

2007.03.06
報道発表資料

Web2.0に向けた新しい情報検索技術 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－

東倉洋一・大山敬三

国立情報学研究所

報道発表に主なメッセージ

- NIIはヤフー株式会社と「Yahoo!知恵袋」データの研究利用契約を締結し、研究コミュニティへの無償提供を開始。
- 利用者発信情報の利用による「情報爆発から新しい価値を汲み出す」
e-Science研究の開始。
- NIIによるe-Scienceデータの研究コミュニティへの提供強化の取組み開始。

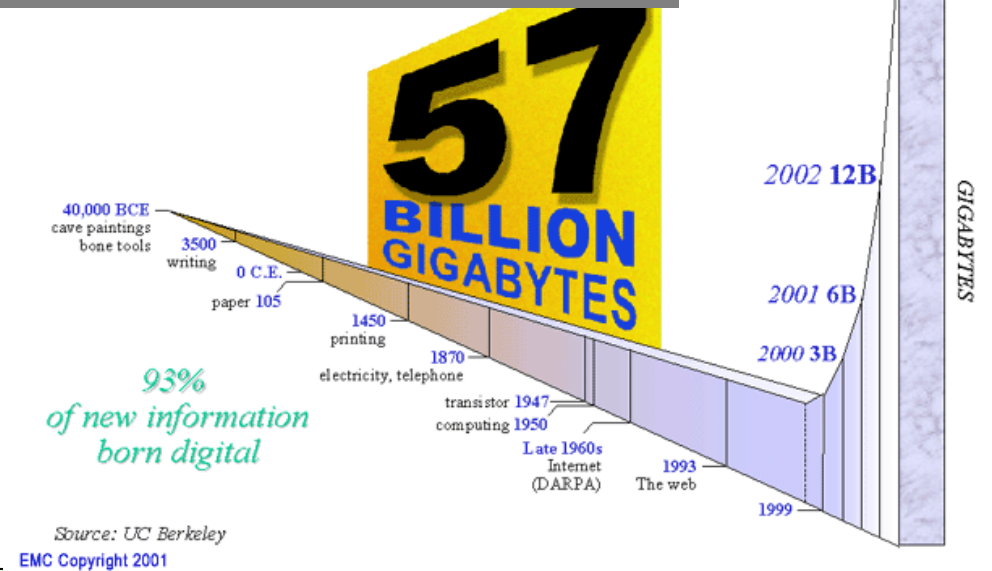
情報爆発時代



受容・処理能力には限界

情報爆発の現状

- 90%以上がデジタル情報
- 今後2年間の情報量は過去の全情報量を超える



情報源:カリフォルニア大学バークレー校
EMC Copyright 2001

情報爆発の課題

- 欲しい情報が探せない
- 人気と重要性が異なる
- マイノリティ情報が埋もれる
- 情報が死蔵される
- 情報の信頼性、信憑性が不明
- 情報の保存が難しい
- 情報保有・提供による社会支配

問題解決に向けた重点プロジェクト (国家施策としての取組み)

