

平成 21 年 6 月 11 日



## 「Yahoo!知恵袋」コーパス第一弾を利用した研究成果例

### 【Q&A コミュニティにおける回答の信頼性指標の構築】

誰もが質問・回答できる場である Q&A コミュニティでは、多数寄せられる回答の中からどの回答が信頼できるものかを、回答そのものに対する知識の乏しい質問者自信で判断することが難しいという問題があります。本研究は、Q&A コミュニティの質問者が信頼性の高い回答を効率的に入手できるよう支援するための「回答の信頼性指標」の構築を目指しています。Yahoo!知恵袋データを対象とした我々の分析から、以下の知見が得られました。

- (i) 回答が一意に定まる質問では、最後の回答が投稿されてからの経過時間と回答の閲覧数が信頼性指標になり得る
- (ii) 回答を裏付けるような外部参照 (URL など) の有無が信頼性指標になり得る
- (iii) 質問と最良回答が多数投稿される時間帯が異なるカテゴリでは、質問と回答の時間帯が信頼性指標になり得る

今後さらに分析を続ける共に、質問者支援に向けた方式を開発する予定です。

■ 問合せ先：奈良先端科学技術大学院大学 大平雅雄 (masao@is.naist.jp)

HCI Research group (hci@is.naist.jp)

■ 参考 URL：[http://se.naist.jp/html/HCI/QA\\_community/](http://se.naist.jp/html/HCI/QA_community/)

### 【Q&A サイトへの質問の作成を支援するための情報の抽出】

Yahoo! 知恵袋に代表されるコミュニケーションサイトがさかんに利用されています。なぜこんなに人気があるのか考えたのですが、そこでは「教えてほしい」「助けてほしい」と考えている人と「教えてあげたい」「だれかのためになれば」と考えている人が出会い、心豊かなコミュニケーションが行われているからではないでしょうか。

こうした心豊かなコミュニケーションを促進するためには、「教えてほしい」「助けてほしい」と考えている人が適切な質問を簡単に作成できることが重要です。そこでわれわれは、ユーザが「教えてほしい」と思っていることについて質問するのに役立つ情報を示すことで、答えてもらいやすい質問を作成するのを支援するシステムを開発することをめざしています。

下記の論文では手始めに、初心者が質問を作成するのに役立つ情報をコミュニケーションサイトに投稿された質問と回答から取り出す方法について検討しました。

磯貝直毅, 西村涼, 渡辺靖彦, 岡田至弘「Q&A サイトへの質問の作成を支援するための情報の抽出」

情報社会学会論文誌, Vol. 3, No. 2, pp. 5-14 (2009).

■ 問合せ先：龍谷大学エクステンションセンター

滋賀県大津市瀬田大江町横谷 1-5 TEL / 077-543-7743 FAX / 077-543-7771

### 【QA サイトコンテンツの信憑性分析および Web クエリの意図発見】

京都大学情報学研究科の田中克己教授の研究室では、Yahoo!知恵袋データを用いて、

- (1) QA サイトコンテンツの信憑性分析の研究
- (2) QA サイトコンテンツを用いた Web クエリの意図発見の研究

を行っています。前者は、Yahoo!知恵袋の回答データが正しいかどうかや別解の発見などを通常の Web から関連する情報をサーチして行う研究です。後者は、Web 検索エンジンに投入されるキーワードクエリに関連する知恵袋質問文をサーチ・分析し、この結果に応じてクエリに対する検索結果をその意図に応じて自動分類するという研究などです。

すでにその研究成果の一部は、学会等で公表しています。例えば、高田 夏希、山本 祐輔、小山 聡、田中 克己：「質問応答コンテンツに対する Web による回答補完」、DEIM Forum 2009 C4-6、2009 年 3 月などです。

- 問合せ先：京都大学 田中克己 (tanaka@dl.kuis.kyoto-u.ac.jp)
- 京都大学 小山 聡 (oyama@dl.kuis.kyoto-u.ac.jp)

### 【コミュニティの活性度評価】

Yahoo!知恵袋は、質問と回答という形態での生活者情報の交換サイトと考えることができます。質問と回答の組で形成されるコミュニティで、複数の回答が連なって議論が盛り上がる要因は何でしょうか？

私たちは、質問の内容だけでなく最初に回答した第 1 回答者の発言が、その後の議論(回答)の方向付けに影響を与えているのではないかと考え、第 1 回答者の分布をカテゴリと時間の 2 つの視点から分析しました。その結果、パソコンとダイエットのようなカテゴリで分布に違いがあること、時間の経過とともに第 1 回答者の分布が拡がり、より多くの回答者が積極的に議論に参加する傾向が見られました。

第 1 回答者の分布などコミュニティの発展や成長過程を捉える指標を明らかにすることで、とかく閉鎖的な社会となりがちな Web コミュニティをオープンに活性化する手法を開発したいと考えています。

- 問合せ先：筑波大学 佐藤哲司 (satoh@slis.tsukuba.ac.jp)
- 参考 URL：<http://ce.slis.tsukuba.ac.jp/>

### 【Web 質問応答システムの開発】

我々が開発した Web 質問応答システムは、質問文を入力すると、その「答」の候補をインターネット上の文章から見つけてきてくれます。例えば、「～するにはどうしたらいいですか?」という質問に対しては、「まず、～を準備してから、～したらいいですよ。」といった回答がなされるように、質問の種類によって回答の仕方が違ってきます。

そこで、Yahoo!知恵袋に投稿された文章から、このような質問の仕方にはこのような回答の仕方をするという情報を抽出し、利用者の質問にうまく答えられるようにしています。

注意していただきたいのは、利用者の質問の答を Yahoo!知恵袋データから取り出しているのではない点です。質問の答はインターネットから探してきますので最近の話題にも答えられます。

- 問合せ先：横浜国立大学 森辰則 (mori@forest.eis.ynu.ac.jp)
- 参考 URL：Web 質問応答システム MinerVA (ミネルバ)

<http://www.forest.eis.ynu.ac.jp/MinerVA-ND/>