

タグ付けスタイルの時間的变化

Temporal Changes of Tagging Styles

高橋 公海 佐藤 進也 中村 哲也 松尾 真人
Masami Takahashi Shin-ya Sato Tetsuya Nakamura Matsuo Masato

日本電信電話(株) NTT 未来ねっと研究所

NTT Network Innovation Laboratories, Nippon Telegraph and Telephone Corporation

The recent information explosion in the Internet makes many users hard to reach and find an appropriate information easily. To achieve an efficient and easy access to a suitable information, a method of content categorization is important. In this paper, we have analyzed the trend change of the way how to generate tags by users to describe the content for web sites. We define two kinds of tags, subjective and objective one. A subjective tag is a user-defined symbol such as a face text or personal impression. An objective one is a common keyword. We observe the trend that a ratio of an objective tag has increased for the purpose of categorizing the web contents. On the other hand, a subjective tag is mainly used for indicating impressions and communicating to others. We also show the result of several user hearing.

1. はじめに

情報爆発にともない、入手可能な情報が増える反面、それらの情報が人間の処理能力を超え、かえって効率が低下する情報過多の問題が深刻化している。最近では Twitter*1 のようなソーシャルメディアも浸透しつつあり、個人が発信する情報がさらに増加することに加え、今後はセンサーデータなどのライフログの増加も見込まれる。このような情報過多の状況の中で、非常に膨大な情報の中から、本当に必要な情報を必要な時に効率よく取り出すには、どのようにすれば良いだろうか。本研究ではこの課題に取り組む。

必要なものを必要な時に取り出せる状態とはすなわち、整理された状態のことである。情報がきちんと整理されていれば、本当に必要な情報を必要な時に効率よく取り出すことが可能となる。しかし、その仕組みをトップダウンに設計し、実現することは難しい。例えば、一見すると部屋に物が乱雑に置かれているように見えても、部屋の持ち主はどこに何があるかきちんと把握しており、必要な時にはすぐに取り出せる場合がある。他人がその状態を見ると整理出来ているとは感じられないかもしれないが、本人にとっては整理された状態であると言える。この例は情報が「整理されている」ということが、情報と人の関わり方に依存しているということを示唆している。

そこで、人がどのように情報を整理しているのかを、人の情報整理行動を通じてボトムアップに探ることを試みる。人は自分にとって使いやすい状態になるよう、情報を整理していく。その過程を追うことによって、どのように整理するのが良いかを探る。本論文では、ソーシャルタギングにおけるリソースへのタグ付けを例に、分析を行った。ソーシャルタギングでは、リソースに対しユーザがタグと呼ばれるメタデータを付与することで、リソースを整理している。そのため、ユーザがリソースに対してどのようなタグを付与しているか、というタグ付けのスタイルを分析することは、ユーザが情報をどのように整理しているかを明らかにするために有効であると考えられる。

これまで、タグと言えばタグ付け対象となるリソースの特

徴を直接的に表現する語が多く用いられてきた。一方で、今回分析の対象とした、はてなブックマーク*2 などのソーシャルタギングサービスでは、顔文字や「これはひどい」「あとで読む」といった主観的な見方を示す語をタグとして付与するユーザも存在する。ソーシャルタギングにおけるタグは、基本的にリソースの分類・整理の効率化を目的として付与され、リソースの有効活用を手助けするものである。しかし、主観的な見方は時間の経過等の要因で変化する可能性が高く、それをタグとして残すことは、分類・検索精度の低下に繋がりがねない。にも関わらず、主観的な見方をタグとして用いるユーザが存在するという事は、そこには何らかの理由があると考えられる。

本論文では、タグ付け対象の特徴を客観的に説明するタグを客観タグ、主観的な見方や顔文字のタグを主観タグと呼び、それぞれが情報を整理する上でどのような用途で使われているのか考察を行う。また、その用途は変化してきているのか。もし変化しているのであれば、なぜ変化しているのか。という点についても、分析とユーザへのアンケートを通じて明らかにする。

分析では、ユーザがリソースに付与するタグの変化を時系列で追ったが、情報の整理手法を試行錯誤しているユーザも見受けられた。ソーシャルタギングでは、ユーザが自由にタグを選択・付与することができるが、それゆえにタグが増えすぎてしまう、表記ゆれが生じるなどの問題もある。今回の分析とアンケートを通じて、ユーザがサービスを利用しながら各自の最適な情報整理手法を模索し、工夫している様子を見ることができた。