- 情報爆発時代に向けた新しいIT基盤技術の研究 (科研費特定領域)
- 情報大航海 経産省プロジェクト
- 情報分析 (情報信憑性) 総務省プロジェクト

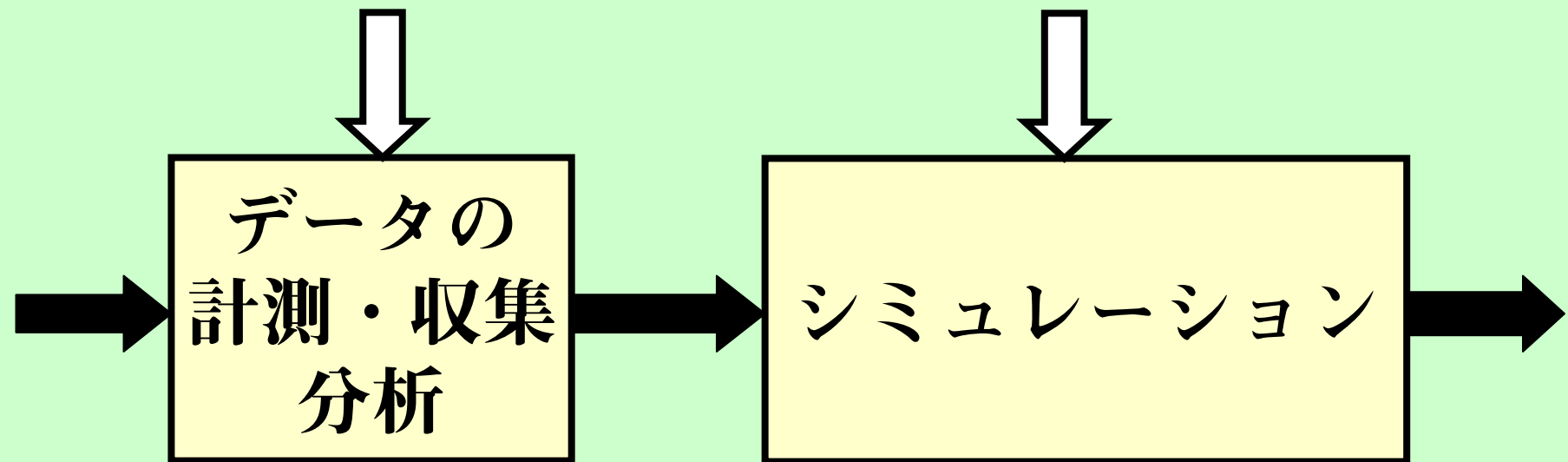
研究アプローチのパラダイムシフト

- 理論研究
