



# Yahoo!知恵袋コーパスのこれから

～さらなる研究支援のためのYahoo!知恵袋コーパスデータ提供～

## Yahoo!知恵袋とは

Yahoo!知恵袋は、疑問に思っていることを質問したり、知っている事柄についての質問に回答することで、参加している方がお互いに知恵や知識を教えあい、分かち合えるサービスである。

質問したい人と回答したい人を結び、利用者同士が知恵や知識を結集し相互協力によって幅広く情報を共有できるサービスです。



## Yahoo!知恵袋コーパスとは

Yahoo!知恵袋コーパスは、投稿データ(ベータ版期間の未削除の質問・回答)を国立国語研究所、国立情報学研究所を通して、広く研究者の皆様提供しているコーパスである。

## Yahoo!知恵袋コーパスのこれから

提供を開始してから2年以上経過しており、コーパスの内容が古くなっている部分もある。また提供後にYahoo!知恵袋に投稿されたデータも多くあり、更に新たにリリースした機能からも研究を進める上で有益そうなデータが集まっている。

これらの状況を鑑み

## Yahoo!知恵袋コーパス第2弾

を提供いたします。

## 想定される研究テーマ

情報科学、社会科学、学際領域など多岐

- 質問応答システム
- 統計的自然言語処理のための統計データ
- コミュニティ抽出、分析
- 日本語学、書き言葉分析...

※Yahoo!知恵袋コーパスデータを利用した研究発表多数

## 提供のデータ素性

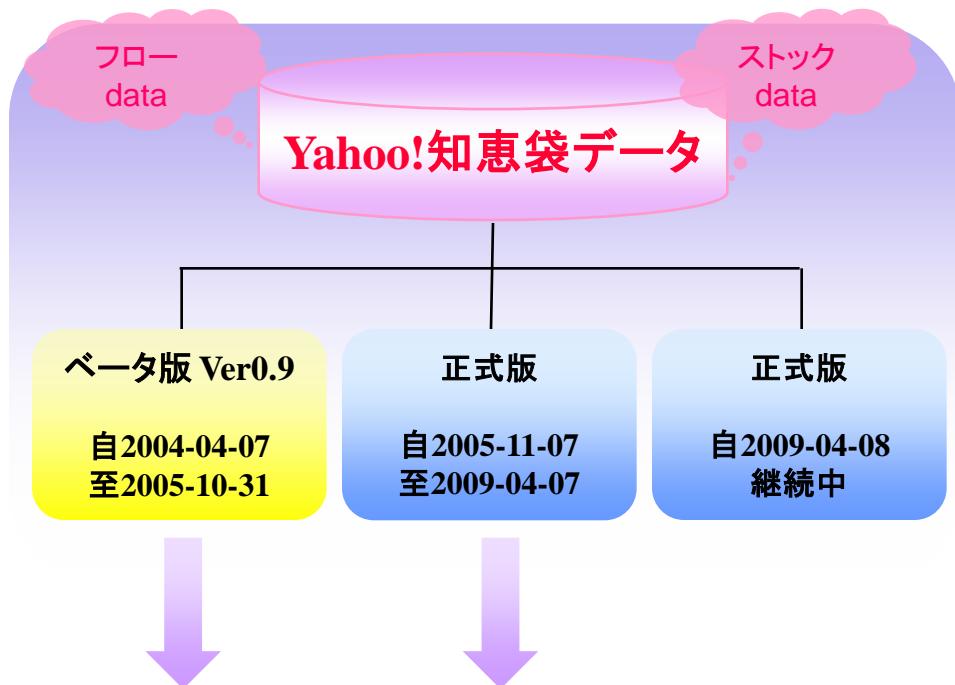
### Yahoo!知恵袋コーパス第1弾

質問数	約300万
回答数	約1300万
期間(年/月)	04/4 - 05/10

- 2005年11月:正式版公開
- 2006年 5月:モバイル版公開
- 2006年12月:リニューアル
- 2007年 9月:モバイル版リニューアル
- 2008年 6月:詳細ページUIリニューアル
- 2008年11月:画像投稿機能リリース
- 2008年11月:専門家回答開始
- 2009年 2月:カテゴリページUIリニューアル

### Yahoo!知恵袋コーパス第2弾

質問数	約2600万
回答数	約7300万
期間(年/月)	04/4 - 09/4



## NIIを通じてデータを無料公開

ヤフー株式会社 Yahoo!知恵袋プロジェクト

国立情報学研究所 IDR事務局 (idr@nii.ac.jp)