NII

SBMがつくるコミュニティ
SBMでつくるコミュニティ
– 研究・サービスの両面から –

国立情報学研究所・株式会社グルコース
大向 一輝

NII

もくじ

- SBMと私
- SBM「が」つくるコミュニティ
  - SBMコミュニティの偏りに関する研究
- SBM「で」つくるコミュニティ
  - コミュニティ・ブックマーク4dk

自己紹介

NII

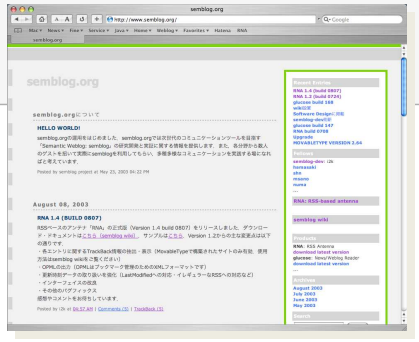
- 大向 一輝（おおむかい いっき）
  - 国立情報学研究所 助教
  - 総合研究大学院大学 助教
    - セマンティックウェブ
    - ソーシャルウェブ
    - 論文検索エンジン「CiNii」
  - 株式会社グルコース取締役（兼業）
    - フィードリーダー「glucose / goo RSSリーダー」
    - mixi station
  - 「ウェブがわかる本」

SBMと私

NII

- SBM研究の源流（2001～）
  - ブックマークエージェント：ブックマークの共有による情報検索の支援 [森00]
    - 最初の論文は1997年
  - Bookmarkからの共通話題ネットワークの発見手法の提案とその評価 [濱崎02]
    - 被験者として手伝う
    - 「誰もブックマークなんか共有しないよ」
      - (´・ω・`)

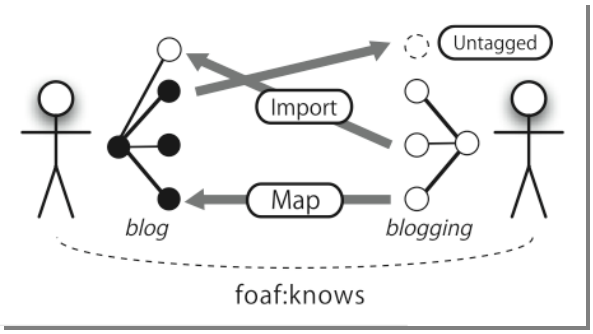
SBMと私



- 未踏時代（2002〜）
  - glucose
  - RNA
    - 日本初（世界初？）のRSS対応アンテナ
      - FOAFでソーシャルネットワーク
      - サイト推薦
      - クリップ
        - フィード中の指定の記事を保存・公開
        - (´・ω・`)

SBMと私

- セマンティックウェブからのアプローチ（2004〜）
  - A Proposal of Community-based Folksonomy with RDF Metadata [Ohmukai05]



  - Ontologies are us: A unified model of social networks and semantics [Mika05]
    - (´・ω・`)

National Institute of Informatics, Tokyo, Japan

SBMと私

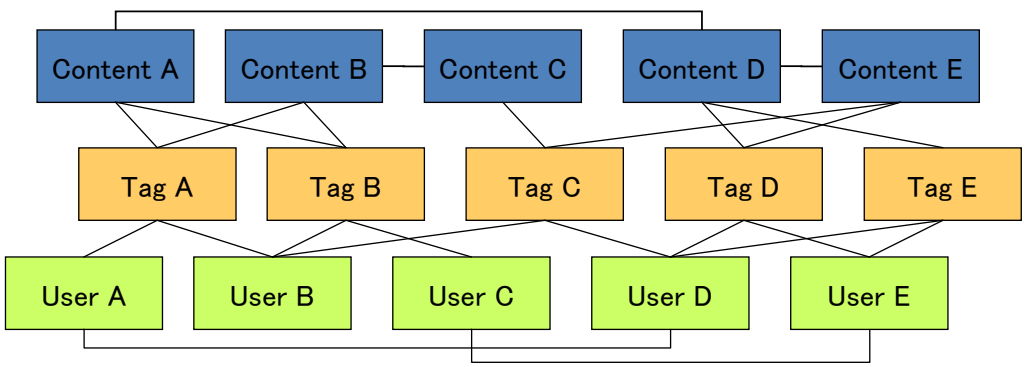
NII

- 私にとってSBMは敗北の歴史
- なのに、まだやっている
- ここから見える風景について、お話しします
  - 研究面から
  - 開発面から

なぜSBMを研究するのか

NII

- 公式回答
  - SBMは3つのネットワークが交わる場所
    - 情報：ウェブ
    - 知識：フォークソノミー・オントロジー
    - ひと：ソーシャルグラフ



