを研究する際の研究方針を検討する。

II SO の目指す知識、共同体観

本節ではまず、SO がどのような意図や目的を持って設立されたかを明らかにする。その上で、SO がその目的を達成するためにどのような方針をとって運営されてきたかを検討する。SO の設立や運営に関する資料は多数存在し、また SO の設立者や運営者へのインタビューを行った研究も存在する³⁾。後述する「meta.stackoverflow.com」において、SO の運営管理を含む様々な事柄がユーザー自身によって議論されている。その中で本研究では運営側の視点に目を向け、SO の創立者が設立の準備段階で個人ブログにおいて発信した SO に関する文書、及び開始後に SO の公式ブログで「コミュニティに対して」もしくは「コミュニティや知識に関する」ことを書いた文書を検討することで、知識やコミュニティに対する考え方を明らかにする。

A SO の出発点

米国では 2000 年代前半から、著名なプログラマーがブログを設立し、様々な技術的情報や考えなどを発信している。そのうちの 2 人である、Jeff Atwood と Joel Spolsky が SO の設立者である。Atwood は、2008 年 4 月 6 日に新しく立ち上げようとしている「ソフトウェア開発コミュニティのウェブサイト」を具現化するようなウェブサイトの名前をブログ上で募集した。その結果、「stackoverflow.com」が最も多くの得票を得た。これが SO の始まりである。

SO の目標が示されたのが 10 日後の 4 月 16 日である⁴⁾。Atwood はフォーラムやオンラインヘルプ、書籍などに情報源が「埋もれて」しまっていること、ブログが有用な情報源であるにも関わらず、それを始める人間が少ないことを鑑み、「簡単に参加でき、簡単に成果を探す

ことができる形でオンラインに成果を出すことができるような何かを作ろう」と提案した。それは「experts-exchange(検索エンジンに質問だけが出てくるが、回答を見るには課金が必要な質問回答サイト)ではなく」「Wikipedia や programming reddit のようなもの」だとしている。その上で、「プログラマーによるプログラマーのための、世界の中の良いプログラミングの知識の総和を集合的に増やす究極的な目的を持ったもの」であり、「良いプログラミングが私達のゴールである」だとサイトの目的を定義している⁴⁾。

9月16日にSOは設立された。その時のAtwoodのブログ記事⁵⁾で、4月16日以来様々な議論を経てきたことが紹介されている。例えばSOが世界の開発者をさらに無口にしてしまうのではないかという批判に対し、全てのプログラマーが完全な知識を持っているわけではなく、少しずつ知識を持ち寄ることができ、サービスがそうあることは開発者の参加を促進することができる旨の発言をしている。

以上のように、SOはプログラミングに関する実用的な知識を、ユーザーが少しずつ持ち寄って増やしていくものとして登場した。次節以降では、実際にそれをどう実現しようとしているかを明らかにする。

B SOのコミュニティ運営に見られる開発者の知識観

開設以降、SOは独特の方針を持った「質問回答サイト」として発展していく。しかし、設立以降もSOの創作者や運営者は共同体に対して質問や回答に関する方針や意見をブログなどで発信している。本節ではブログ記事(<https://stackoverflow.blog/>)や公式のドキュメントの中から中心的な主張を取り上げながら、SOにおいてどのような知識、及び知識のストックが望ましいとされてきたかについて提示する。

1 質問の基準:有用性と主観的/客観的な質問

SOのヘルプページにおいては、質問できる事柄について「特定のプログラミング上の問題」や「実践的で回答可能な、ソフトウェア開発にとって特有の問題」などを挙げられている。その上で、質問が実践的であることを維持するべく、「主観的な」質問を避けるよう述べている。例えば好みを求める質問、仮説的で答えのない質問などがこれにあたる。一方で、全ての「主観的な」質問が良くないわけではなく、理由や方法の説明を喚起するもの、事実や情報源に基

づく意見など「建設的で主観的な質問」は良いものとされる。

もう1つ重要な観点が、「少なくとも他の1人が関心を持つような質問」をすべきであるということである。プログラミングに関する問題にはある特定の状況でしか発生しないものが多い。しかし、SOでは特定の状況などに依存した質問は排除され、あくまで他の人がアクセスできるように質問を書くべきという方針である。

2 質問の「重複」と知識の多様性

従来から、プログラミングに関するディスカッション・フォーラムなどでは同じ技術的質問が何度もされることが問題となっていた。SOでもこれはWeb検索をしにくくするという観点から問題視され、技術的改良を経て質問の重複を一般ユーザーが報告し、モデレータが質問の無効化や、類似の質問へのリンクなどの処置を行えるようになった。

しかし一方で、質問を安易に「重複」として処理すべきでないという注意喚起も行われている。同じ問題を複数の仕方で提示することは、検索性を高める。また、どこまでが「重複した」質問なのかは曖昧であり、同じ機能に関する事柄でも似た別の問題が数多く存在する。例えば同じエラーが出たとしても、原因や対処の方法はその場その場で異なる場合が多い。SOの趣旨は、プログラマたちが自分の知識を少しずつ持ち寄ることにあり、似た質問回答によって解決できなかったことを質問することで、新たな知識が持ち寄られる可能性がある。