本論文の構成は次の通りである。2章では関連研究について述べる。3章で、客観タグ・主観タグに着目したタグ付けスタイルの分析とその考察について説明する。4章はまとめと今後の課題である。

2. 関連研究

主観情報の付与に着目した研究として、メタ認知促進支援を目的とした主観アノテーションシステムが挙げられる [伊藤 07]。このシステムでは、ユーザが Web 上にアップロードした写真

連絡先: 高橋 公海, NTT 未来ねっと研究所, 〒 180-8585
東京都武蔵野市緑町 3-9-11, t.masami@lab.ntt.co.jp

*1 <http://twitter.com/>

*2 <http://b.hatena.ne.jp/>

等のデータに対し、アバターを介して感情や意志などの主観情報を付与しておくことができる。それにより、Web上のデータと関連する体験の再現・振り返りが容易になり、メタ認知が促進されるというものである。伊藤らはメタ認知促進支援を目的とした主観情報の付与を提案しているが、本論文では、情報を分類・整理するという観点から、主観情報を付与する意義について分析・考察を行っている。

ソーシャルタギングに関連する研究は数多くあるが、実データを活用して行われた分析の中で代表的なものとして、Golderらの分析 [Golder and Huberman 05] が挙げられる。実際に del.icio.us^{*3} のデータを用いて、ユーザがリソースに付与するタグの種類・度数分布のパターンを抽出し、ソーシャルタギングにおけるタグの性質に関する仮説を提唱した。del.icio.usにおけるタグの機能を7項目(詳細は3.1参照)に分類し、ソーシャルタギングにおけるタグがどのような機能を持っているか考察した点は特に興味深い。本論文ではユーザがどのようなタグを利用しているかに関してアンケートを行っているが、Golderらの分析を裏付ける回答(詳細は3.3.3参照)も見られた。Golderらはdel.icio.usを対象に分析を行っていたが、サービスの違いによるタグの機能の差異はそれほど見られないことが分かった。

深見らは、ソーシャルタギングが日本で利用され始めた初期段階において、ソーシャルタギングという手法の有効性とその情報流通形態について分析を行っている [深見 07]。情報探索や分類という観点からソーシャルタギングを捉え、ソーシャルタギングを利用しているユーザへのヒアリングを実際に行うなど、内容が多岐に渡るため大変興味深い。これに対し本論文では、ソーシャルタギングにおけるタグ付けスタイルの変化を時系列で分析することによって、情報の整理手法はどのように変化し、今後どうなっていくのかという点を明らかにすることを目指している。このような分析は、ソーシャルタギングが利用され始めた初期の段階では難しく、ソーシャルタギングを数年利用しているユーザが多く存在する今だからこそ、可能な分析であると考えられる。

岩田らは、ソーシャルタギングにおける主観タグをラベルノイズと考え、それを含むデータに対しても頑健なトピックモデルを提案している [Iwata 09]。この手法では、コンテンツとラベルが生成される過程をモデル化し、確率的EMアルゴリズムを用いてモデルを学習することにより、各ラベルがコンテンツに関係するかどうかを推定している。実験ではソーシャルタギングサービスの実データを用いて、内容に関連しないラベルを分離することができるか検証しており、提案手法が有効であることが示されている。本論文の実験では、主観タグと客観タグを分離する際に、検索エンジンのヒット数を利用している。しかし、岩田らの手法を利用することで、より精度よく主観タグと客観タグを分離できる可能性もある。この点については、今後の課題である。

3. タグ付けスタイルの分析

本章ではまず、主観タグと客観タグについて、より詳しく説明する。次に、検索エンジンのヒット数を用いて主観タグと客観タグを分離する方法を述べる。その後、本論文で行った分析結果とユーザへのアンケートを通じた考察を述べる。