なぜSBMを研究するのか

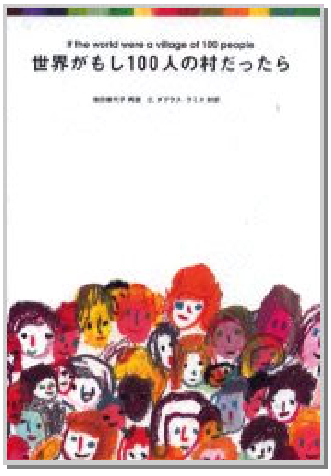
NII

- 非公式回答
  - どうしてこんなに偏っているのだろう？
    - 「なんとか村（？）」
    - 民俗学的興味
  - 巨大コミュニティとしてのSBM
    - 「意図せざる協働」[深見07] の結果
    - データからコミュニティの偏りを理解したい
      - 手がかりとして3つのネットワークを使う

偏りを知るために

NII

- 「全人類の興味・関心の平均」との距離を測る
  - これは無理
- 「偏りがないと意味をなさない仕組み」の有効性を検証する
  - 背理法的に
- できれば人様に役に立つような形で
  - 理解と推薦



研究のアプローチ

NII

- 既存の数理モデルをあてはめるのが目的ではない
- フィールドワーク的
  - コミュニティにどっぷりつかる
  - シンプルな仮説を作る
  - 仮説の検証・改良
- 高度さ・複雑さはあとまわし

データセット

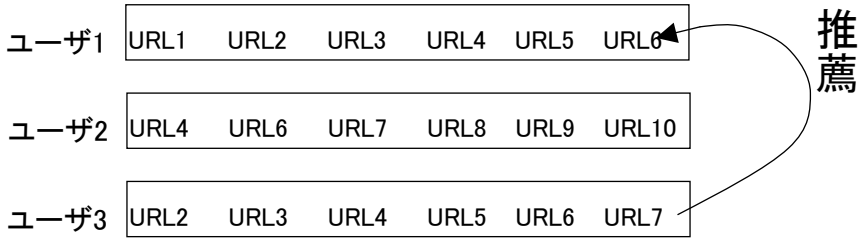
NII

- ECナビ様のご提供
  - Buzzurl（当時はECナビ人気ニュース）
  - 期間：2005/10/25〜2006/12/5（407日間）
  - ユーザ数：7,579
  - ブックマーク数：139,602
  - ユニークURI：81,123
  - 備考
    - ユーザは匿名化（ただしIDあり）
    - タイムスタンプあり

SBMと推薦

NII

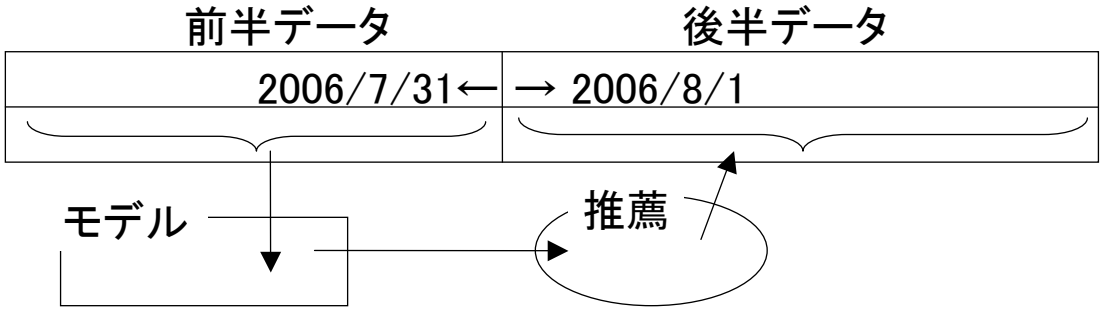
- 内容ベース
  - 任意の2アイテムの類似度を算出し、高いものを推薦する
  - アイテムの内容だけで推薦を行うことができる（人が介在しない）
- 協調フィルタリング
  - 同じアイテムを好むユーザを見つけ、そのユーザが好む他のアイテムを推薦する
  - ユーザの嗜好データが事前に必要となる（コールドスタート）
- とりあえず協調フィルタリングを採用してみる



予備実験

NII

- シミュレーション
  - 前半のデータで協調フィルタリングのモデルを作成
  - 後半のデータとの一致を見る
  - 実際の推薦ではないことに注意



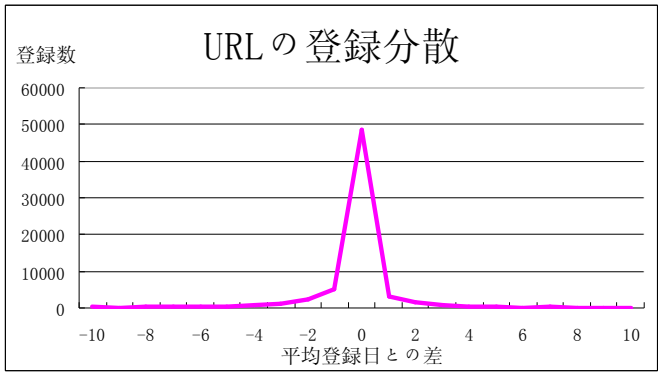
- 精度 (precision) = 適合数 / 推薦数
- 再現率 (recall) = 適合数 / 検証数
- F1値 = 2 * 精度 * 再現率 / (精度 + 再現率)

