



Yahoo!知恵袋NII公開データ

- 大規模Q&Aデータを用いた研究に向けて -
ヤフー株式会社 山本 健一

Yahoo!知恵袋とは

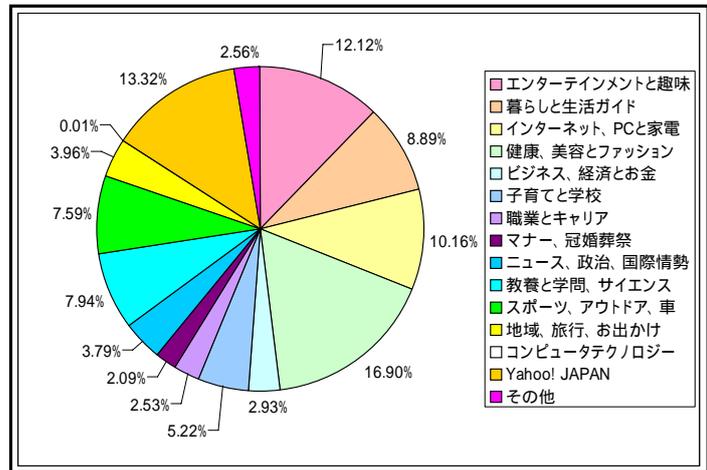
Yahoo!知恵袋は、疑問に思っていることを質問したり、知っている事柄についての質問に回答することで、参加している方がお互いに知恵や知識を教えあい、分かち合えるサービスである。質問したい人と回答したい人を結び、幅広く情報を共有することを目指す。



データの素性

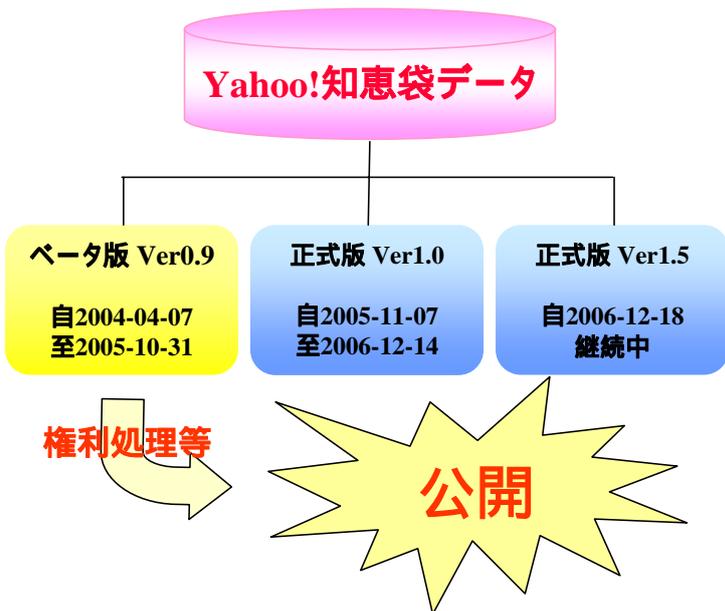
質問数	約300万
回答数	約1300万
文字数	約16億
投稿者数	約24万

多岐にわたる話題



NIIを通してデータを無料公開

Yahoo!知恵袋では、お客様により質問文とそれに対応する回答文が日々蓄積される。これらのデータの一部をNIIを通して広く研究者に公開している。



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...

ワークショップ開催: 11月29日

情報社会学会とヤフー株式会社との共催で、**知識共有コミュニティワークショップ**を開催する。知恵袋データをキーとし、多様な分野の研究者が一堂に会することで、新たなコラボレーションを期待する。

ワークショップCFP

http://www.infosocio.org/cfp_workshop_a2008.html

知恵袋データ入手先

<http://research.nii.ac.jp/tdc/chiebukuro.html>