

ソーシャルブックマークサービスにおける アノテーション情報の機能分析

Characteristics of annotated information in social bookmark services

深見 嘉明*1

FUKAMI Yoshiaki

*1 慶應義塾大学大学院 政策・メディア研究科

Keio University Graduate School of Media and Governance

To handle vast amount of information stored or transferred on the internet, there have been many efforts to organize them. Social tagging at social bookmark services is bottom-up approach to annotate metadata to divide web contents into classes. In this research, I examine the possibility of social tagging to build bottom-up taxonomy, in other words, folksonomy. We find users attempt to align usage of vocabulary for tag to provide utility to search information each other. Stored data shows vocabulary are consolidated to some extent without any control. There is a possibility to form folksonomy, contributed by function of tagging support and tools for communication among users.

1. はじめに

常時接続環境の普及やウェブログなどユーザが容易にコンテンツをアップロードできるツールの開発により、ウェブに蓄積・流通する情報は増加する一方である。それらを有機的に結びつけ、効率的に有効に活用するために、ウェブ上の情報を構造化する試みが各方面で行われている。

大量の情報に対する構造化の試みとして、古くから行われてきたものに図書分類がある。図書分類では出版物に対し、日本十進分類法など定められた基準にのっとり、図書館員・司書といった専門家が分類コード、つまりメタデータを与えている。しかし、図書よりも遥かに膨大な数のドキュメントが存在するウェブコンテンツの構造化を、トップダウンで基準を設け、限定された人数の専門家が行うことは非現実的であり、メタデータ付与とタスクの分散化が必要となる。

最近分散型メタデータ生成事例として、ソーシャルブックマークサービス（以下SBM）が注目されている。SBMではサイト登録時に、タグと呼ばれるキーワードや1行程のコメントを付与することができ、登録内容は原則的に公開される。タグはユーザが登録コンテンツの分類に活用しているとすると、ユーザが協働してメタデータを生成していると捉えることができる。

ユーザ全体によるタグ付けはソーシャルタギング、その結果として生成されたタグ集合が表す分類体系はFolksonomy[Mathes 04]と呼ばれている。しかしSBMにおける個人の分類は、個人個人それぞれのルールに従って行われているため、何ら統制されずに付与されたタグが、情報を構造化するに足りるのか危惧される。

本論文ではSBMのユーザに対しアンケート調査を実施し、プラットフォーム上での協働意識がどの程度存在するのか、またソーシャルタギングによって生み出されたタグ集合が構造化された分類体系を表し得るのかについて分析した。またSBMに蓄積されたデータを解析し、ボトムアップ分類で生じるとされる問題点がどの程度生じているのかについて検証し、Folksonomyという概念が現実的であるのかについて考察した。

2. ソーシャルタギングで生成されたタグ集合の特徴

前節で述べたとおり、ソーシャルタギングは誰からも統制されないままなされる。そのため、生成されるタグ集合は以下に挙げた3つの特徴を有している [Golder 05]。

- polysemy:多義性
- synonymy:同義性・同義語の存在
- basic level variation:基礎水準の混在

多義性とは、同一の単語を各ユーザが異なる意味合いで用いている場合があるということである。例えば“阪神”という単語は、プロ野球についての話題で用いられる際は“阪神タイガース”という球団を意味し、交通手段について話していれば“阪神電鉄”という大阪梅田・神戸三宮間を結ぶ私鉄路線を指すであろう。

SBMでは、同一語彙のタグは自動的にリンクが張られまとめられるが、多義性により異なるものを指すものがまとめられることがある。

同義性・同義語の存在とは、同じ内容のものを指す複数の語彙があることを指す。例えば、日本のプロ野球について論じているときに“阪神”、“タイガース”、“猛虎”という単語は、全て阪神タイガースというセントラルリーグに所属するチームを指すであろう。

基礎水準とは認知言語学の用語であり、概念分類（カテゴリ）における中心部分の個人差を指す。基礎水準の混在とは、世界の認知・解釈の際に基準とするレベルが異なるということである。例えば、ボーナスで購入したいものを夫婦で話し合うときに、夫は車を買いたいと思っており、妻は薄型テレビの購入を考えている。夫は「車かテレビか」という選択肢から車を選び取りたいと考えているのに対し、妻は既にテレビの購入は既定事項と考え、液晶テレビかプラズマテレビかについて話し合いたいと思っている。この夫婦の話し合いは当然かみ合わないものとなる。

