

解禁日時：2010年3月8日報道発表終了後

National Institute of Informatics

平成22年3月8日  
国立情報学研究所  
国立国語研究所  
ヤフー株式会社

## ウェブユーザーに配慮した大規模データ研究利用ガイドラインの発表 — 「Yahoo!ブログ」のデータおよび研究利用促進のためのガイドラインの公開 —

国立情報学研究所（所長：坂内正夫（さかうち まさお）以下、NII）と国立国語研究所（所長：影山太郎（かげやま たろう）以下、国語研）およびヤフー株式会社（代表取締役：井上雅博（いのうえ まさひろ）以下、ヤフー）は、「Yahoo!ブログ」のデータの一部を日本語コーパスとして公開、そしてWeb上の利用者発信情報（以下、CGM（Consumer Generated Media：消費者生成メディア））を対象とした情報検索、情報分析、情報活用などの研究および現代日本語のコーパス言語学的研究促進のためのガイドラインを発表しました。

ウェブの普及にともない、増大を続けるブログなどのCGMの研究利用は、現代日本語の研究や、情報爆発から新しい価値を汲み出すために研究データとして注目される一方、個人情報保護の観点から研究利用に関する困難さが指摘されています。特に、ブログは個人の意見や個人の特長に結びつく可能性のある表現が多く含まれるため、その適切な取り扱いに関するガイドラインの重要性が増しています。

NIIと国語研及びヤフーは、平成20年度よりNIIの東倉洋一副所長を中心に、研究利用に関するガイドラインの研究グループを発足し、WEB上の大規模CGMであるヤフー株式会社「Yahoo!ブログ」を事例として、コンテンツ生成者であるユーザが安心できるデータ利用に関するガイドラインについて研究して参りました。

ガイドラインに沿って作成したデータの一部（500万語）を国語研から一般向けに日本語コーパスとして公開を開始いたします。

NIIからは、研究者向けのデータとして提供を開始いたします。

ブログは、誰でも自由に情報発信を行うことができ、論説から独り言までを含む、世の中を反映する鏡のようなものです。ブログが社会に及ぼす影響も日々増大し、言語学、国語教育、日本語教育、辞書編集、自然言語処理、さらにはブログの実態把握や意見分析、話題分析を通じたマーケティングや世論調査などへの利用の可能性も広がっています。

NIIの東倉洋一副所長を中心としたNII、国語研、ヤフー株式会社および東京大学大学院情報理工学系研究科（以下、東大情報理工）などの研究者グループの共同研究の成果は、ブログを研究に利用するためのガイドラインとしてまとめたことです。本ガイドラインは、今回の国語研及びNIIからのブログデータの公開と同時にNIIより公開を開始いたします。今

解禁日時：2010年3月8日報道発表終了後

National Institute of Informatics

回のガイドラインが発表されたことにより、コンテンツ生成者であるユーザおよびデータを利用する研究者が、安心して提供・利用できる環境を整えられます。

なお、今後の展開として、NIIでは、ヤフーより提供された「Yahoo!ブログ」の全データ公開のための共同研究プロジェクト\*1を発足し、今回作成したガイドラインに沿ってデータを処理した上で、研究者向けデータとして、H22年度公開に向けて活動をします。

\*1: 共同研究プロジェクトメンバー

国立情報学研究所・コンテンツ科学研究系・教授 大山敬三

国立情報学研究所・副所長・教授 東倉洋一

国立情報学研究所・コンテンツ科学研究系・教授 安達淳

国立情報学研究所・音声メディアグループ 大須賀智子

国立国語研究所・言語資源研究系・教授 前川喜久雄

国立国語研究所・言語資源研究系・准教授 山崎誠

京都大学・大学院情報学研究科・教授 黒橋禎夫

東京工業大学・精密工学研究所・教授 奥村学

東京大学・生産技術研究所・教授 喜連川優

東京大学・大学院情報学環／大学院情報理工学系研究科・教授 辻井潤一

東京大学・大学院情報理工学系研究科・教授 石塚満

東京大学・大学院情報理工学系研究科・特任助教 木戸冬子

ヤフー株式会社・メディア事業統括本部・メディアサービス本部 ソーシャル企画部  
チームリーダー 寺岡宏彰

ヤフー株式会社・メディア事業統括本部・メディアサービス本部 ソーシャル企画部  
東保知子

解禁日時：2010年3月8日報道発表終了後

National Institute of Informatics

(用語解説)