予備実験

NII

- まったく効果なし
  - F1＝0.04%
- SBMに登録されるデータは常に変化する
  - ほぼ1〜2日の間に登録が終わる
  - 前半のデータを後半で見せても無駄
  - 「人気ニュース」だから？

ユーザ数	597
適合数	21
推薦数	59,700
検証URL数	33,876
精度	0.04%
再現率	0.06%
F1値	0.04%



SBMのための協調フィルタリング

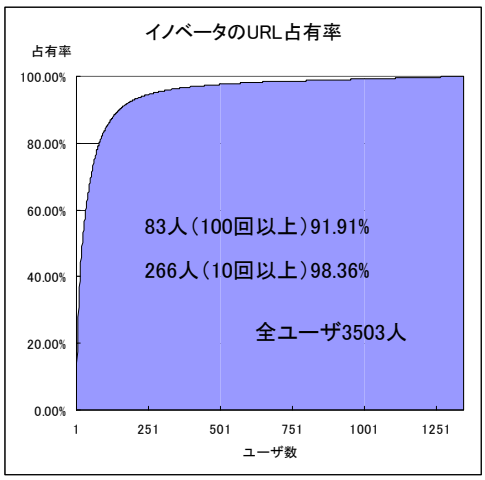
NII

- 最新のコンテンツを推薦するために
  - 富豪的アプローチ
    - リアルタイムに新規のデータを取り込み、モデルを再計算
    - 膨大な計算コスト
  - 「偏り」仮説
    - イノベータとフォロワーの存在
      - いち早くコンテンツをブックマークする人々
      - イノベータについていく人々
      - どちらも興味・関心は変わらない
    - イノベータの情報をフォロワーに配信
      - 推薦の対象をコンテンツからユーザに転換

SBMのイノベータ

NII

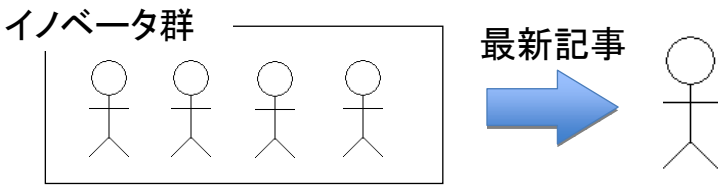
- イノベータ層が実際に存在するのかどうかを検証
- 究極の簡略化：1ゲッター
  - すべてのURIについて「1さん」だけを抽出
  - 1ゲッター集団の占有率
    - 100回以上：83人：91.9%
    - 10回以上：266人：98.4%
    - 全ユーザ：3503人
  - 「スクープ」の影響？