このような主眼となる水準の差が世界の認知レベルにおいて起こることが基礎水準の混在である。タグ集合がサイトの分類体系を反映していると考えられるなら、基礎水準の混在は体系が整理されていないことを意味する。

いずれも Folksonomy が現実のものとして機能するには大きな障害だが、特に同義語は Folksonomy 成立以前に、蓄積情報の共有に対する障害要因となる。SBM においてタグは情報検索の最も重要なキーとなるからである。タグは検索クエリとして重要な要素である。またタグを付与された情報は自動的にまとめられ、同一タグが付与された一覧ページが生成される。タグは類似情報をまとめる機能を持ち、同一語彙のタグが付与されることは情報間の関連性を表す。つまりタグは蓄積情報を共有し活用する上で重要な役割を担っているのである。

しかしタグはシステムによって自動的に判別され、リンクが生成される。具体的には、Google という企業については”Google”、”google”、”GOOGLE”、”グーグル”とさまざまな表記をされる。タグは機械的に処理されるため、これら4つの語彙はそれぞれ別のもので扱われる。つまり”Google”とタグ付けされた情報と”グーグル”とタグ付けされた情報はリンクされず、それぞれのタグが付与されたサイトはまとめられることはなく、ユーザが両方の情報を参照するにはタグのバリエーションに気づいた上でそれぞれのリンクページにアクセスする必要がある。

3. ソーシャルタギングによる構造化の実効性に対する検証

3.1 調査設計のポイント

前節で述べたように、タグ付けをする際にユーザが同一概念に対し異なる語彙を用いると、自動リンク生成機能がうまく機能せず、分類体系が成立し得ないどころか蓄積情報共有の障害要因となる。多く SBM では情報登録時に、他のユーザが用いた語彙を候補を表示するなど、タグ付け補助機能が実装されており、これが活用されていれば自然と語彙は統制されていくであろう。またユーザの間に情報を共有するという意識があれば、できるだけ他ユーザと使用語彙をあわせていこうという意識が出てくるであろう。

本論文では SBM ユーザにアンケート調査を実施し、サービスに対しロイヤリティ（帰属心）がどの程度生成されているか、またタグ付け時に使用語彙を他ユーザとあわせていこうとする意識がどの程度あるのかについての実態把握を行った。同時に SBM に蓄積された実データを解析することで実際にタグとして用いられる語彙のバラつきを検証した。

3.2 対象とするサービス

調査対象としては株式会社 EC ナビが運営する Buzzurl(図1)*1を選んだ。SBM は個人の情報収集・整理・分類のためのツールとしての役割と、情報蓄積やアノテーションによるコミュニケーションツールとしての側面がある。

Buzzurl は、ニュース情報について共有するというコンセプトで設計されている SBM であり、他のサービスと比較して明確に”情報共有”のための場としての性格が示されている。具体的には、同じ傾向のサイトを蓄積しているユーザを推薦する”シンクロユーザ”機能や、他のユーザがサイトに付与したコメントに対しレスがつけられるなど、ユーザ間インタラクションの発生を支援する機能が実装されている。

Folksonomy の成立可能性について検証するにあたり、ユーザ間インタラクションの発生は前提となると考え、特にコミュニケーション支援機能が充実している Buzzurl を分析対象として選定した。



図 1: Buzzurl トップページ

3.3 Buzzurl のタグ付け補助機能

前述のとおり、ユーザの語彙選定に大きな影響を与えるのがタグ付け補助機能である。Buzzurl ではサイトを登録する際、UI はタグに用いる語彙の候補として同一サイトに対し他のユーザが用いた語彙のリストと、自身が過去に用いた語彙のリストが表示され、選択することができる(図2)。

