

タグを用いた WEB サイト検索システム

石原進一[†] 松本和弘[†] 内田理[†] 中西祥八郎[†]

Website Search System Using Tag

Shin-ichi Ishihara[†] Kazuhiro Matsumoto[†] Osamu Uchida[†] Shohachiro Nakanishi[†]

1. はじめに

近年、ブロードバンドの普及に伴い幅広い年代層にインターネットが浸透してきた。インターネット利用者が情報検索を行う際、現在主にポータルサイトが利用されているが、これらのポータルサイトでは、キーワード検索や予めカテゴリー分けされている情報を掘り下げる形で情報を絞り込んでいくことが多い。したがって、膨大な検索結果から自分が欲しい情報を探さなければならない。

このような方法で検索して、ユーザが自分の要求にあったサイトを見つけるには、多大な労力が掛かる場合がある。そこで、ユーザは自分がもう一度、もしくは頻繁に見たいと思ったサイトはWEBブラウザのブックマークに登録する。こうすることで、毎回検索する手間もなく簡単に同じWEBサイトを閲覧することが可能となる。

さらに最近では、オンラインブックマークというサービスが提供されている[1][2]。オンラインブックマークとは、現在WEBブラウザについているブックマーク機能をサイト上に実装したものであり、インターネット環境があればどこからでも自分のブックマークを閲覧することができる。

2. オンラインブックマークとタグによる検索

2.1 オンラインブックマーク

オンラインブックマークではユーザは興味を持ったサイトを登録する傾向が強い。そのため、自然とユーザが興味を持たないWEBサイトは淘汰されていく。

このような性質を持つオンラインブックマークサイトを利用することで、ユーザにとって有効なWEBサイトを検索結果として得ることが容易になる。

現在、オンラインブックマークサービスでは分類に従来のWebディレクトリのような固定的・階層的な分類法ではなく、ユーザが自由に設定でき

る「タグ」と呼ばれる単語やフレーズを利用しているものが多い。複数のユーザがそれぞれ1つのサイトに自由にタグをつけるという方法で分類を行うことにより、自分と同じ関心を持つユーザのブックマークを閲覧することや、自分の思いつかなかった切り口により目的に合ったサイトを発見することができる。

しかし、Flog[1]やkaiba.jp[2]などオンラインブックマークサイトに実装されているタグシステムを利用した検索は、検索されたタグがWEBサイトに1つでもついていれば検索結果として出力されてしまう。そのため多くの種類のタグがつけられているWEBサイトはそのすべてのタグの検索結果として表示されてしまう。また、多くのユーザによく使用されるタグで検索を行った場合、ユーザは膨大な量の検索結果のサイトから自分の目的に合ったWEBサイトを探さなければならない。

そこで本研究では、オンラインブックマークサイト内における、タグを利用したWEBサイト検索手法を提案する。本システムを利用することで、タグシステムを利用しているオンラインブックマークサイトにおいて、ユーザの要求にあったサイトを効率よく検索することが可能になる。

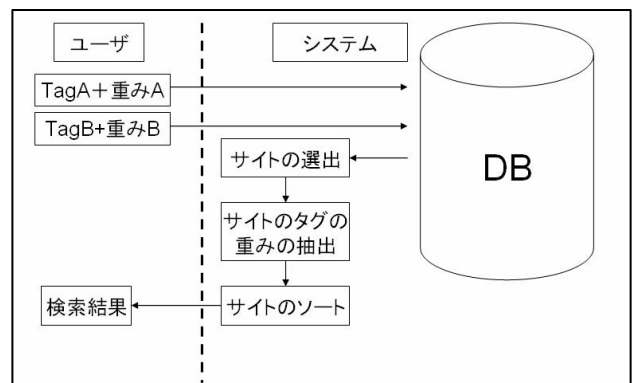


図1. システム構成

2.2 タグによる検索

タグを用いたオンラインブックマークサイトを利用するには、まずユーザはWEBサイトを自分で登録する。登録する際は、WEBサイトのURLとWEBサイトを分類するための単語(タグ)を入力す