イノベータからの推薦

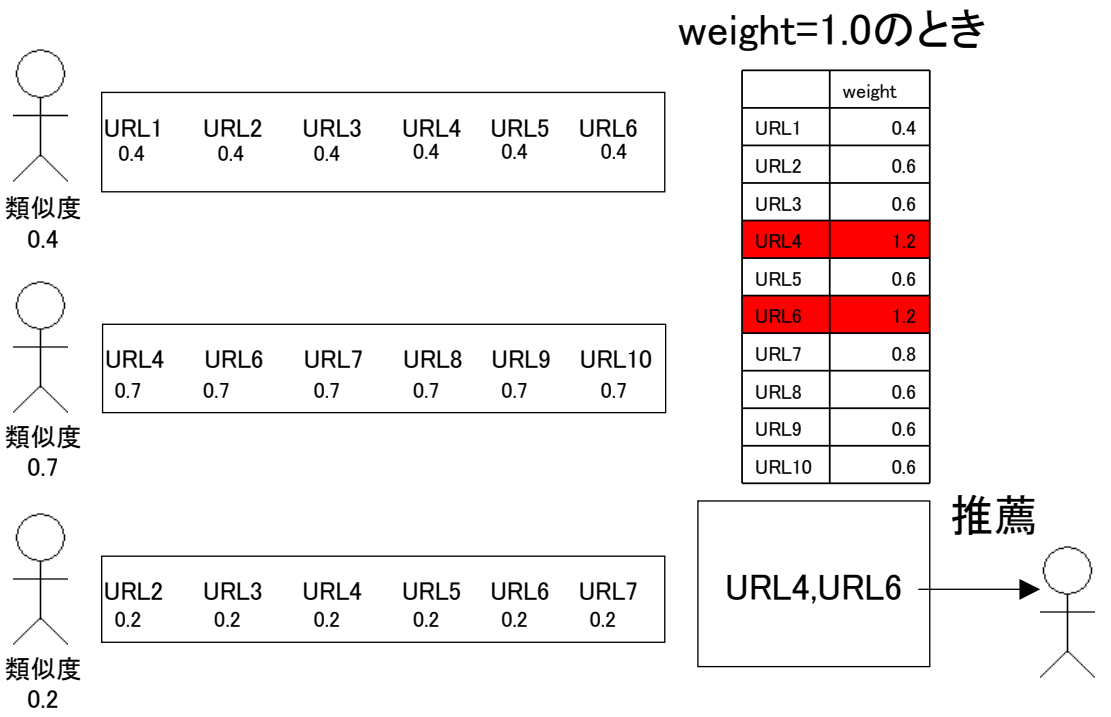
NII

- シミュレーション
  - 前半データ
    - 各ユーザとイノベータ群の類似度を計算
  - 後半データ
    - 類似度の高いイノベータ群がブックマークしたコンテンツのうち、各ユーザがどの程度受理したかを測定
  - コンテンツの推薦基準
    - 類似度上位n人のイノベータのコンテンツのみ
    - 類似度x以上のイノベータのコンテンツのみ
    - 各コンテンツにイノベータの類似度で重みづけし、積算



イノベータからの推薦

NII



シミュレーション結果

NII

- F1値：7〜8%
  - 精度と再現率はトレードオフ（閾値によって変化）
  - 実際の推薦問題ではない
- 計算量の激減
  - 協調フィルタリングは1回走らせるだけ
    - 期間中にモデルを再構築しても結果はほぼ同じ
    - 嗜好は変化しない

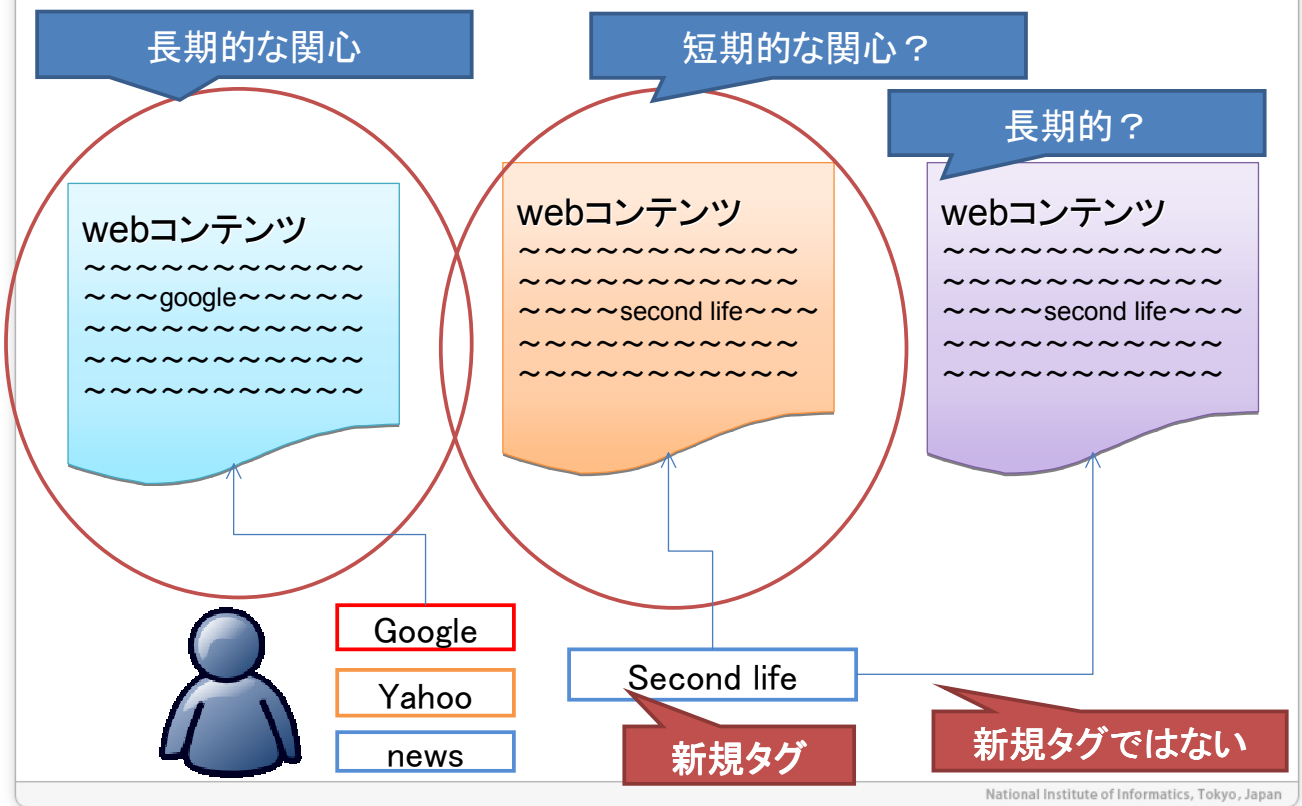
weight	0.2	0.3	0.4	0.5	0.6
ユーザ数	597	597	597	597	597
有効ユーザ数	483	483	483	483	483
適合数	838	561	396	297	231
推薦数	30932	11765	5547	3005	1771
検証URL数	4267	4267	4267	4267	4267
精度	2.71%	4.77%	7.14%	9.88%	13.04%
再現率	19.64%	13.15%	9.28%	6.96%	5.41%
F1値	4.76%	7.00%	8.07%	8.17%	7.65%

