

150万ブログを抱えるYahoo!ブログでインターネット全体に公開されている投稿記事データを、国立情報学研究所・国立国語研究所を通し、研究素材としてご提供いたします。

1 ブログ投稿データ提供の意義と提供イメージ

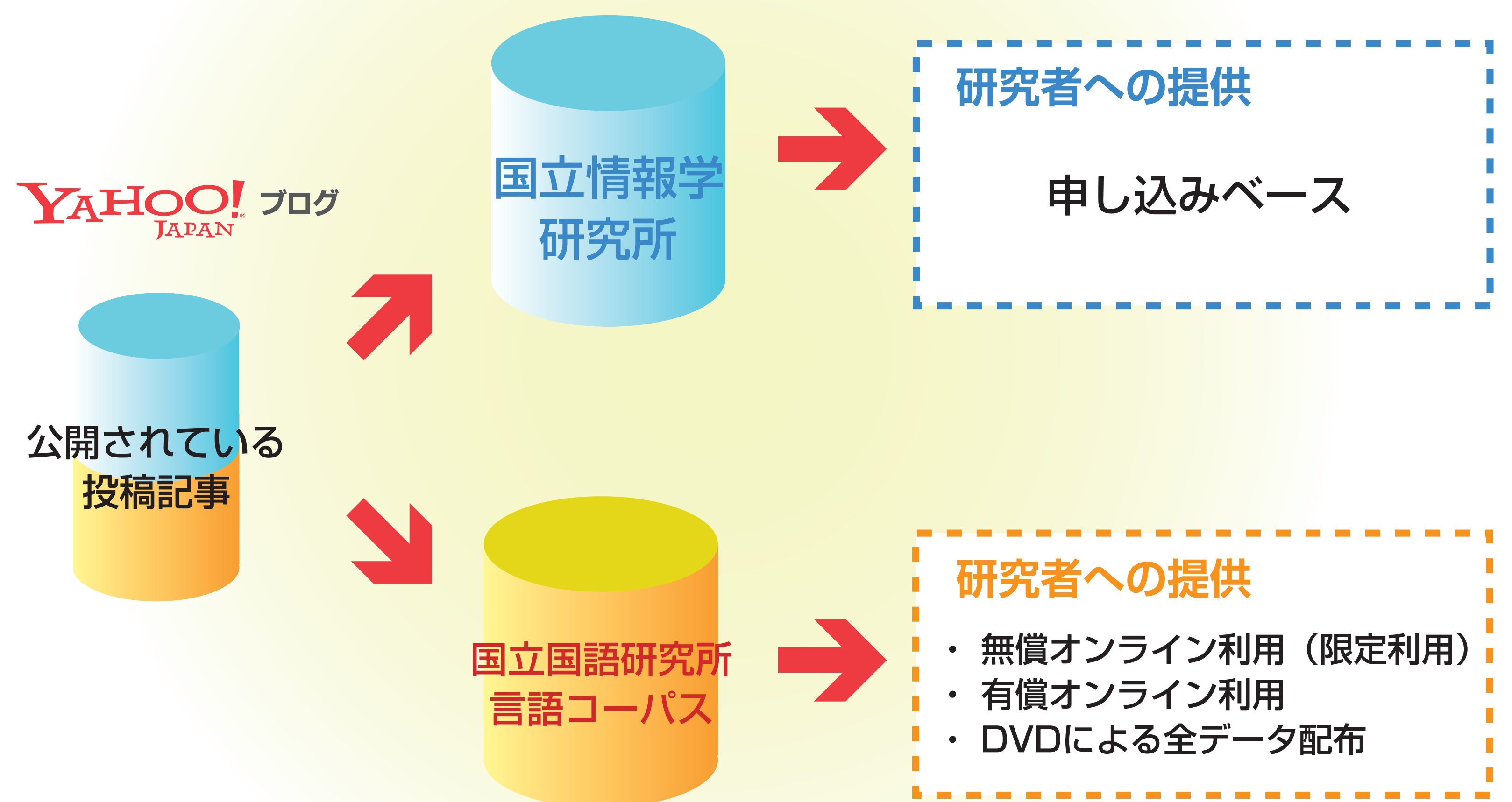
ブログサービスの出現後、一般ユーザーによる投稿データが爆発的な勢いで増え続けています。文法や既存の文字にさえ縛られないブログ投稿データは、まさに生きた言語そのものとして多くの研究者から注目されはじめています。データ提供は、国立情報学研究所・国立国語研究所を通してのみ提供予定です。データ利用の申し込み方法、提供データ詳細などは、各研究所にお問い合わせください。

投稿サンプル

そのライブは、まるでサーカス。
Piano, Bass, Drumsの三人が、これでもかと思えないテクニックを繰り広げ続ける。
3人の緊張感と、恍惚とした表情、これ以上に気持ちいい瞬間はないという笑顔。
何を見ても、何を聴かされても、「すごいすごいすごいすごい！」と口の中で小さくつぶやくしかなかった。
これほどまでに驚かされたライブパフォーマンスは、たぶん、はじめて。

後輩と、一風堂経由のダーツ。
久々にラーメン食べて幸せ(´∀`)
一風堂以上に久しぶり(つ~か年単位で久しぶり)のダーツは、シャフトが折れるくらい白熱w
戦績は3勝4敗。
後輩がすんげ〜ウマくていろいろ教えてくれて、超楽しかったです(´∀`)

データ提供のイメージ



2 大規模テキストコーパス整備における個人情報等取り扱いの検討

特定の個人に結びつく可能性のある表現などの情報の取り扱いを検討し、研究者コミュニティにおけるブログデータの取り扱いに関するガイドライン共同研究に参画します。

主な参画メンバー

国立情報学研究所	コンテンツ科学研究系・教授	大山 敬三
	副所長・教授	東倉 洋一
	コンテンツ科学研究系・教授・学術基盤推進部長	安達 淳
東京大学	生産技術研究所・教授	喜連川 優
東京大学大学院	情報理工学系研究科・教授	辻井 潤一
	情報理工学系研究科・教授	石塚 満
	情報理工学系研究科・特任助教授	木戸 冬子
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・准教授	奥村 学
国立国語研究所	研究開発部門・グループ長	前川 喜久雄

ブログユーザーと投稿の特徴

- ・若年層など個人情報開示のリスクについて意識の低いユーザーが含まれる
- ・個人的な内容を扱った日記が多い
- ・報道引用なども多く、容疑者など個人名をそのまま記載しているものが見られる
- ・社会活動や言論活動を目的とし、時には個人名が扱われることもある

投稿記事の引用発表に伴うリスクについて研究し、投稿者のプライバシーを守ることが求められます。

YAHOO! ブログ とは…

Yahoo! JAPANが提供するブログサービス。テキストや画像の投稿機能、コメント・トラックバック機能に加え、ブログ同士をつなぐお気に入り機能、相手を限定して記事を公開できる公開先限定機能、投票機能など、コミュニティ性の高いブログサービスです。

<http://blog.yahoo.co.jp/>

Yahoo!ブログの歴史

- 2005/1/31 サービス公開
- 2006/1 50万ブログ突破※
- 2007/1 100万ブログ突破※
- 2008/2 150万ブログ突破※
- モバイルサービス開始

※ 1ユーザー1ブログ。削除されたものを除く