

ソーシャルタグの上位下位関係に基づく 散策的ブラウジング支援

木村 清 堯^{†1} 湯本 高 行^{†1}
新居 学^{†1} 高橋 豊^{†1}

本稿ではユーザが明確な目的を持たない、興味に基づいたブラウジングを散策的ブラウジングと定義し、この状態における情報推薦手法を提案する。提案手法では閲覧した Web ページを基準にソーシャルブックマークから発見し、推薦する。推薦の対象とする Web ページは汎化関係にある Web ページ、類似関係にある Web ページ、主要な一部分が異なった Web ページである。推薦する Web ページは入力ページに付与されたタグを用いて、タグ間の上位下位関係を利用して取得する。タグを入力として Web ページを取得する実験では、それぞれアプローチで求める Web ページ以外の Web ページを含んでいた。これは、閲覧した Web ページと推薦された Web ページの比較が不十分であるためと考えている。

Assisting Stroll Browsing based on Upper-lower Relation between Social Tags

KIYOTAKA KIMURA,^{†1} TAKAYUKI YUMOTO,^{†1}
MANABU NII^{†1} and YUTAKA TAKAHASHI^{†1}

In this paper, we propose a method of information recommendation for stroll browsing. Stroll browsing is browsing state in which users don't have clear goal and they browse web pages on the basis of their interest. The proposed method discovers and recommends web pages on the basis by browsed page from social bookmarks. The recommended web pages have generalization relation or similarity relation with the browsed page, or partly contain different content from the browsed page. The proposed method recommends web pages using upper-lower relation between tags of the browsed page. We consider it's not enough to compare browsed page with recommended pages.

1. はじめに

現在、Web ブラウジングが一般的になり、日常生活を過ごす上で様々な用途に使用されるようになった。ユーザは必要に応じて検索エンジンを用いて所望する知識を得たり、ニュースサイトにてニュースを購読したりすることで、興味のある分野の情報を容易に把握することができる。しかし、Web 上に情報が爆発的に増加してきたことで、ユーザは求める情報の閲覧に至るまでに膨大かつ多種多様な情報に目を通すこととなり、無視できない負担が生じている。

このような背景から、情報発見までの負担を軽減するための研究や、ブラウジングをより快適に行うための研究が活発に行われており、閲覧している Web ページに対して関連する内容の Web ページを推薦するものや、ブラウジングにより発生する閲覧履歴からユーザの好みを判定し、それに応じた Web ページを推薦するものなどが存在する。しかし、前述した Web ブラウジングの一般化により、Web 上から情報を探し出す以外の強い目的を持たないブラウジングが目立ってきたことで、ユーザの興味を惹く情報の推薦が重要であるが、膨大な情報がある中で閲覧中の内容との関係を分類して推薦するようなことは行われていない。

本稿では、強い目的を持たない Web ブラウジングを散策的ブラウジングとした上で、ユーザが興味を示す Web ページ群を、閲覧した Web ページとの関係により分類した形で提示する手法について述べる。推薦する Web ページ群は、閲覧したページに対して、上位のジャンルについて述べられている Web ページ（汎化ページ）、類似した内容の Web ページ（類似ページ）、同じジャンルではあるが別の内容について述べられている Web ページ（関連ページ）であり、それぞれを取得する手法について述べている。

推薦する Web ページはソーシャルブックマーク（以下、SBM）に登録されている Web ページを用いる。加えて、SBM で Web ページの分類に用いる“タグ”と呼ばれるものを用いて SBM 内からの Web ページの取得などに用いる。

以下、2 章では関連研究を、3 章では散策的ブラウジングについて述べた上で支援手法について述べる。そして、4 章で実験とそれらの考察を、最後に 5 章でまとめとする。

^{†1} 兵庫県立大学大学院工学研究科
Graduate School of Engineering, University of Hyogo

2. 関連研究

ユーザの散策的閲覧行動に対する研究としては、是津らがユーザの興味を持つページを閲覧していく際のブラウジングスタイルが、都市での散策行動に類似しているとして、“Webでの散策”というコンセプトを提案している¹⁾。そしてこれに基づいて、閲覧中のWebページの周辺空間を呈示するナビゲーション手法を提案している。この手法では、実際の都市での周辺空間に合わせて呈示することが特徴となっている。本研究で述べる散策に対する支援で推薦するWebページ集合は、この研究で閲覧ページに対する内容的な周辺空間に属するものと考えられる。