- 実験研究
- 計算論的研究
- e-Science研究 (Data-centric研究)

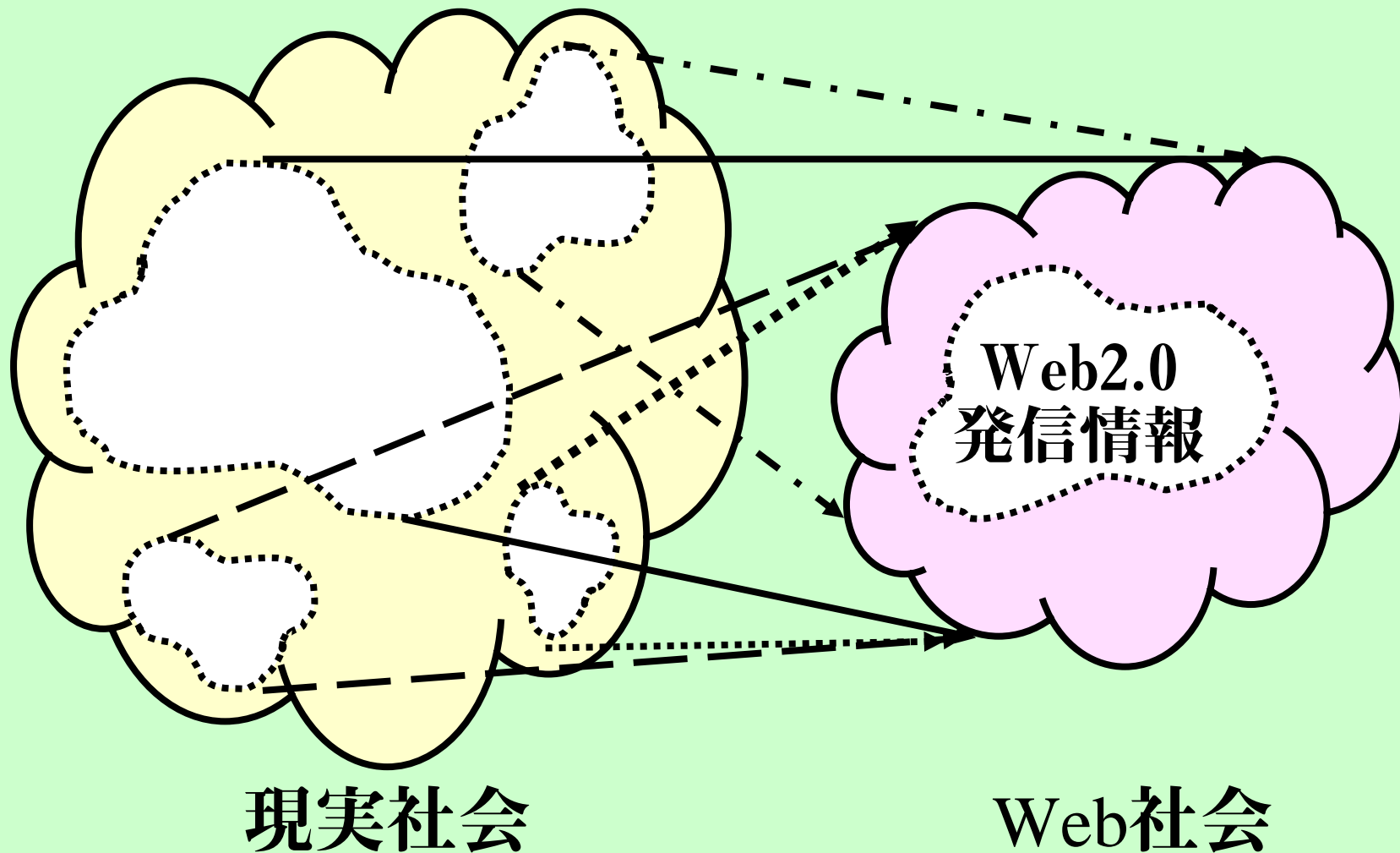
理論、実験、計算論的研究の融合と多様な情報源からの大量情報の獲得・活用により
新しい価値を創成する研究