3 利用者による質問回答の管理

SOの質問回答は、様々な形で利用者によって管理される。まず、重要な管理者である「モデレータ」が利用者の中から毎年投票によって選出される。また、SO内において利用者が回答などを行うとその貢献度が自動的に算出され、貢献の度合いによって編集、重複した投稿の「クローズ」、ベストアンサーの変更などより多くの編集権限を得ることができる。さらに、質問回答やSOでの活動自体を議論するための「メタ」質問回答サイトが設置されており、「本線」であるSOから「SOに関する」議論を排除し、自治によってサイトを運営する基盤を提供している。

その中でも重要なのが、「編集」であり、これは「SOの核」ともされている。ある質問回答で質問者が満足したとしても、他の利用者にとってわかりやすいとは限らない。このため、主な編集の役割は、文法の修正や質問回答の明確化、

情報源の追加などの「改善」である。

4 「検索」への志向

Googleなどの検索エンジンにSOの質問回答が現れることは企画段階の文書から考慮されており、SOの「窓口」は検索エンジンであるとすら言われている。そして、SO以外の情報源も含めてより良い検索結果をもたらすように、SOの内部を十分に整備することが重要であるとしている。これは、先に述べた質問回答の多様性にも関連しており、1つの問題に別の表現を行えると、同じ情報ニーズの様々な表現に対応できる。また、重複した質問が少ないと必要なだけの検索結果を出すことができる。このように、先に挙げた方針は個別の質問回答のみならず、SOと外部の窓口となる検索と強く関連している。

C まとめ

SOは、従来のプログラミングに関する知識を互いに教え合うサービスの問題点として、問題解決がその場限りになってしまい、検索されるようなアクセスできる情報源としては不十分であるという問題を提起し、他の人にも役立つように質問回答を整備する方針を設定した。それを共同体で「改善」できるものとし、個別の質問回答を越えた「プログラマーにとって」役立つ情報源であることを目指している。のみならず、プログラミングという主題特有の、問題に対して様々な表現と様々な解決策があることをあくまで損なわないようにしている。

III 質問回答サイトに関する先行研究

A 概要

SOは質問回答サイトとして研究されていることが多い。このため、質問回答サイトについて何が研究されているかを、先行研究から明らかにする。質問回答サイト研究は2005年から活発になり始め、計算機科学、図書館・情報学など多分野からの研究が蓄積されている。主要なレビュー論文が2つ存在する。Gazanは2010年までの研究を対象に、分野を問わず質問回答サイトに関する研究をまとめている⁶⁾。Srbaらは2015年までの研究を対象に、計算機科学分野の中で隆盛している人工知能や機械学習によるアプローチに限ってまとめている⁷⁾。2011年以降にも質問回答サイト研究のテーマは様々な方向に広がっているため、先行するレビュー論文を踏まえながら現状を把握するために2011年以降の文献を収集、検討した。

B 文献の収集と分類

図書館・情報学分野ならびに計算機科学分野における2011年以降の質問回答サイトに関する文献を収集した。質問回答サービスはヴァーチャル・レファレンス・サービス(VRS)を含め多種多様であるが、本研究では「オンライン上の」「人による」「参加者なら回答できる」サービスを質問回答サイトとした。論文誌の他に主要な会議録も含め、153件の英語文献を収集した。

質問回答サイトでは例えば質問と利用者の属性など、相互に関連する要素が多い。一方、回答の客観的な品質から利用者の行為に至るまで様々な関心の元で研究が行われている。その中で「質問回答サイト研究」を概観するために、「何に注目したか」という成果に注目した。その上で、回答、質問、利用者の特徴や役割、利用者の心理や行動、行為の過程、サイト全体の共同体や概念枠組の6つに先行研究を分類した。

1 回答に注目した研究

回答の内容の品質は、質問回答サイト研究の主要な研究テーマである。主な質問回答サイトではベストアンサーの選択や投票機能があるため、利用者による品質の基準と客観的な基準の双方から研究されている。初期の研究では利用者による品質の評価基準を内容面や情報源、感情面などで類型化し、その頻度が分析されている。その後、質問と回答の関係のネットワークなどの様々な要因も見出され、数理モデルによって質の高い回答の自動判別を行う研究も登場した。また、医療情報など特定の主題領域に特化した品質の評価基準の研究も行われており、SOもその1つに当たる。一方で、客観的な品質モデルを用いた研究も存在し、VRSなどとの比較も行われている。

2 質問に注目した研究

質問の内容は利用者の行動や回答の内容な別の要素に付随したものと扱われることが多い。情報探索型の質問や助言を求める質問などの質問の種類の違いが、質問回答サイトの種類によってどう違うかなどが分析されている。このほか、質問の種類が回答の速度にどう影響するか、そもそも回答されない質問はどのようなものかなど研究も行われている。SOの研究では、質問の種類分析によりプログラマーが持つ課題などが明らかになっている。