佐々木ら²⁾はSBMにて、あるユーザが用いているタグをベクトル化し、ベクトルを補完するという観点から、他のユーザのブックマークデータでの、そのタグに関係を持つWebページを推薦する手法を提案している。本研究とは異なりタグを階層的には用いていないが、同様に幅広い情報を推薦することが可能と考えられる。

丹羽ら³⁾はSBMのデータを用いたWebページの推薦手法を提案している。この手法では、各タグをタグ間の類似度によってクラスタを作成してタグ情報の抽象化を行い、タグとユーザ間の関連度を求める。これによって各タグクラスタにおけるそれぞれのユーザに適したWebページを推薦することができる。本研究と同様にSBMのタグの階層構造を用いているが、本研究ではクラスタは作成せずにタグの上位下位関係のみを用いている。

3. 支援手法

3.1 散策的ブラウジング

散策的ブラウジングとは、強い目的を持たない、ユーザ自身の興味に基づいたブラウジングである。具体例としては以下が挙げられる。

- 日常的に閲覧しているニュースサイトを閲覧している
- タイトル、概要から興味を持ったブログ（または個別の記事）を閲覧している
- 今日の運勢を確認している
- 動画共有サイトにて動画を鑑賞している

以上のようなブラウジングを本稿では散策的ブラウジングとしている。いずれの例でも、ブラウジングにおいて強い目的を持っておらず、娯楽的要素のあるブラウジング、緊急性の低いブラウジングといえる。

3.2 推薦するWebページ

前述したように散策的ブラウジングでは、ユーザの閲覧する際の目的はあいまいである。このような場合においては、ユーザが興味を示すであろう多分野のWebページや、汎化された内容のWebページなどを提示する支援を行えばよい。これは、ユーザが興味を抱くであろうWebページであれば、推薦対象としては問題なく、様々な情報が含まれたものを推薦することが重要であるためである。重要なことは良質なWebページ群を用いることであり、それを解決するためにSBMに登録されているWebページを用いる。

SBMにはユーザが頻繁に閲覧するWebページや、有益な情報が記載されたWebページを保存する用途などで登録されるため、Web上のページの中で良質なWebページが厳選されている。またSBMに保存されているブックマークに付与されているタグには“programming”や“料理”といったジャンルを意味するものが数多く存在しており、これを用いることでユーザが閲覧している内容に対して様々なアプローチで情報を推薦することが可能と考える。本稿では、入力である閲覧したWebページに対して、以下のように分類してWebページを推薦する。

- 汎化：閲覧している内容に対して、上位となるジャンルに対して述べているWebページを推薦する
- 類似：閲覧している内容と同ジャンルのWebページを推薦する
- 関連：同じ系統ではあるが、述べられている主要な一部分が別の内容となっているWebページを推薦する

例として、PerlプログラミングについてのWebページを閲覧している場合には、汎化ページとしてプログラミング全般についてのWebページ、類似ページとして同様にPerlでのプログラミングについてのWebページ、関連ページとしてはPerlではなくRubyでのプログラミングについてのWebページを取得、推薦する。

このようにWebページを分類して取得するにあたり、SBMのタグを活用する。タグ集合に対し、一部のタグを追加や削除、置換を行い、それを用いてSBMに対し問い合わせを行うことで、幅広いWebページを取得する。

それぞれのアプローチでWebページを取得において、共通の処理として入力とするブックマークに付与されているタグのうち、同ジャンルであり、上位下位の関係にあるタグの組（以下、タグペア）を判定、取得する。