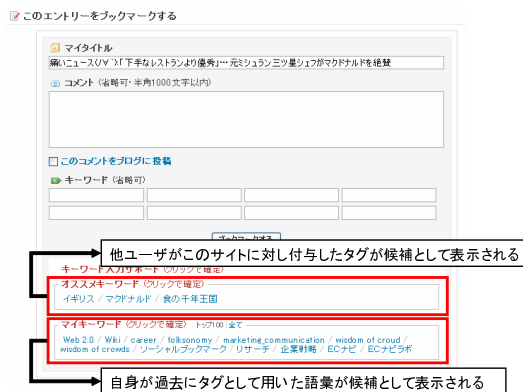


図 2: Buzzurl のタグ付け補助機能

4. アンケート調査から見るタグ付け時の意識

4.1 調査設計

調査は、Buzzurl においてサイトの登録時にタグ付けを行っているアクティブユーザに対しメールでアンケートを依頼した。アクティブユーザの定義は以下のとおりである。

- いずれかのサービスで直近 1 ヶ月以内にレビューを書いている
- いずれかのサービスで合計 10 以上のエントリを登録し、かつ合計 10 以上のタグを付与している

アンケート回収はウェブフォーム上にて実施、78ss を回収した。設問内容は、日常的に利用しているメディア、SBM を含むウェブサービス利用状況、SBM 利用時の意識、SBM に蓄積された情報の活用状況等である。本論文では、タグ付け時の意識や利用サービスに対する考え方に関する設問に絞って結果を紹介する。

*1 Buzzurl: <http://buzzurl.jp/>、データ提供時のサービス名称は EC ナビ人気ニュース。

4.2 対象者の基本属性

性別は男女共に50%と半数ずつである。年代は男性で30代(男性の64%)に集中し、女性は20代(21%),30代(44%),40代(31%)に広がっている。職業は社会人が全体の6割を占めている。他は学生が1割,専業主婦とパートアルバイトがそれぞれ15%程度である。

4.3 SBM 利用理由

Buzzurlは実際にコミュニケーションプラットフォームとして捉えているのであろうか。SBM利用理由に対する設問で、コミュニケーションツールとして捉えているとする回答は2割以下であった(図3)。

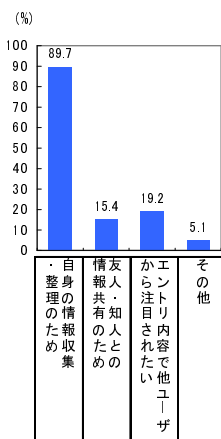


図 3: SBM ユーザアンケート:SBM 利用理由

4.4 SBM の捉え方

その一方、主利用サービスに対する考えについての設問に対する回答では(図4)、約4割の対象者が自分が登録したサイトがランキング上位にくるとうれしい」と、28%が他利用者の参考となるよう、投稿内容を考えると回答している。他利用者からの評価を気にしたり、他利用者の役に立つような形で利用しようとする意識を持つユーザが3,4割程度存在しているのがわかる。一方、自身の登録内容に対し付与タグ・コメントが気になるとするのが26%、常に参考にして利用している利用者がいるのが13%と、他ユーザが付加した情報に興味関心を持つという意見も見られる。

自分が情報を蓄積する際は他利用者が閲覧・利用することを想定し、また他利用者の蓄積内容も興味関心をもって確認する層が一定数存在するという事は、ユーザが互いの存在を意識し、その上で情報の発受信を行っているということになる。意識には明確に立ち現れてこないものの、各人個別の情報管理作業がコミュニケーションとして機能しているといえるだろう。

結果的に、利用サービスに愛着を覚え(46%)、他利用者に共感(36%)、趣味嗜好があう利用者が多いという回答が2割強ある。つまり、利用のきっかけや主目的はあくまでも情報整理のためのツールとして活用するためだが、蓄積された情報が全て公開されているという前提のため、他ユーザが蓄積した情報を日常的に活用することで、「お互い様」という相互扶助的な感覚が芽生え、他者にも役立つようなアノテーションのスタイルとなっていくようである。