[†] 東海大学電子情報学部情報科学科
School of Information Technology and Electronics,
Tokai University

る。各ユーザがタグでWEBサイトを分類することで、同じタグで登録されたWEBサイトを検索することができるようになる。さらに、多くのユーザが同じWEBサイトに様々な種類のタグをつけることにより、様々な角度から目的にあったWEBサイトを見つけることができるようになる。

本研究ではタグシステムを用いたオンラインブックマークサイト内でのWEBサイト検索システムの構築を試みた。本システムを利用するには検索に使用するタグを2つ入力して行う。システムは保存しているすべてのデータの中から入力された2つのタグのついたWEBサイトを、検索結果として表示する。

2.3 WEBサイト検索の優先度

このシステムで問題となるのは2つのタグを使って抽出した後のWEBサイトの優先順位である。抽出されたWEBサイトには検索に使用したタグ以外にも必要とされていないタグがついている為、必ずしもユーザが望むサイトであるとは限らない。今回の方法では検索時にユーザが2つのタグに重みをつけて検索することで、検索結果の優先順位を変化させる。

本システムにおけるWEBサイト検索方法は以下のようなルールに従って行われる。

- (1) DBからユーザが検索時に入力したTagAおよびTagBの両方のタグがついたWEBサイトを検索対象とする。
- (2) 検索対象となったWEBサイトにつけられているタグの数を集計し、上位2つがTagAおよびTagBであるWEBサイトを抽出する。
- (3) 抽出されたWEBサイトにTagAもしくはTagBをつけているユーザの数をそれぞれカウントし、そのWEBサイトにおけるTagAとTagBの和を分母としてTagA、TagBそれぞれの割合を求める。
- (4) (3)で計算した2つのタグの割合とユーザの入力した2つのタグの割合を比較し(式(1))、類似度の高い順に検索されたWEBサイトを出力結果とする。

$$\sum \left\{ 2 - \left| \frac{A(\text{site})}{A+b} - \frac{A(\text{user})}{A+B} \right| \right\} \quad (1)$$

3. システム評価実験

3.1 実験方法

提案システムの評価をするため以下のような実験を行った。まず似たようなカテゴリーのWEBサイトの中から50サイトを選出し、各サイトに

適しているであろうタグを5個ずつ選択肢として用意する。そして20人の被験者にこちらで用意したタグの中からサイトを分類するためのタグとして最もふさわしいものを、5つの中から1つ選んでもらう。次に、このようにして集めた50サイトにつけられたタグのなかからサイトを検索するための任意のタグを2つ選び、それらの割合を選択することによって、抽出されたサイトの優先度にどのような変化があるかを考察した。

3.2 結果と考察

今回の実験で使ったサイトから「音楽」と「アーティスト」というタグを用いて「音楽」から見た割合を2割、4割、7割と変えて検索してみたところ「バンド」、「シンガー」、「アイドル」のオフィシャルサイトが表1のような順番で検索結果として表示された。

表1. 実験結果

検索結果(割合は音楽)		
2割	4割	7割
アイドル	シンガー	バンド
シンガー	アイドル	シンガー
バンド	バンド	アイドル

2つのタグと、その割合を利用することによってサイトの属性を考慮した検索ができるようになったが、今回の実験では被験者の数が少なかったため、サイトにつけるタグが限定されていた。今後さらに登録ユーザを増やし、ユーザが自由な分類でタグをつけることによって、属性を考慮した検索が可能になると思われる。

4. まとめと展望

本研究によって、各ユーザがWEBサイトにつけたタグの割合はWEBサイトの属性を表現する指標の一つとして用いることが可能であることがわかった。今後の課題としては、上記のような特性を利用することで、より効率的なサイトの検索アルゴリズムを検討することがあげられる。

参考文献

- [1] Flog: <http://www.flog.jp/>
- [2] webshot: <http://s.phpspot.org/>
- [3] ネット世界をタグで分類する「フォークソノミー」
<http://hotwired.goo.ne.jp/news/culture/story/20050207204.html>