

L_060

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クソノミー¹⁾という言葉が注目を集めている。フォークソノミーとは、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、その中でフォークソ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ある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。

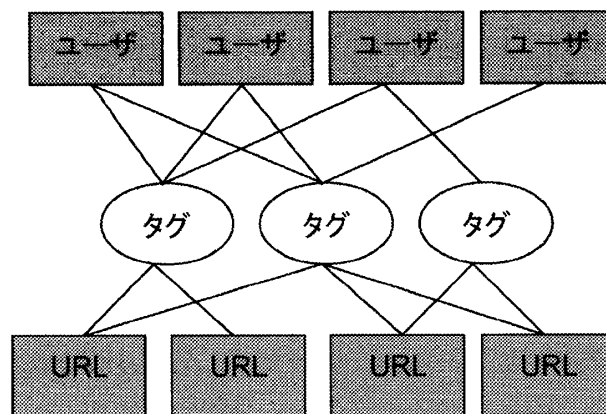


図1. ソーシャルブックマークの概念図

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.us³⁾ 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、カタカタ、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 情報環境学研究所

‡東京電機大学 情報環境学部