3 利用者の特徴や役割に注目した研究

質問回答サイトの利用者の中で、異なる役割が自発的に生まれることが指摘されている。専

専門性の高さはわかりやすい例であるが、専門家の他に知識のまとめ役が存在することも指摘されている。専門性はプロフィール、回答数などの指標でモデル化され、ある主題に関する専門家を判別する研究も行われている。SOの研究でも同様に初心者と熟練者の違いは重要になり、利用者の専門性の自動判別などが行われている。

4 利用者の心理や行動に注目した研究

質問や回答などの行動や、それを行う社会的心理的要因は、情報行動研究の枠内で関心を持たれてきた。回答の品質や信頼度、質問回答の動機、情報源の提示、議論などの質問回答以外の行為といった、質問回答サイトの様々な側面が利用者の行動の特性によって説明されている。行動を支援する研究もなされており、例えば情報源を気にするには対立する情報をあえて提示するなどのことが有効であることが明らかになっている。SOにおいては質問回答以外に編集の傾向の分析も行われており、先に挙げた方針が品質に寄与していることが明らかになっている。

5 質問回答サイトで行われる行為の過程の研究

質問回答サイトで質問行為が始まって回答や評価に至る過程を、モデル化したり概念枠組を提示する研究も行われている。質問、回答、コメントなどをペアとして見る研究はその端緒である。また、質問と回答の間に動機や文脈、技術などを取り入れて過程をより質問回答のモデルを詳細化する研究も存在する。一方で、社会的学習共同体などのモデルでより抽象的にプロセスを説明する研究も存在する。SOにおいては、回答が時間を追ってよりよい知識となっていくように修正されていく時系列的パターンが明らかになっている。姉妹サイト MathOverflow の問題解決の過程を類型化した研究も行われている。

6 サイト全体の統計や共同体、概念枠組の研究

特定の、もしくは複数の質問回答サイトを総合的に捉える研究としては、まず「質問回答サイトとは何か」を定義し、比較を行う研究が存在する。質問回答サイトの分類の提案、VRS との比較などが見られる。また、質問回答サイトを人々が集まる「共同体」、特に人々が学び合う実践共同体として捉える研究がなされている。知識の集まりとして捉える研究も行われており、伝統的な情報検索と異なり人との交流による「ソーシャルサーチ」であるというモデルを提示したもの、図書館のような知識の保管庫から探していくものではなく口伝による「村パラダイム」であることを提唱したものが存在する。SOにお

いては、利用者当たりの質問回答数が他の質問回答サイトと同じくべき分布に従うこと、メーリング・リストから SO に質問が移行していく過程などが明らかになっている。

C まとめ

以上のように、既存の質問回答サイト研究においては、個別の質問回答に焦点を当て、その質問や回答の内容、利用者の動機や品質の評価などの要素を捉えようとする研究が多い。また、個別の質問回答を越えた研究としては質問回答サイトを共同体として定義した研究が見られ、個々の人々の交流が共同体を形成し、その上で知識が流通していく特徴が明らかになっている。

IV 結論

既存の質問回答サイト研究では、SO のような質問回答を参加者自身が知識のストックとして組織していくという側面の議論が欠けている。SO における参加者による知識の管理に目を向けた研究は 2014 年以降現れつつあり、編集の時系列的な特性や、「メタ」の議論への参加者がどう SO の質問回答を行っているかの統計的研究も登場している。しかし、参加者による知識ストックの管理が実際の質問回答という活動の中でどのように行われるかに関しては、実践に焦点を当てた理解が必要である。

謝辞

本研究は平成 29 年度慶應義塾大学大学院博士課程学生研究支援プログラムの研究助成を受け実施しました。

参考文献

- 1) Reagle, Joseph. Good Faith Collaboration: The Culture of Wikipedia, MIT Press, 2010.
- 2) Stack Exchange Inc. "Stack Overflow". <https://stackoverflow.com/> (accessed 2017-09-11).
- 3) Mamykina, Lena; Manoim, Bella; Mittal, Manas; Hripcsak, George; Hartmann, Bjorn. "Design lessons from the fastest q&a site in the west". 2011 Annual Conference on Human Factors in Computing Systems (CHI '11). ACM, 2011, p.2857-2866.
- 4) Atwood, Jeff. "Introducing Stackoverflow.com". CODING HORROR. 2008-04-16. <http://blog.codinghorror.com/introducing-stackoverflow-com/>, (accessed 2017-09-11).
- 5) Atwood, Jeff. "Stack Overflow: None of Us is as Dumb as All of Us". CODING HORROR. 2008-09-16. <http://blog.codinghorror.com/stackoverflow-none-of-us-is-as-dumb-as-all-of-us/>, (accessed 2017-09-11).
- 6) Gazan, Rich. Social Q&A. Journal of the American Society for Information Science and Technology. 2011, vol.62, no.12, p.2301-2312.
- 7) Srba, Ivan; Maria Bielikova. A comprehensive survey and classification of approaches for community question answering. ACM Transactions on the Web (TWEB), 2016, vol.10, no.3, 18p.