3.3 Webページからのタグペアの生成手法

本手法での入力はSBMに登録されているWebページであり、取得するWebページの

基点となるものである．よって入力ページを基点ページとし p_b , これのブックマークに付与されているタグの集合を $Tag(p_b)$ として以下のように表す．

$$Tag(p_b) = \{t_1, t_2, \dots, t_n\} \quad (1)$$

それぞれのアプローチで Web ページを得るにあたり共通の処理として，基点ページ p_b のタグ集合から上位下位の階層関係をもつタグを 2 つ選択しタグペアを作成する．ここでは上位階層にあたるタグを t_u , 下位にあたるタグを t_l として，これを一組としたものをタグペアと定義し， $[t_l, t_u]$ と表記する．後の処理ではこのタグペアを用いて Web ページを取得する．基点ページ p_b から取得したタグペアの集合を $TagPair(p_b)$ として以下に示す．なお， $t_l \prec t_u$ は上位タグ t_u と下位タグ t_l に上位下位関係があることを示している．

$$TagPair(p_b) = \{[t_l, t_u] | t_l, t_u \in Tag(p_b), t_l \prec t_u\} \quad (2)$$

タグペアをブックマークから取得する基準として，以下の 2 点を用いて判定する．

- 2 つのタグが同じジャンルのタグであるか
- 2 つのタグに上下関係が見られるか

まず，同ジャンルのタグの判定には Jaccard 係数を用いる．そして，上下関係を指定したタグを付与している一意な Web ページ数によって判定する．Jaccard 係数による判定はしきい値 θ_1 を用いて以下のように行う．

$$Jaccard(t_l, t_u) \geq \theta_1 \quad (3)$$

Jaccard 係数の算出は以下のように行った．ここでは $\theta_1 = 0.01$ とした．

$$Jaccard(t_l, t_u) = \frac{|BM_{Tag}(t_l) \cap BM_{Tag}(t_u)|}{|BM_{Tag}(t_l) \cup BM_{Tag}(t_u)|} \quad (4)$$

$BM_{Tag}(t)$ はタグ t が付与されているブックマークの集合である．そして，上下関係の判定として，以下の条件を満たすことを条件とした

$$|Page(t_l)| < |Page(t_u)| \quad (5)$$

$Page(tag)$ はタグ tag が付与されている Web ページの集合である．

以上の手法で取得したタグペアを用いて，それぞれの Web ページを取得する．

3.4 汎化ページの取得手法

基準ページと汎化関係にある Web ページを取得する手法について述べる．これによりタグペアの上位タグについて，重点を置いて述べられている Web ページを取得する．ここでは，基点ページ p_b と汎化関係にある Web ページの集合 $GeneralizedPage(p_b)$ を，基点ページから生成したタグペアの上位タグ t_u を用いて取得する．

$$GeneralizedPage(p_b) = \{p | p \in Page(t_u)\} \quad (6)$$

ただしこの集合の各要素 p は，以下に示すタグ集合 $FrequentTag(p)$ の要素 t が上位タグ t_u の下位タグには不適切であることを条件とする．この条件により上位タグのジャンルをさらに限定して記述された Web ページを除去することが可能である．

$$FrequentTag(p) = \{t | \arg \max_{t \in Tag(p), t \neq t_u} DF(t, BM_{Page}(p))\} \quad (7)$$

なお， $BM_{Page}(p)$ は Web ページ p のブックマークの集合であり， $DF(t, BM_{Page}(p))$ は $BM_{Page}(p)$ の各要素におけるタグ t の出現頻度である．

3.5 類似ページの取得手法

ここではタグペアの上位下位タグの両方を用いて SBM に問い合わせを行い，Web ページを取得する．基点ページから作成した任意のタグペア tp を用いて，基点ページ p_b と類似関係にある Web ページ集合 $SimilarPage(p_b)$ を取得する．

$$SimilarPage(p_b) = Page(t_l) \cap Page(t_u) \quad (8)$$

$$tp = [t_l, t_u], t_l \prec t_u \quad (9)$$

3.6 関連ページの取得手法

タグペアの下位にあたるタグと同等の，上位タグのジャンルに属している別のタグを発見し，SBM に問い合わせることで Web ページを取得する．

基点ページ p_b と関連関係にある Web ページを取得するにあたり，基点ページの任意の上位タグ t_u を用いて，新たな下位タグを兄弟タグ t_s として取得する．そしてこの兄弟タグと上位タグを含む Web ページを関連ページの集合を $RelevantPage(p_b)$ とする．