3.1 主観タグと客観タグ

今回の分析では、ユーザがリソースに付与したタグを主観タグと客観タグに分け、それぞれの時系列での変化を追った。

しかし、ソーシャルブックマークサービスでユーザが付与する全てのタグについて、客観・主観のどちらに基づいているかを正確に分離することは困難である。厳密に言えば、タグはユーザ自身が考え、付与するものであるため、全てのタグはユーザの主観に基づいているとみなすこともできる。この点について本論文では、リソースの特徴を直接説明するタグを客観タグ、リソースに対する主観的な見方を示すタグを主観タグとして分離している。例えば、<http://www.google.co.jp> を対象のリソースとした場合に、その特徴を直接示す客観タグとしては「google」「検索エンジン」等が挙げられる。これに対し、「便利」「よく使う」といったタグは、ユーザの主観的な見方を示していると考えられるため、本論文では主観タグとみなしている。

2章でも述べたが、Golderらはdel.icio.usにおけるタグの機能を次の7項目に分類している。本論文における客観タグは(1)に、主観タグは主に(5)に対応している。

- (1) Identifying What (or Who) it is About. : リソースの主題を表すタグ。主に名詞が用いられる。
- (2) Identifying What it Is. : リソースの種類に関するタグ。article, blog など。
- (3) Identifying Who Owns It. : リソースの所有者や制作者を示すタグ。
- (4) Refining Categories. : ユーザ独自のカテゴリ分類。
- (5) Identifying Qualities or Characteristics. : リソースの質や特徴を示すタグ。主に形容詞が用いられる。
- (6) Self Reference. : リソースとユーザとの関係を表すタグ。主に *my* で始まるもの。
- (7) Task Organizing. : タスクに関連するタグ。toread など。

今回の分析では、タグが客観・主観のどちらであるかを判定する方法として、検索エンジンのヒット数を用いた。次節では、その判定方法について述べる。

3.2 検索エンジンのヒット数を用いた分析

3.2.1 分析手法

タグ付けスタイルの変化を分析するため、検索エンジンのヒット数を用いて主観タグ・客観タグを分離し、それぞれの変化を追った。具体的には、あるリソースに対し、そのリソースに付与されたタグをクエリとして検索を行い、検索結果が何件あるかというヒット数を用いて、タグが主観的か客観的かを判断した。例えば、<http://www.google.co.jp> を対象のリソースとし、それに対しあるユーザが「検索エンジン」「便利」という2つのタグを付与していたとする。その場合、「site:www.google.co.jp "検索エンジン"」「site:www.google.co.jp "便利"」というクエリで検索を行う。その結果、ヒット数が1件以上であれば対象のリソースの特徴を直接示す語がタグとして付与されているとみなし、そのタグは客観タグとして扱う。ヒット数が0件であれば、対象のリソースの特徴を直接示していない語がタグとして付与されているとみなし、そのタグは主観タグとして扱う。

検索エンジンのヒット数を用いるアプローチは、次の2つの仮定に基づいたものである。(1)客観タグに使われる語は、リソースの特徴を直接示す語が多く用いられるため、検索を行った場合にヒットする可能性が高い。(2)主観タグに使われる語は、リソースの特徴を直接示すような語ではないため、検索を行ってもヒットする可能性が低い。これらの仮定から、リソースに付与されたタグをクエリとして、リソースを検索した際に、検索ヒット数が1件以上の場合には客観タグ、0件の場合には主観タグと判断し、時系列でその変化を追った。なお、検

*3 <http://www.delicious.com/>

索ヒット数の調査には、Yahoo!Web 検索 API^{*4}を利用した。

3.2.2 対象データ

今回の分析で用いたデータは、はてなブックマークの10ユーザ分、合計約2万エントリのデータである。1つのエントリには、ブックマークしたURL、ブックマークしたページのタイトル、ユーザid、タグ、ブックマークした時刻、コメント等の情報が含まれている。データの収集期間は2006年8月～2011年2月であるが、ユーザごとにサービスの利用開始時期が異なっている。最も利用期間が短いユーザは8ヶ月間であるが、全ユーザのサービスの利用期間は平均すると約3年であり、タグ付けスタイルの変化を分析するには十分な利用期間であると考えられる。分析後、ユーザへのアンケートを行うことを踏まえ、今回は10ユーザのみを対象に分析を行った。