興味・関心の偏り

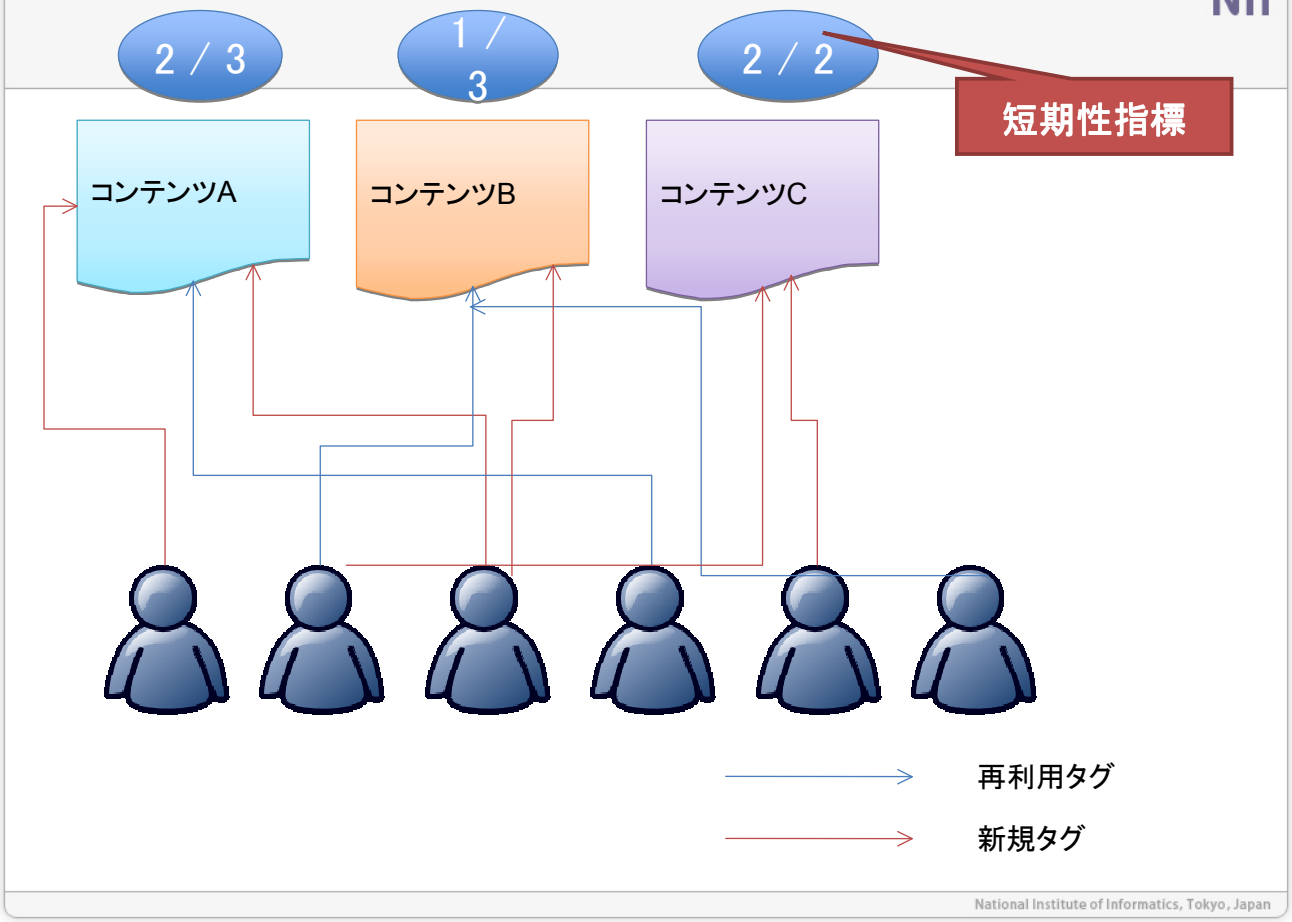
NII

- SBMがつくるコミュニティには長期的な興味・関心がある
  - ブックマーク数の多寡では判断できない
- 各コンテンツが長期的関心に合致するかどうかを判断する
  - 個々のユーザのタグ付与行動履歴を観察
  - 総体として表れるコミュニティの嗜好