$$RelevantPage(p_b) = Page(t_s) \cap Page(t_u) \quad (10)$$

$$tp = [t_s, t_u], t_s \prec t_u \quad (11)$$

兄弟タグ t_s を取得する手法について述べる．基点ページから作成した任意のタグペア tp を入力として兄弟タグの集合 $SiblingTag(t_u)$ を作成する．以下にその条件と共に示す．

$$SiblingTag(t_u) = \{t_s | t_s \in Large(CoOccurrenceTag(t_u), Confidence(t \rightarrow t_u), num), t \in CoOccurrenceTag(t_u)\} \quad (12)$$

$Large(CoOccurrenceTag(t_u), Confidence(t \rightarrow t_u), num)$ は，上位タグ t_u と同じブックマークで共起しているタグの集合 $CoOccurrenceTag(t_u)$ の各要素 t について，相関ルール $t \rightarrow t_u$ の確信度 $Confidence(t \rightarrow t_u)$ が上位 num 件の集合を表している．なお， $num = 15$ とし，確信度は以下のようにして算出した．

$$Confidence(t \rightarrow t_u) = \frac{|BM_{Tag}(t) \cap BM_{Tag}(t_u)|}{|BM_{Tag}(t)|} \quad (13)$$

ただし、兄弟タグ t_s はしきい値 θ_2, θ_3 を用いて以下の条件を満たすものに限る。

$$Support(t_s \rightarrow t_u) \geq \theta_2 \tag{14}$$

$$|User(t_s)| \geq \theta_3 \tag{15}$$

$Support(t_s \rightarrow t_u)$ は相関ルール $t_s \rightarrow t_u$ の支持度であり以下のように表す。 N は全ブックマーク数である。なお、 $\theta_2 = 0.00001$ とした。また、 $User(t)$ はタグ t を SBM 中で使用しているユーザの集合である。ここでは $\theta_3 = 50$ とした。

$$Support(t \rightarrow t_u) = \frac{|BM_{Tag}(t) \cap BM_{Tag}(t_u)|}{N} \tag{16}$$

4. 実験

実験で用いる SBM は livedoor クリップの 2009 年 12 月時点のものを用いて、タグの大文字小文字は区別せずに扱った⁴⁾。

4.1 タグペアの判定

任意のタグの組を作成し、それらがタグペアであるか（同ジャンルであり、上下関係を見受けることができるか）を判定した。表 1 に結果を示す。なお、表中の “x” は評価したタグの組がタグペアであることを、“x” はタグペアではないことを表している。

表 1 に示したタグペアの判定結果から、提案手法によりタグペアとして [Perl, Programming] や [カレー, 料理], [Firefox, Software] などを発見した。しかし [Firefox, Software] については、上位下位の関係にはあるが、想定していたタグ “Browser” のさらに上位の階層にあたるタグ “Software” がタグ “Firefox” タグペアと提案手法では判定してしまった。概念的に離れているとの解釈もできるため何らかの条件を追加する必要があると考える。また、タグ “Firefox” とタグ “Browser” の関係を手法ではタグペアと判定することができなかった。これは式 5 で、タグ “Firefox” が付与されている Web ページ数を意味する $|Page(Firefox)|$ の値がタグ “Browser” の値 $|Page(Browser)|$ を上回ったためである。SBM はユーザが登録したブックマークの情報のみを有しているため、各ジャンルに対するページ数が同等ではない。よって何らかの正規化が必要である。

4.2 汎化ページの取得評価

任意の上位タグを用いて、汎化ページの取得を行った。表 2 に用いたタグとそれぞれの取得件数を、表 3, 4 に取得した Web ページの抜粋を示す。

表 3 に示している取得した汎化ページには、複数の Web ブラウザを評価している Web ページや、特定のブラウザに依存していないブラウザ操作補助ソフトウェアの紹介 Web ペー