3.2.3 分析結果

まず、2006年8月～2011年2月の月別の全てのタグ数・主観タグ数・客観タグ数の推移を図1に示す。縦軸は毎月付与されたタグ数、横軸は時間を示している。図1から、客観タグは比較的一定して付与されていることが分かる。一方で、主観タグは全体のタグ数の増減と対応して増減している。

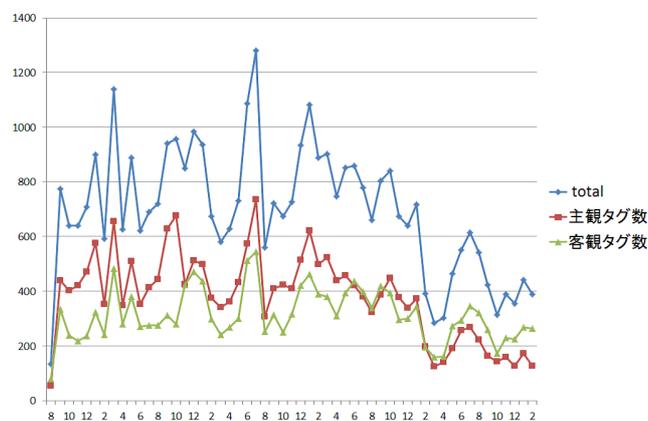


図1: 毎月付与された全タグ数・主観タグ数・客観タグ数の推移 (全ユーザ)

次に、毎月付与された全タグ数に占める客観タグ割合の推移を図2に示す。縦軸は毎月付与された客観タグの割合、横軸は時間を示している。図2から、時間の経過とともに、客観タグの割合が増加する傾向が見られる。

前述の通り、今回の分析では客観タグの割合を、その月に付与された全てのタグのうち、客観タグが何%含まれているかで単純に算出している。そのため、以前から客観タグをよく付与しているユーザが、主観タグを段々付与しなくなった場合に、客観タグの割合が増加していくように見える。図2で客観タグの割合が増加している傾向が見られたが、これは主観タグを段々付与しなくなったために起こった変化ではないかと考えられる。ユーザごとの分析も行ったが、ほぼ同様の傾向となったため、ブックマーク数の多い一部のユーザの影響が大きいといったことは考えにくい。次節では、このような結果となった理由について、ユーザへのアンケートを通じて考察を行う。

3.3 ユーザへのアンケート

ユーザへのアンケートは、前節で分析を行った10名のユーザのうち8名から回答を得た。調査日程は2011年3月7日～



図2: 毎月付与された客観タグ割合の推移 (全ユーザ)

20日である。サンプル数が非常に少ないことによるデータの精度の低さを念頭において考察を行うこととする。

前節の分析から、多少の個人差はあるものの、客観タグの割合が徐々に増加する傾向にあることが分かった。そこで、主観タグ・客観タグの使用頻度、使用目的とそれらの変化について自由回答で尋ねた。

3.3.1 客観タグ

客観タグに関しては、「後からエントリを探しやすいため、以前からよく使っている」と回答したユーザが5名、それ以外に「以前はあまり使わなかったが、最近をよく使うようになった」と回答したユーザも1名おり、この点については前節の分析結果を裏付けるものである。アンケートの中では、客観タグについて「ページのタイトルに含まれておらず、かつ内容に直接関連するタグを付けることによって後で発見しやすくなる」という事に気付いたため、積極的に付けるようになった」と回答したユーザも存在した。このように、ユーザがソーシャルタギングサービスを利用する中で、後で情報を活用しやすいように情報の整理手法を工夫している様子も見受けられた。

3.3.2 主観タグ

主観タグの使用頻度に関しては、全員が「以前にそれらのタグを付与したこともあったが、最近ではあまり使わなくなった」もしくは「利用開始当初から全く使っていない」と回答していた。主観タグを付与しなくなった理由としては、「それらのタグを手掛かりにしてエントリを探すことがなかったため、使わなくなった」という理由が多数であった。やはり、主観タグは情報の分類・検索という用途ではあまり有用ではないようである。では、主観タグは何を目的として付与されているのだろうか。