NII



National Institute of Informatics, Tokyo, Japan



コミュニティの興味・関心

- 「短期性指標」下位のコンテンツ
  - コンテンツ・タグともにウェブ系が大半を占める

「短期性指標」上位のコンテンツ

  - コンテンツ・タグともに統一性なし

コンテンツタイトル	新規性指数	主なタグ
ITmedia:Yahoo!,eBayの提携がMicrosoftに及ぼす圧力	0	Microsoft,Yahoo!,Business
CNET:映画ダウンロードの先のAppleとGoogleの関係	0.059	動画,Apple,Google
ITmedia:ヤフーがCGM化の大号令MySpaceは「連携も」	0.065	Yahoo!,CGM,MySpace
CNET:「ECナビ人気ニュース」に新機能を追加	0.071	News,CNET,ECナビ
MYCOM:代表者に聞く「キャラホビ」今年の動向	0.075	ガンダム,キャラ,ホビ

コンテンツタイトル	新規性指数	主なタグ
ブクマで夢をかなえよう!プレゼントキャンペーン	0.90	NitendoDS旅行,iPod,iMac
百式 － ネガティブ目標（Joes Goals.com）	0.87	目標,酒,タバコ,管理
山口もえ感涙！ほのぼの披露宴	0.85	山口もえ,結婚式,おめでとう
ECナビ人気ニュースβ版/ニュースに特化したSBM	0.79	ECナビ人気ニュース,タグ,祝
asahi.com：アドウェイズが上場 上場企業最年少	0.78	IPO,アドウェイズ,最年少

タグの偏り

NII

- タグの主観性・客観性 [Golder06]
  - コンテンツの主題（i.e. Google）
  - コンテンツの種類（i.e. Book）
  - コンテンツの作成者（i.e. ITmedia）
  - 分類のための記号（i.e. ☆☆☆）
  - 意見・感情（i.e. これはすごい）
  - コンテンツとユーザの関係
  - ユーザのタスク（i.e. あとで読む）