データの量と質の飛躍的増大

- Web
- 新センサー
- 先端計測装置
- 先端分析装置
- データマイニング
- モデル
- スーパーコンピューティング



Web社会は現実社会の「鏡」

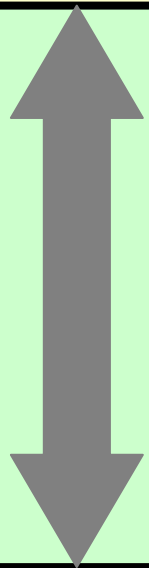


情報の多様性への着目

- テキストから多様な情報メディア
(イメージ、サウンド)
- 画像から映像まで
- 音響、音楽から音声まで
- 言語 (Web言語) の変化と多様性

新しいWeb言語等を研究対象に

書き言葉

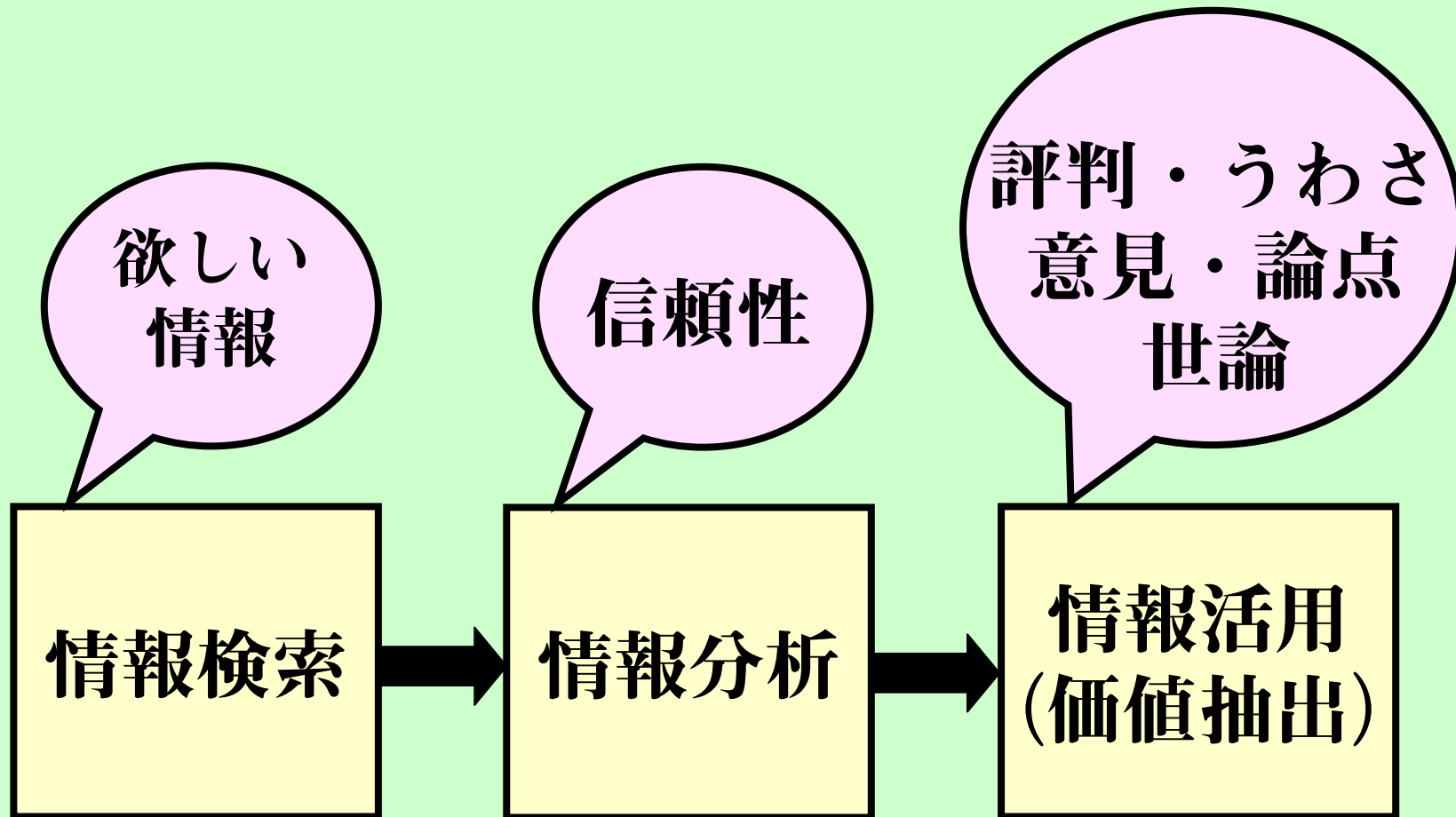


話し言葉

- ニュースサイト、ネット小説、
学術論文、書籍関連データ、
報告書、説明広報資料……
- ネット日記
- **質問・回答文 (Q & A サイト)**
- ブログ
- 電子メール
- 携帯メール

データの研究利用が必須!

情報爆発から新しい価値を汲み出す



利用者発信情報の研究利用

【従来の問題点】

- プロバイダー等の情報サービス提供者の企業情報としてのみ利用され、非公開扱い
- データに含まれる個人情報の壁

【問題解決への第一歩】

- ヤフー株式会社の「Yahoo!知恵袋」の提供による研究コミュニティへの貢献

【今後の方向】

- 多くのプロバイダー等による多様なデータの提供を期待

「Yahoo!知恵袋」とは？

(2004年4月にサービス開始)

- 質問したい人と回答したい人をむすび、知恵と知識を参加者同士で共有
- 日本最大の知識検索サービス
- 質問総数1000万件、回答総数3500万件以上
- 参加登録者約92万人

【提供データ】

質問と回答の本文約16億字。
(質問約311万件、回答約1347万件)

NIIからのデータ提供と利用

- 情報検索、情報分析、情報活用などの研究目的利用者へのデータ提供
(2007.4予定)
- NIIでの情報関連プロジェクトでの利用
- NTCIR:情報検索・アクセス技術の比較と性能評価のための研究基盤関連での利用

今後の課題

- **利用者発信情報（特に、ブログ、Webメールなど）の研究利用の拡大と充実。**
- **新しいWeb言語の解析手法を含む利用者発信情報の分析・活用技術の確立。**
- **多様なe-Scienceデータの研究利用の促進と、情報爆発研究分野におけるe-Science研究による未来価値の創成。**