

Web 2.0 に向けた新たな情報検索技術の研究を開始 — 「Yahoo!知恵袋」の研究利用による情報アクセスの新展開 —

平成 19 年 3 月 6 日

国立情報学研究所

(要点)

国立情報学研究所（所長：坂内正夫（さかうち まさお）以下、NII）は、Web2.0の普及にともなったWeb上の利用者発信情報を対象とした情報検索、情報分析、情報活用などの研究を開始します。このため、NIIは、平成19年3月6日（火）にヤフー株式会社と知識検索サービス「Yahoo!知恵袋」データの研究利用契約を締結し、このデータを情報検索、情報分析、情報活用などの情報関連技術研究コミュニティに対して無償提供を行います。

Web2.0の普及にとともない、飛躍的な増大を続けるブログなどの利用者発信情報の研究利用は、情報爆発から新しい価値を汲み出すために必須のデータです。しかし、従来、これらは個人情報を含む上にプロバイダー等の所有による非公開データであったため、研究コミュニティからは手の届かない状況にありました。今回、「Yahoo!知恵袋」データの研究利用が実現することは、この壁を破るものであり、情報爆発研究の新しい第一歩を踏み出すものです。

NIIでは、これを契機に、「Yahoo!知恵袋」データの研究利用を、大量で多様なデータの活用によって新しい研究分野を切り開くe-Science (Data-centric Science) の一環と捉え、従来からの学術論文、図書、研究報告などの研究成果の学術情報サービスに加えて、e-Scienceデータの研究コミュニティへの提供を強化します。

(具体的内容)

NIIでは、情報の未来価値として、情報爆発から「情報や情報分析結果から新しい価値を発見したり、汲み出したりする」ことの重要性に着目し、このための研究の開始と、これに必要なデータ利用に関する準備を進めてきました。ここで新しい価値を汲み出す情報源の一つとして注目したのが、従来の書き言葉を主体としたWeb上のテキスト情報に対して、「書き言葉と話し言葉」の中間的存在として、急速に増大している利用者発信テキスト情報（ネット日記、質問・回答文、ブログなど）です。

しかし、利用者発信情報の利用には、下記のような壁がありました。

- ① 研究用に利用できる利用者発信情報データベースが存在しない（データベースはプロバイダーなどが企業戦略として所持しており非公開）
- ② 利用者発信情報は本質的に個人情報を含む（個人情報やプライバシー保護の原則）
- ③ 利用者発信情報を構成する言語には、従来の言語解析（構文解析や意味解析など）が通用しない

URL:<http://www.nii.ac.jp/>

National Institute of Informatics

この①と②の壁を超えて、情報検索、情報分析、情報活用などの情報関連技術研究の新展開への一歩を踏み出したのが、ヤフー株式会社が運営する知識検索サービスの質問・回答文で構成される「Yahoo!知恵袋」データの研究利用の実現です。また、「Yahoo!知恵袋」データのNIIによる研究コミュニティへの無償提供の開始によって、情報関連技術研究コミュニティが、③の壁への挑戦を開始するデータベースを入手し、この分野のe-Scienceとしての新しい研究課題と研究成果を生み出すことを可能にします。

NIIでは、これらの研究課題の企画・実施を推進するため、情報爆発科研費特定領域研究、情報分析、情報信頼性評価などの研究プロジェクトでの利用を進めるとともに、「情報検索システムの性能評価」という切り口で推進してきたワークショップ型共同研究「情報検索システム評価用テストコレクション構築プロジェクト（NTCIR）」での活用を予定しています。

(今後の課題)

「Yahoo!知恵袋」データの研究利用は、Web上に日々生まれる「生きた情報」としての多様な利用者発信情報を活用したe-Scienceへの第一歩でしかありません。このデータベースを最大限に利用することによって、新しい研究価値を生み出すと同時に、多様な利用者発信情報（ネット日記、質問・回答文・ブログ・Webメールなど）の研究利用の促進に向けて活動を続けます。

ヤフー株式会社からのコメント： 岡本 真

1996年のサービス開始以来、Yahoo! JAPANは日本のインターネット利用の拡大と共に歩んできました。これまで多くのお客様にご利用いただいた成果を、研究目的でのオープンなデータ公開という形でシェアさせていただくなど、学術コミュニティに対するさまざまな協力によって、インターネットのさらなる発展と可能性の探求に貢献していきたいと考えております。

(用語解説)

【情報検索システム評価用テストコレクション構築プロジェクト（NTCIR）】

国立情報学研究所で主催するワークショップ型共同研究で、国内外の学術・研究機関や企業の参加を得て、情報検索、テキスト要約・質問応答・情報抽出などの情報アクセスのためのテキスト処理技術の研究を、「検索システムの性能評価」を重視して推進してきました。NTCIRでは、国内外の数十グループが参加する評価ワークショップを定期的に行い、共通の実験データを用いたシステム評価基盤を提供するとともに、システム間の相互比較、研究上のアイデアの交換等を実施することによって、情報検索技術と性能向上とテキスト処理研究分野の発展に貢献してきました。

URL:<http://www.nii.ac.jp/>

National Institute of Informatics

【Yahoo! JAPAN】

ヤフー株式会社が運営する Yahoo! JAPAN は、1 か月あたり約 4,671 万人のユニークカスタマー数 (※1) と、1 日 13 億 7000 万ページビュー (※2) のアクセスを数えるインターネットの総合情報サイトで、検索、コンテンツ、コミュニティー、コマース、モバイルなど多くのサービスを提供しています。

(※1) 2007 年 1 月の Nielsen/NetRatings 「NetView AMS JP」における家庭からの視聴率 88.2%、職場からの視聴率 91.5%というデータをもとに、家庭、または職場からのインターネットユーザーを約 5,284 万人 (NetRatings Japan「インターネット基礎調査」より)として Yahoo! JAPAN のユニークカスタマー数を算出。

(※2) ページビューは、従来ウェブサーバーからの配信回数を計測した値を採用していましたが、2007 年 3 月からブラウザでの表示回数を CSC により計測した値を採用しています。

【Yahoo!知恵袋】

「Yahoo!知恵袋」は、質問したい人と回答したい人とを結び、知恵や知識を参加者同士で共有する、日本最大の知識検索サービス (Q&A サイト) です。2004 年 4 月のサービス開始以来、質問総数 1000 万件以上、回答総数は 3500 万件を数え、現在は約 94 万人が登録するナレッジコミュニティです。

■本件問い合わせ先

東倉 洋一 教授・副所長

大山 敬三 コンテンツ科学研究系教授・研究主幹

取材窓口／その他問合せ

国立情報学研究所 (NII: エヌアイアイ)

国際・研究協力部 広報普及課 企画・広報係 担当: 小野・早川
〒101-8430 東京都千代田区一ツ橋 2-1-2 (学術総合センター 18 階)

TEL: 03-4212-2135 (直通) FAX: 03-4212-2150

e-mail: kouhou@nii.ac.jp

URL: <http://www.nii.ac.jp/>