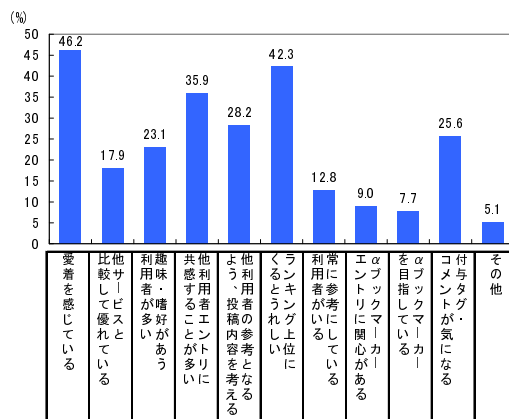


図 4: SBM ユーザアンケート:主利用 SBM に対する考え

4.5 タグ付け時の語彙選定

ではタグ付け時に協調的な行動はされているのであろうか。タグ付け時の語彙の選び方としてできるだけ他ユーザにあわせるという回答は26%、時には場合によるという回答を含めると実に76%が他ユーザの使用語彙を気にすると回答している。つまりかなりの割合で語彙選定に関する協調意識が見られる。では登録されたデータで、実際に語彙は収斂しているのだろうか。

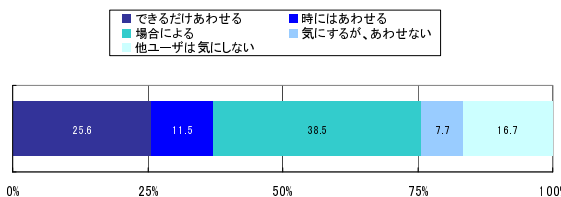


図 5: SBM ユーザアンケート:タグに用いる語彙の選定意識

5. 蓄積データから見るタグ付け時使用語彙の実態

5.1 Buzzurl のデータ

今回分析に用いたデータは2005年10月~2006年12月までに登録されたブックマークデータである。なお当該期間、サイトの名称はECナビ人気ニュースであった。当該期間にサイトを登録したユーザは1,221人、登録されたURLは74,128種類、のべ登録数は139,602であった。139,602のブックマークに対し264,058個のタグが付与されている。用いられた語彙は31,076種類、1ブックマークあたりの平均付与タグ数は1.9であった。

なお今回の分析では"Google", "google"のように、人間が読めば明らかに同一対象を指すと理解できてもシステム上別の語彙として取扱われるものは別の語彙として扱う*2。

*2 システムにより大文字・小文字や全角・半角などの取り扱いは異なる。そうした仕様の影響を排除するため、今回は文字コード上異なるものは別のものとしてカウントした。

5.2 語彙の分布

表 1 に使用頻度上位 10 語彙と類義語の使用数を挙げた。1 位の”Google”は他に”google”など 6 つの同義語が存在するが、”Google”の使用頻度が飛びぬけている。類義語では”google”の使用頻度が高いが”Google”と比較すると 3 分の 1 以下であり、他は 10 分の 1 にも満たない。また、5 位”画像”、7 位”芸能”は明確な類義語が存在せず、他の順位でも一つの文字列形態に頻度が集中しているのが見てとれる。アンケート調査結果に現れた、ユーザの意識と実装されているタグ補助機能の影響か、使用語彙は自然と収斂しているようである(表 1)*3。

| 順位 | 語彙 | 類義語 | 度数 |
|----|--------|------------|------|
| 1 | Google | Google | 3078 |
| | | google | 893 |
| | | Go o g l e | 11 |
| | | GOOGLE | 2 |
| | | グーグル | 111 |
| | | 米 Google | 5 |
| | | 米グーグル | 2 |
| 2 | ネタ | ネタ | 3038 |
| | | ネタ | 1 |
| 3 | おもしろ | オモシロ | 2491 |
| | | おもしろ | 1 |
| 4 | プロレス | レスリング | 2419 |
| | | レスリング | 4 |
| 5 | 画像 | | 2277 |
| 6 | SNS | SNS | 1955 |
| | | sns | 75 |
| | | sns | 56 |
| 7 | 芸能 | music | 1773 |
| | | Music | 13 |
| 8 | サッカー | サッカー | 1685 |
| | | football | 47 |
| | | 音楽 | 1593 |
| 9 | 広告 | ad | 391 |
| | | Ad | 1 |
| | | AD | 1 |
| | | AD | 1 |