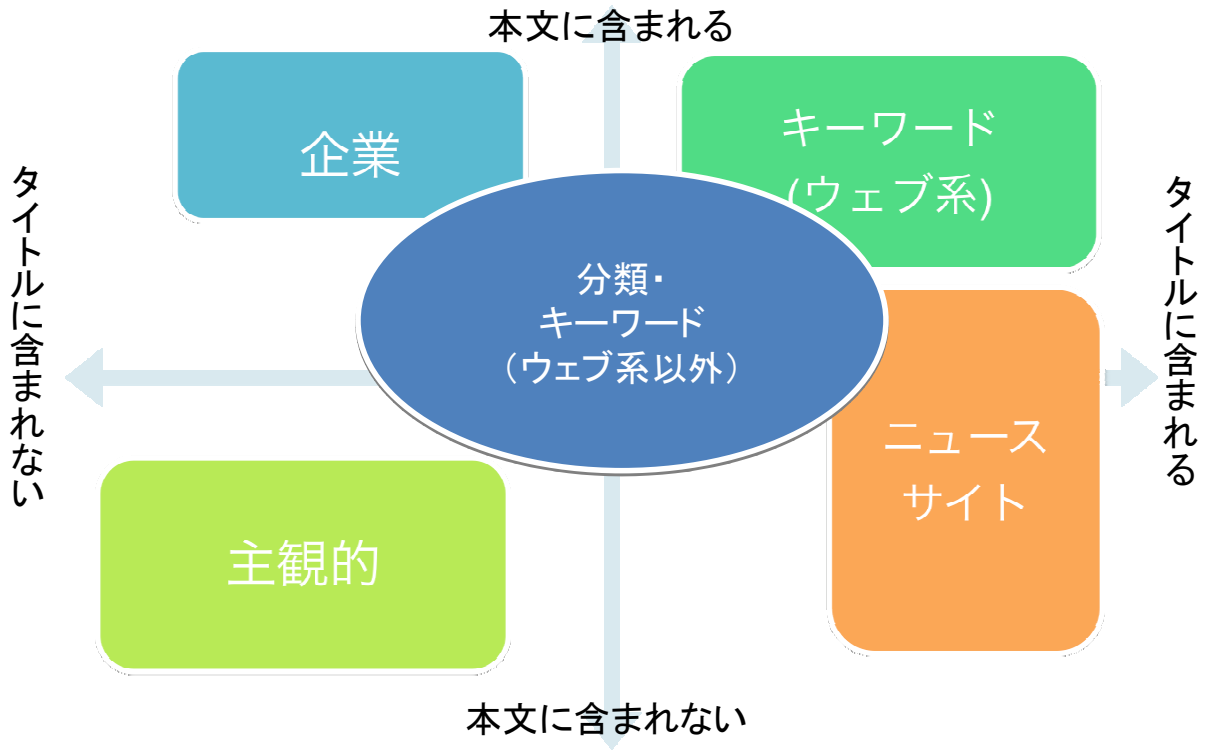
客観的

主観的

- タグの文字列がタイトル・本文に含まれるかどうか「だけ」で判定してみる

タグの偏り

NII



National Institute of Informatics, Tokyo, Japan

ここまでのまとめ

NII

- SBMコミュニティに避けがたく存在する「偏り」
  - 人・タグ・情報
  - 内部から「偏り」の姿を見たい
  - 「偏り」を積極的に活用したい
- シンプルなアプローチ
  - メタデータの分析のみ
  - 計算量を抑えて実用化しやすく

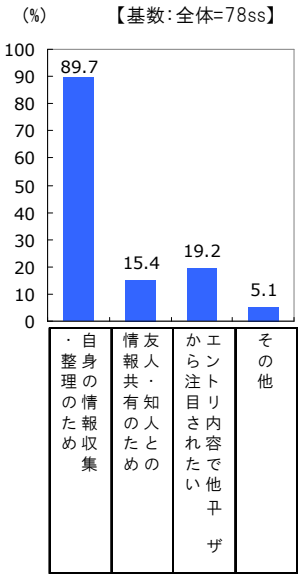
何のためのSBM？

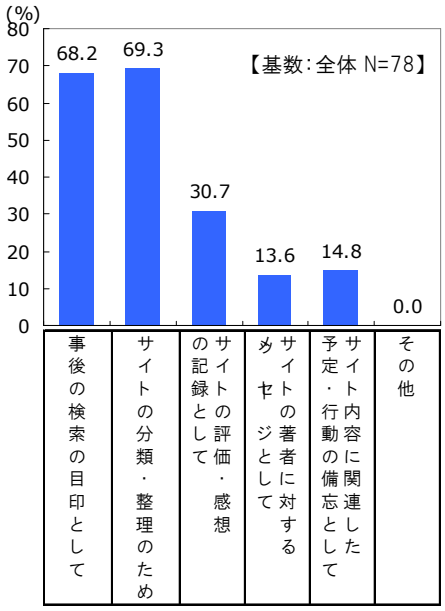
NII

- SBMが提供する機能
  - 個人向け
    - 保存・整理
    - 投票
    - コメント
    - 表現（スクラップブック）
  - 集合知として
    - 未知との遭遇
    - 外部評価

SBMに期待されるもの

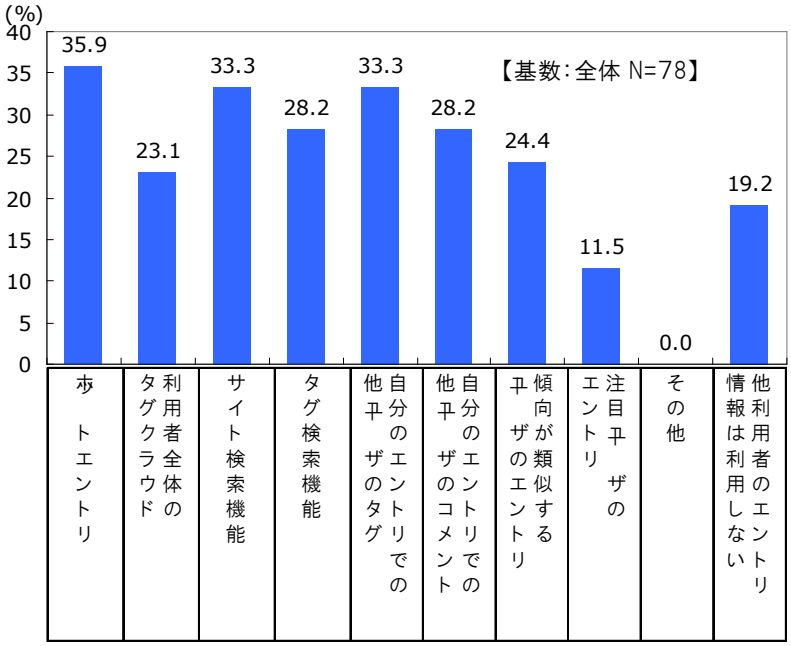
- SBMユーザに対する質問紙調査 [深見07]
  - 有効回答数：78
  - SBM・タグの使用法





