

ソーシャルブックマークにおけるタグの関係性分析

Analysis of relation among tags in social bookmark

土屋 吉寛 †

Yoshihiro Tsuchiya

和田 雄次 ‡

Yuji Wada

1. はじめに

現在、ウェブの世界でホットなキーワードとして「フォーカスノミー」¹⁾という言葉が注目を集めている。フォーカスノミーとは、多くの人々が思い思いに入力した情報に基づく、ネットワーク的かつフラットな情報の構造であり、ディレクトリ構造のような諸概念を階層的に組織する従来の分類学的発想に対して言う言葉である。

近年はGoogleやAmazonなどビジネス上での成功事例もあり、Web2.0ブームの元に様々なサービスが開発されており、その中で「フォーカスノミー」もタグという実装で各サービスの中で利用されている。

今回はそのようなサービスの中でもソーシャルブックマークのタグの関係性を分析することによって、同意語を表すタグやタグ同士の類似性の抽出するため、またURLがどのようなものを表しているかを、タグを元に自動的に抽出するための実験を行った。

2. ソーシャルブックマーク

ソーシャルブックマークとはWeb上で他のユーザーとブックマーク情報を共有するためのサービスである。そしてその共有を助ける手段として現在多くのソーシャルブックマークはタグという機能を備えている。タグというのは各ユーザーがそれぞれのURLに対して自分が思ったままのキーワードを付加することで情報を分類する手段である。その場合のユーザー、タグ、URLを表した概念図が図1である。これにより分類された情報は各ユーザーのそれぞれが集めた情報を基にして構成されているため、網羅性に関してはGoogle

等のロボット型検索エンジンに劣るが、適合性に関しては分野や仕様用途によっては有効であると言われている。

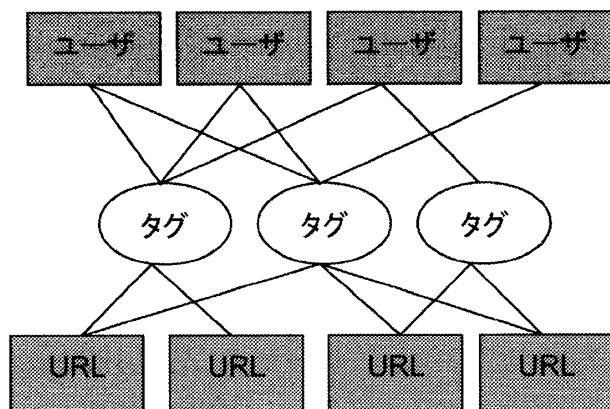


図1. ソーシャルブックマークの概念図

3. はてなブックマーク(2)

今回は分析対象としてはてなブックマークを使用した。このサービスは日本企業である株式会社はてなが提供するサービスで、特徴として日本語でタグの仕様が可能で、他のはてなのサービスとも連携していることもあり、日本国内ではソーシャルブックマークのディファクトスタンダードとして使用されている。

はてなブックマークを del.icio.us⁽³⁾ という海外のソーシャルブックマークと比較すると、主に使用されているタグに特長がある。タグの表現にひらがな、カタカナ、漢字、英語の混在した多くのバリエーションが存在する。そのバリエーションの多さの副作用としてタグに対して意図しない誤字・脱字が見られる。またスペースの使い方も個人によって使い方の差異がある。

そこで今回の実験では各URLとそれに対するタグの関係性を分析することによって、タグ同士の近似度やタグの付けられたURLから内容の推測が出来るかどうか実験してみた。

† 東京電機大学大学院 情報環境学研究科

‡ 東京電機大学 情報環境学部

4. 実験

今回の実験は以下の手順で行った。

1. 特定のタグの選出
 2. ブックマーク数が多い URL の抽出
 3. その URL に対するユーザーとタグを抽出
 4. URL とタグの関係においてコレスポンデンス分析を行う.

特定のタグは、ある程度使用率の高いタグの中から複数個選出して実験を試みた。そのタグを使用している URL の中からブックマーク数の多い上位 25 件の URL を抽出し、それぞれの URL で使用されているタグの中から、ある閾値で選別されたタグを対象にして分析を行った。コレスポンデンス分析という手法は主にマーケティングの分野でブランド・ポジショニングなどに使われる手法で、2 つの変数間の差や類似度を表現する手法です。

5. 結果

図2は「あとで読む」タグを元に情報を抽出し、コレスポンデンス分析を行った結果である。図2の上部にデザイン関係のタグが集中しており、また「css」と「CSS」という大文字、小文字しか差のないタグに関してはほぼ重なった状態で表示されている。実際に座標の値を確認しても css が $(x,y)=(0.040,4.700)$ に対し CSS(x,y)=(0.020,4.641) となり、大変近い値となる。また今回抽出された URL の中でデザイン関係を主に取り扱っている URL は一つしかなく、その URL に関してはそのデザイン関係のタグが集中している場所にプロットされていることから、今回の実験としては良好な結果が出ていると思われる。しかしタグに閾値をまったく設けない、もしくは閾値が適正値より低いと、独自タグが乱立した結果、精度が低下することがある。適正な閾値を設けることによって、他のタグを基準に分析を行った場合もおおむね良好な結果が出ることが確認できた。

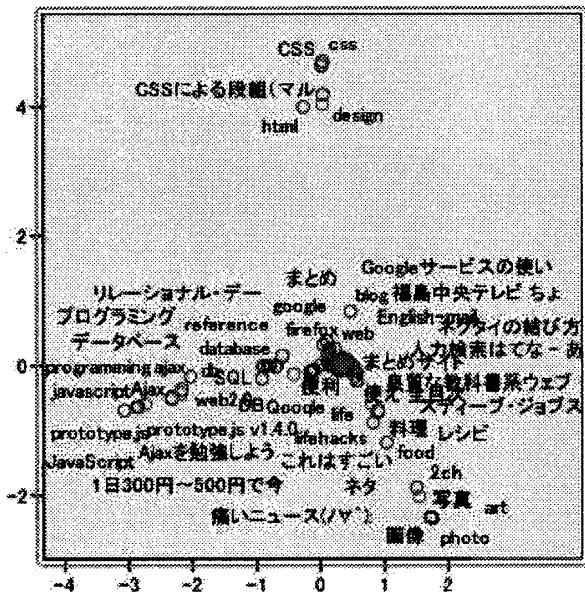


図2. 分析の結果例

6. まとめ

今回の実験によって、はてなブックマークからタグの関係性をある程度抽出出来ることがわかつた。今後はもう少し網羅的に分析することにより、サービス全体に対して分析を行うことで、同意語問題の解決手法を研究していきたい。また今回はタグと URL の要素に関して分析を行ったが、ユーザーも含めて別の切り口からも分析をしていきたいと考えている。さらに今回分析対象にしたタグを見る限り、サービスにおけるコミュニティの特色が出ていると考えられる。例えばはてなブックマークであれば、コンピューター技術に興味がある日本人が好んで使う傾向がある。そこで今後は他のソーシャルブックマークサービスと比較することにより、各ソーシャルブックマークのコミュニティの違いなども調べていきたい。

参考文献

- 1) J Golbeck, B Parsia, J Hendler : Folksonomies—Cooperative Classification and Communication Through Shared Metadata, (2004).
 - 2) はてなブックマーク : <http://b.hatena.ne.jp/>
 - 3) del.icio.us : <http://del.icio.us/>
 - 4) Peter Mika : Ontologies are us:A unified model of social networks and semantics (2005)