表 1: Buzzurl, タグとしての使用頻度上位 10 語彙と類義語

5.3 使用語彙の時系列分析

上位の中でも比較的類義語の登録数が多い 1 位の”Google”と 6 位の”SNS”については月ごとの使用回数推移を分析した。”Google”は当初は半角小文字の”google”の使用回数の方が多かったが、2006 年 4 月に逆転して以降正式名称である G が大文字の”Google”が多数を占めるようになり、一度定まった傾向は固定化されるようである(図 6)。”SNS”は当初から半角大文字の”SNS”が多数を占め、その傾向は変わることがなかった(図 7)。

分析事例数が少ないため断定はできないものの、一度多数を占めた語彙はその後ますます採用されるといって、ログイン現象が生じている可能性がある。つまりユーザコミュニティ内で、自発的に語彙が統制されているのではないと思われる。

6. 結論

アンケートから SBM ユーザは、利用するサービスに対しロイヤリティを形成しているとともに、ユーザ間の協調行動志向が生まれ、タグ付け時に使用語彙の統一意識につながる事が判明した。実際に登録されたデータの解析からも語彙が統制されていることが判明し、プラットフォーム設計などを通じユーザ間に強調意識が生まれれば、タグにおける多義語の発生を抑制することができるかと判断できる。ただし、SBM の利用は

*3 なお、表 1 中 3 位”おもしろ”の類義語にある”おもしろ”は、”ろ”の後ろに全角スペースが含まれたもの。見た目は”おもしろ”と全く変わらないが、システム上は別の語彙として扱われる可能性があるため、別途集計した。

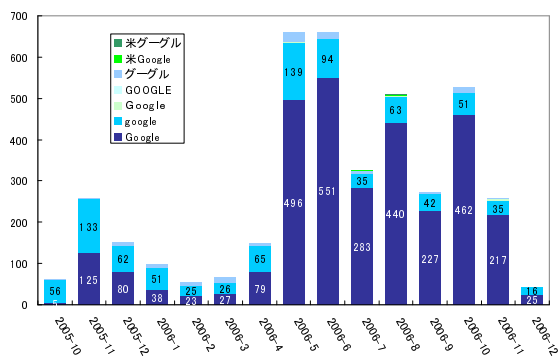


図 6: 使用語彙時系列変化:Google

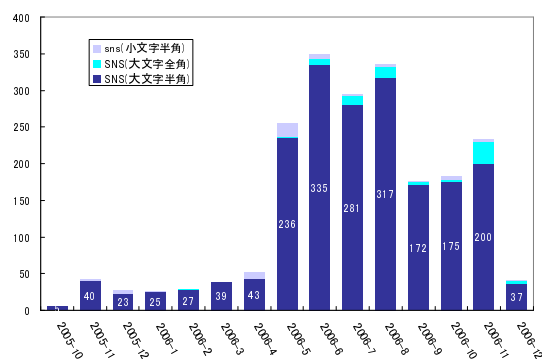


図 7: 使用語彙時系列変化: SNS

まだ大きく広がっておらず、今回の分析は基数の少ないデータをもとに行わざるを得なかった。また現在の利用者は初期採用者であり、サービスの普及に伴いフォロワーの利用が増加した際、ユーザ特性が変化する可能性もある。今後ユーザ数増加の後、十分な基数で調査分析を行うことにより、今回出した結論が一般的なものであり続けるかの検証を改めて行う必要がある。加えて、同義性以外の多義性や基礎水準の混在といった Folksonomy が成立する上での障害がクリアされるのかについても、別途検証が必要である。

7. 謝辞

今回の調査分析にあたり、蓄積データを提供して下さるとともに、ユーザアンケート実施に協力いただいた、株式会社 EC ナビの對馬様、藪田様にお礼申し上げます。

参考文献

[Golder 05] Golder, S. and Huberman, B. A.: The Structure of Collaborative Tagging Systems, *Arxiv preprint cs.DL/0508082* (2005)

[Mathes 04] Mathes, A.: Folksonomies-cooperative classification and communication through shared metadata, *Technical Report, University of Illinois Urbana-Champaign* (2004)