SBMに期待されるもの

- 他のユーザが投稿した情報の利用



SBMの現在

NII

- 多用途に耐えるシステム
  - 使いこなす気持ちよさ(?)
  - アナロジーで説明できなくなっている
- 機能分化の可能性
  - かつてブログ・SNSが通ってきた道
  - digg・tumblr…
- 新しい対象・新しい用途を考える

現実のコミュニティ

NII

- 現実のコミュニティに目を向けてみる
  - 研究室・職場・家など
    - クローズ・半クローズなコミュニティに囲まれている
    - 情報共有は切実な問題
  - FYIは「絶対読め」の意味だが…
    - メール:一方向的(反応がない)
    - イントラ:コミュニティは1つではない
  - ほしいもの
    - 双方向的・複数コミュニティ対応
    - (力づくでも)読ませるツール

4dk

NII

- コミュニティ・ブックマーク
  - http://4dk.jp/
- FYIメールを現代化する
  - 宛先のある情報のやりとり
    - 重要な情報＝伝えたい情報
  - 宛先＝コミュニティ
    - コミュニティは1つではない
  - 読まれたかどうかがメタ情報となる
    - リアルのコミュニケーションとの連動
- 名前の由来
  - 「読んどけ」

4dkの機能

NII

- 複数グループの作成・参加
- ダッシュボード
  - すべてのグループの状況を一元管理・フィルタ
- 視聴率
  - グループ内で何人が読んだのか
- 朝刊メール
  - 前日のキャッチアップ
- 夕刊メール
  - 他のグループからのおすすめ
- 検索
  - 「誰が投稿・コメントしたのか」で思い出す

4dkデモ

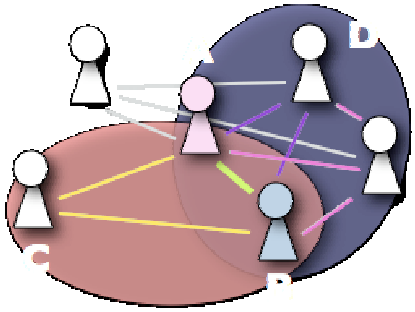
運用結果

- 2ヶ月ほど運用してみて
  - 既存のSBMとは異なる利用モデル
    - 圧倒的にクローズドが多い
    - 1人グループ・2人グループの存在
- ちょっと反省
  - まだまだ欲張りすぎ
    - 16通りのグループ設定
  - Ajax使いすぎ
    - ユーザビリティ改善

本音と課題

NII

- ソーシャルってなんだろう
  - 「不特定多数と私」だけではないはず
  - 「私を取り巻くコミュニティ」の集まり
- 切実な情報共有
  - 「読んでもらわないと困る」と「あとで読みたい」は質的に異なる
- 課題
  - コミュニティの本質的な面倒さ
    - ソーシャルグラフ作り
    - グループ設定
  - 既存のコミュニティとの連携



本当のまとめ

NII

- SBMがつくるコミュニティ
  - 偏りの理解と応用に関する研究
- SBMでつくるコミュニティ
  - 情報を伝えることに特化したブックマークサービス

ありがとうございました！

NII

- 大向 一輝（おおむかい いっき）
- i2k {at} nii.ac.jp / i2k {at} glucose.jp
- http://d.hatena.ne.jp/i2k
- http://4dk.jp/i2k

参考文献

NII

- 森幹彦, 山田誠二: ブックマークエージェント : ブックマークの共有による情報検索の支援, 電子情報通信学会論文誌, Vol.J83-D-I, No.5, pp.487-494 (2000).
- 濱崎雅弘, 武田英明, 松塚健, 谷口雄一郎, 河野恭之, 木戸出正継: Bookmarkからの共通話題ネットワークの発見手法の提案とその評価, 人工知能学会論文誌, Vol.17, No.3, pp.176-284 (2002).
- I. Ohmukai, H. Takeda, M. Hamasaki, K. Numa and S. Adachi: Metadata-Driven Personal Knowledge Publishing, The Semantic Web - ISWC 2004: Third International Semantic Web Conference, LNCS 3298, pp.591-604 (2004).
- P. Mika: Ontologies are us: A unified model of social networks and semantics, Journal of Web Semantics, Vol.5, No.1, pp.5-15 (2007).
- 深見嘉明, 國領二郎: 意図せざる協働 ―ソーシャルブックマークにおけるボトムアップメタデータ生成による情報 共有―, 情報社会学会誌, Vol.2, No.2, pp.6-19 (2007).
- 大石剛司, 亀田尭宙, 深見嘉明, 大向一輝, 武田英明: ソーシャルブックマークにおけるユーザのタグ付与行動分析に基づくコンテンツ分類, 情報処理学会第70回全国大会講演論文集 (2008).
- 大力慶祐, 大向一輝, 武田英明: ソーシャルブックマークにおけるイノベータに注目した情報推薦手法の提案, 人工知能学会全国大会(第22回)論文集 (2008).