**【特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21世紀の日本語研究の基盤整備」】**

略称「日本語コーパス」は、平成18年度から22年度まで5年間のプロジェクト。

領域代表者：前川喜久雄（国立国語研究所言語資源研究系）

本研究には次のふたつの目標があります。

(1) 現代日本語のコーパス言語学的研究の基盤を整備するために、大規模な現代日本語書き言葉コーパスを構築すること

(2) 構築途上のコーパスを様々な領域で利用することによってコーパス日本語学の可能性を探り、同時に構築中のコーパスを評価すること

**【コーパス】**

言語分析を行うための基礎資料として、書き言葉や話し言葉の資料を組織的に収集し、研究用の情報を付与したうえで電子的に保存したもの。偏りのない形で対象の全体像を反映したデータとなっていることが望まれる（均衡コーパス）。

**【Yahoo! JAPAN】** <http://www.yahoo.co.jp/>

ヤフー株式会社（市場名：東証1部/JASDAQ、銘柄コード：4689、本社：東京都港区、設立年月日：1996年1月31日、代表取締役：井上雅博）が運営するYahoo! JAPANは、1か月あたり約5154万人のユニークカスタマー数※と、1日20億3000万ページビューのアクセスを誇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。※Nielsen Online「NetView」、2010年1月、家庭もしくは職場からのアクセスによる。

**【Yahoo!ブログ】** <http://blogs.yahoo.co.jp/>

「Yahoo!ブログ」はブログを簡単に作成、更新できるサービスで、Yahoo! JAPAN IDを取得していれば、誰でも無料で利用できます。2005年1月にサービスを開始した国内最大級のブログサービスです。

**【東京大学大学院情報理工学系研究科】** <http://www.i.u-tokyo.ac.jp/>

「大学院情報理工学系研究科（Graduate School of Information Science and Technology (IST)）」は、情報理工学分野の知識を統合的に備える研究拠点として、また、情報科学技術の教育研究を充実させるために2001年4月に設置されました。本研究科は、理学系研究科と工学系研究科から改組再編して、コンピュータ科学専攻、数理情報学専攻、システム情報学専攻、電子情報学専攻、知能機械情報学専攻の5専攻で発足しましたが、2005年4月には創造情報学専攻を設置して、あらたな展開を図っています。研究科では、これら情

解禁日時：2010年3月8日報道発表終了後

National Institute of Informatics

報理工学の基礎から応用まで幅広い分野の研究を融合させて、未来の情報科学技術を確立することを目指しています。

■本件問い合わせ先

国立情報学研究所

国立情報学研究所

コンテンツ科学研究系 教授 大山 敬三

E-mail : oyama@nii.ac.jp

国立国語研究所

言語資源研究系教授・系長 前川喜久雄

E-mail : kikuo@ninjal.ac.jp

---

取材窓口／その他問合せ

国立情報学研究所 企画推進本部広報普及チーム 担当：佐久間

〒101-8430 東京都千代田区一ツ橋 2-1-2

TEL:03-4212-2131 (直通) FAX:03-4212-2150

E-mail : kouhou@nii.ac.jp

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

国立国語研究所 管理部研究推進課 課長補佐 黒田清彦

〒190-8561 東京都立川市緑町 10-2

TEL: 042-540-4489 (直通) FAX: 042-540-4333

E-mail: [kiy-kuro@ninjal.ac.jp](mailto:kiy-kuro@ninjal.ac.jp)

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