

## 知識共有コミュニティを創り出す人たち(3) 「回答者」データから見るコミュニティ内の「知識」

○三浦 麻子 (神戸学院大学人文学部), 川浦 康至 (東京経済大学経済学部), 地福 節子, 大瀧 直子, 岡本 真 (ヤフー株式会社; 非会員)

キーワード: 知識共有コミュニティ, Web 調査, 情報提供者

### 問題

インターネットは未曾有の規模の情報データベースとして、今や人間社会に不可欠な存在である。インターネット（多くは Web）上の公開情報は、偏在するのではなく、ハイパーリンクを通じて集約される。莫大な情報はやがて知識として結晶化し、それらとの接触はわれわれに創発を生み出す機会を提供してくれる。中でも近年、利用者たちによってオンライン上でリアルタイムに情報の登録や更新がおこなわれ、一種のコミュニティとしても機能する情報データベースサービスに対する注目が集まりつつある。これをここでは知識共有コミュニティとよぶ。世界規模で展開されるフリー百科事典ウィキペディア・プロジェクトは、こうした知識共有コミュニティの顕著な例であるが、その情報の信頼性は、伝統ある著名な百科事典ブリタニカと比較検討され、実際それに比肩する内容をもつと評価される (Giles, 2005) ほどの影響力をもつ存在に成長している。

こうした知識共有コミュニティは、無数のインターネット利用者のボランティア的貢献によって支えられている。彼らがなぜこうしたコミュニティに関わろうとするのか、またそこから何を得ているのかを考えることは、そこで共有される知識の姿を明らかにするための大きな手がかりとなる。本研究はこうしたコミュニティサービス大手「Yahoo! 知恵袋」の利用者を対象とした意識調査をおこなったものである。特に本発表では、知識提供、すなわち「回答」をおこなった参加者のデータにもとづいて、コミュニティ内で共有されている知識の姿を明らかにすることを試みる。

### Yahoo! 知恵袋の概要と特徴

Yahoo! 知恵袋は、ヤフー株式会社が運営している利用者参加型の知識共有コミュニティサービスである<sup>1</sup>。無料発行の Yahoo! JAPAN ID を持つ者であれば誰でも登録・参加でき、日常のあらゆる疑問を質問したり、またそれらの質問に対して回答できる。2004 年 4 月 7 日にベータ版公開・運用が、2005 年 11 月 7 日に正式版サービスが開始された。

Yahoo! 知恵袋は、参加者同士の対人コミュニケーションを重視する知識共有コミュニティである。利用者は、質問を投稿する積極的な情報獲得希望者、自らがもつ情報を質問に対する回答というかたちで投稿する情報提供者、投稿された情報を閲覧する受動的な情報獲得者の立場から Yahoo! 知恵袋に関わることができる。質問の内容やジャンルに制

約はなく、自由に情報提供を求めることができる。また、正解のある質問だけでなく、いわゆる ill-defined な問題解決を求める質問もできる。コミュニティの究極的な目標は集積される知識が百科事典的な意味で有用なサービスとなることだが、「参加者同士で楽しく知恵や知識を教えあい、分かち合」うための対人コミュニケーションの場となることも重視した知識共有コミュニティである。

### 調査

**概要** Yahoo! 知恵袋上の利用者たちを対象に、情報獲得希望者（質問者）対象と情報提供者（回答者）対象の 2 種類の質問紙を作成し、Web 上で調査をおこなった。

**実施方法** 調査は、2005 年 12 月 6 日～20 日の 15 日間に、Yahoo! 知恵袋サイト上で実施された。調査対象者は、当該期間中に Yahoo! 知恵袋サービスにログインし、質問か回答のいずれかを投稿した利用者である。ある利用者の Yahoo! 知恵袋での質問/回答の投稿が終了した時点で、各 1 回限りで調査協力依頼画面が呈示された。両方の調査に回答した利用者については時間的に早い回答のみをデータとして採用した。最終的に質問者用調査 5515 件、回答者用調査 2474 件が分析対象となった。特に本報告では、このうち回答者用調査のサンプルに関する分析結果を主に報告する。

**質問項目** すべての調査協力者に問うた項目は、1) 基本的属性（性別・年齢）、2) コミュニティでの行動と意識（知恵袋利用開始時期・最近 1 ヶ月の知恵袋利用頻度・貢献度への関心・他の利用者とのコミュニケーション欲求・今後の利用意向）、3) 情報欲求に関する尺度（川浦, 1998）から抜粋した 12 項目（4 件法）である。共通設問に加えて回答者用調査のみで問うた項目は、1) 回答した質問の内容（正答の有無・緊急度と重要度の推測）、2) 回答を投稿した理由を問う尺度（13 項目・4 件法）、3) 投稿した回答の評価、回答内容について自ら調査したかどうかとその手段、である。なお、複数の質問項目で構成されている尺度は、回答者ごとに項目の呈示順序をランダムに変化させ、順序効果を相殺した。

**調査協力者のコミュニティ参加状況** 質問紙調査に回答したすべての調査協力者について、コミュニティ参加状況の指標となる客観的データを収集した。調査協力時点で投稿した質問/回答の内容、当該時点までの質問投稿数、回答投稿数、投稿した質問の被回答数、および、コミュニティへの登録日時、最初の質問投稿日時、最初の回答投稿日時、公開 ID 数である。初回登録以降に、知恵袋専用の公開 ID を変更している調査協力者の場合は、過去のすべての公開