アンケートによると、主観タグを付与する目的は、他のユーザとのコミュニケーションや意思表示であるという回答を得た。また、「以前は意思表示のために使うこともあったが、最近ではそういったことはTwitterでつぶやくため、ブックマークのタグとして残すことは無くなった」というユーザも居た。これは、以前はソーシャルブックマークが他ユーザとのコミュニケーションの場としての役割も果たしていたが、最近ではTwitterがその役割を担うようになったことを示しているのではないかと。つまり、ソーシャルブックマークはコミュニケーションの場というよりは、Web上の情報を収集し、後で活用しやすいように整理しておくというブックマーク本来の役割が大きくなり、主観タグのような情報の分類・検索精度を下げるタグが徐々に減少していく結果になったのではないかと考えられる。

*4 <http://developer.yahoo.co.jp/webapi/search/>

ただし、主観タグの割合の減少は、主観タグ自体が使われなくなったことを意味しているわけではないと考えている。場は Twitter に移ったが、リソースにコミュニケーションや意思表示を目的としたメタ情報が付与されることは行われている。1章で述べたように、情報の整理について人と情報との関わり方という視点で捉えたときに、主観タグを付与することは何らかの意義があると思われる。実際に、コミュニケーションや意思表示を目的としたメタ情報を付与することによって、どのような意義があるのかを明らかにすることは今後の課題である。

3.3.3 その他のタグ

アンケートでは、主観タグ・客観タグ以外でよく利用するタグについても自由回答で尋ねた「あとで読む」といった予定や行動の備忘録、著者などリソースに関連する人物名、制作者へのメッセージといったタグを利用するというユーザが存在することが分かった。これらは、Golder らのタグの機能分類における (7) Task Organizing, (3) Identifying Who Owns It, (5) Identifying Qualities or Characteristics にそれぞれ該当し、Golder らの分析を裏付けるものである。はてなブックマークでは顔文字をタグとして用いるユーザも見られたが、アンケートによると「リソースに対する個人的な感想や解釈を示すために使っている」との回答を得た。これは、Golder らのタグの機能分類における (3) Identifying Qualities or Characteristics に該当するものと考えられる。これらのタグの機能分類は del.icio.us を対象に考案されたが、サービスの違いによるタグの用途の差異はそれほど見られないことが分かった。

4. まとめと今後の課題

本研究では、主観タグ・客観タグに着目したタグ付けスタイルの変化の分析と、ユーザへのアンケートを通じた考察を行った。今回の分析では、どのユーザも客観タグの割合が徐々に増加する傾向にあるという変化が見られた。情報の分類・整理という観点では、やはり客観タグを付与するユーザが多く、検索精度向上のため意識的に客観タグを付与するようになったユーザも存在することが分かった。

また、主観タグを残す目的として、情報の分類・整理ではなく、他のユーザとのコミュニケーションや意思表示という用途がメインであることも明らかになった。主観タグの割合が減少傾向にある要因としては、コミュニケーションの場がソーシャルタギングから Twitter などに移り、ソーシャルタギングサービスはあくまで Web 上の情報を収集・整理という用途で利用されるようになっていったことが考えられる。その結果、主観タグのような情報の分類・検索精度を下げるタグは徐々に利用しなくなっていった可能性がある。

今後の課題としては、データ数を増やし、より大規模な実験を行い、今回の分析結果と考察の妥当性について検証を行う必要がある。また、コミュニケーションや意思表示を目的としたメタ情報の付与は、情報整理という観点で実際に何に役立っているのかを明らかにすることが挙げられる。さらに、本論文での分析を踏まえ、情報をどのように分類・整理していくのが良いか。という点をより明確に示す必要があると考えている。

参考文献

[Golder and Huberman 05] Golder, S. and Huberman, B. A., The Structure of Collaborative Tagging Systems, Journal of Information, Vol. 32, No. 2, pp. 198-208, 2005.

[深見 07] Yoshiaki Fukami, Social Tagging -New Conformation of Information Transaction Generated by Unintended Collaboration-, Master Thesis 2007.

[伊藤 07] 伊藤 冬子, 廣安 知之, 三木 光範. アバターによる主観アノテーションの提案, 2B4-11, 人工知能学会第 21 回全国大会論文集, 2007.

[Iwata 09] Tomoharu Iwata, Takeshi Yamada, Naonori Ueda. Modeling Social Annotation Data with Content Relevance using a Topic Model, Advances in Neural Information Processing Systems (NIPS2009), 835-843, 2009.