<sup>1</sup> URL: <http://chiebukuro.yahoo.co.jp/>

ID 分のデータを合算したものを分析に利用した。

## 結果と考察

**基本的属性** 回答者用調査の協力者の基本的属性は、平均年齢 34.3 歳、男女比は男性 54.7%、女性 45.3%であった。

**回答対象の質問について** まず、自らが直近に回答した質問については「決まった答や正解があるはず」が 1048 件 (42.4%)、「特に決まった答や正解はない」が 1426 件 (57.6%) であった。同様の質問について、質問者は自身の質問は「正解あり」67.5%、「正解なし」32.5%であると答えている。両データに対応はなく、また互いの認知の歪みを考慮に入れる必要はあるが、「正解なし」質問の方が答えられやすい傾向があることが考えられる。また、当該質問に関する回答者の「質問者にとっての重要度の推測」は「急いでいる」が 36.0%、「重要である」が 59.0%であった (同・質問者による「緊急度の認知」は 73.4%、90.1%)。質問者のもつ重要度や緊急度は、回答者により過少評価されており、その傾向は緊急度において特に強いことが考えられる。

**回答時の行動** 「正解あり」質問に回答した調査協力者に、回答時の事前知識と正解確認の有無を問うた結果を表 1 に示す。質問の正解が既知であったとする割合が大多数で、その真偽を特に確認せずに回答したとする調査協力者がそのうち 2/3 程度を占めていた。さらに、事前知識の有無によらず、回答時に何かを「調べた」場合の手段を多肢選択で問い、各手段の利用頻度と、各協力者が利用した手段の総数、および各情報源が利用されたか否かのデータにもとづいてクラスタ分析 (最近隣法・平行ユークリッド距離) をおこなった結果を表 2, 3 と図 1 に示す。表 2, 3 を見ると、真偽の確認をした回答者の場合も、その際に利用した手段はごく少数に限られており、またサーチエンジンを用いた場

合がほぼ半数であったことがわかる。また、図 1 に示したデンドログラムから、利用される情報源のパターンはネット/非ネット込みで 1) 動的情報、2) 静的情報、そして 3) サーチエンジンの 3 クラスに分類できると考えられる。

これらの結果から、回答者によって提供される情報、すなわち知恵袋に蓄積されていく「知識」は、回答者がその質問に接した時点で既に持っていた情報が主体であることが明らかになった。このことは、Yahoo! 知恵袋コミュニティでは、参加者の経験や記憶が「知識」として多く共有されていることを示しており、こうした「知識」は通常のサーチエンジンによる検索で得られる Web 上のドキュメントに基づく「知識」とは異なる特徴をもつことが考えられる。また「セカンドオピニオン」の確認については、おこなっている回答者は相対的に少数で、おこなっている場合のリソースの多くはサーチエンジンの検索結果であった。とはいえ、この点のみをもって、知識共有コミュニティの「知識」の信頼性の基盤は多分に脆弱なものであるとか、またそれがこの種のコミュニティにとってクリティカルな問題であると結論づけたりするのは性急であろう。インターネット上の情報共有コミュニティの『「みんなの意見」は案外正しい』(スロウィッキー, 2006) か否かを検討するためには、利用者にとって何が「正しい」情報なのかをも含めて、今後さらなる検討が必要である。

## 参考文献

- 1) Giles, J. 2005 Internet encyclopedias go head to head, *Nature*, **438**, 900-901.
- 2) 川浦 康至 1998 情報欲求と情報行動 広告月報 1998 年 1 月号, 42-47.
- 3) ジェームズ・スロウィッキー 2006 「みんなの意見」は案外正しい 角川書店

表 1 「正解あり」質問への回答者の正答既知/未知と確認の有無

| 答の既知/未知と確認の有無  | 度数  | 比率   |
|----------------|-----|------|
| 答既知・真偽を再確認せず回答 | 619 | 59.1 |
| 答既知・真偽を確認して回答  | 300 | 28.6 |
| 答未知・調べずに回答     | 77  | 7.3  |
| 答未知・調べて回答      | 52  | 5.0  |

\*\*\*\*\*HIERARCHICAL CLUSTER ANALYSIS\*\*\*\*\*

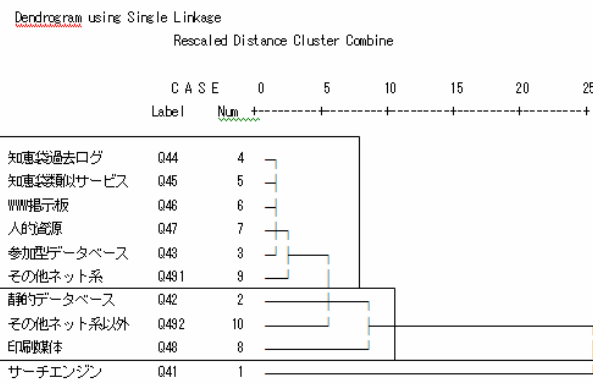


図 1 各情報源のクラス分析結果

表 2 正解確認の際の各情報源の利用頻度

| 調べた手段*    | 答既知 | 答未知 |
|-----------|-----|-----|
| サーチエンジン   | 174 | 35  |
| 書籍新聞雑誌    | 58  | 3   |
| その他ネット以外  | 44  | 4   |
| 静的データベース  | 39  | 5   |
| その他ネット系   | 21  | 3   |
| 参加型データベース | 16  | 7   |
| 人に聞く      | 14  | 0   |
| 知恵袋内過去ログ  | 11  | 3   |
| 知恵袋類似サービス | 6   | 3   |
| 掲示板       | 8   | 1   |

\*答既知の場合の度数に基づき降順ソートしてある

表 3 調査協力者の各情報源の利用総数

| 利用手段数 | 度数  | 比率   |
|-------|-----|------|
| 1     | 270 | 76.7 |
| 2     | 58  | 16.5 |
| 3     | 20  | 5.7  |
| 4     | 2   | 0.6  |
| 5     | 1   | 0.3  |
| 6     | 1   | 0.3  |