表 1 タグの上位下位判定結果

下位タグ	上位タグ	正解	判定結果	下位タグ	上位タグ	正解	判定結果
Perl	Programming			Firefox	Browser		x
Perl	Ruby	x	x	Firefox	Opera	x	x
Perl	CPAN	x	x	Firefox	Software		
Perl	Book	x	x	Firefox	Editor	x	x
Perl	Gmail	x	x	Firefox	google	x	
Perl	Browser	x	x	Firefox	Programming	x	x
Perl	Editor	x	x	Firefox	Book	x	x
Perl	Tips	x	x	Firefox	tips	x	
Perl	Health	x	x	Firefox	css	x	
カレー	料理			Emacs	Editor		x
カレー	レシピ			Emacs	Software		x
カレー	recipe		x	Emacs	vi	x	x
カレー	cook		x	Emacs	vim	x	x
カレー	food		x	Emacs	Programming		x
カレー	Health	x	x	Emacs	lisp		x
カレー	Book	x	x	Emacs	perl	x	x
カレー	2ch	x	x	Emacs	Book	x	x
カレー	tips	x	x	Emacs	2ch	x	x

表 2 汎化ページの取得結果

上位タグ	取得件数
Browser	1426
レシピ	556

表 3 タグ “Browser” を用いた汎化ページ取得結果 (抜粋)

タイトルおよび URL
ITmedia News : 今学期の「ブラウザ通信簿」, がんばった子は誰? http://www.itmedia.co.jp/news/articles/0608/04/news074.html
3分 LifeHacking “ 素 ” の Web ブラウザを起動する方法 - ITmedia Biz.ID http://bizmakoto.jp/bizid/articles/0708/16/news064.html
[CSS] ブラウザ間の差異を無くすリストのスタイルシート — コリス http://coliss.com/articles/build-websites/operation/css/866.html
そろそろウェブサイト側からも AutoPagerize を導入していったらどうか - Liner Note http://note.openvista.jp/2008/autopagerize-without-greasemonkey/
Mozilla Re-Mix: 開いているタブの URL リストを簡単に作成できる Firefox アドオン「URL Lister」 http://mozilla-remix.seesaa.net/article/125155458.html

表 4 タグ “レシピ” を用いた汎化ページ取得結果 (抜粋)

タイトルおよび URL
男が作れる超簡単料理 http://otoko-cooking.com/
こうちゃんの簡単料理レシピ - Yahoo! ブログ http://blogs.yahoo.co.jp/booooy2005
一人暮らしの作り置き:アルファルファモザイクだった http://alfalfa.livedoor.biz/archives/50951128.html
@nifty: デイリーポータル Z: 松本商店の美味しい干物レシピ http://portal.nifty.com/2008/10/22/c/
“飲む点滴” 甘酒の可能性を考える (Excite Bit コネタ) - エキサイトニュース http://www.excite.co.jp/News/bit/00091205339541.html
【2ch】 ニュー速クオリティ: 実はずごいな「ためしてガッテン」レシピ http://news4vip.livedoor.biz/archives/51250785.html

ジを取得することができた。しかし、Firefox の拡張機能の紹介といった限定的なものも取得している。これは汎化ページの式 7 から得られるタグ集合のいずれの要素も、入力とした上位タグとはタグペアの関係にならないことを条件としているためである。現手法では“Firefox” < “Browser” と判定できない。よってこの Web ページは式 7 から得たタグ集合に“Firefox”が含まれていたが、前述の理由により除くことができなかったと考える。

4.3 類似ページの取得評価

以下に示すタグペアを入力して、類似ページの取得を行った。表 5 に用いたタグペアとそれにより取得した Web ページの取得件数を示す。表 6, 7, 8, 9 に取得した Web ページの抜粋を示す。

表 5 類似ページの取得結果

タグペア	取得件数
[Perl, Programming]	421
[カレー, レシピ]	26
[Firefox, Browser]	542
[Emacs, Editor]	48

表 7 に示した Web ページはどれもカレーのレシピについての記事である。しかし他の結果では扱っている内容に幅が出た。表 8 に示すように、Firefox の情報をまとめた Wiki のように幅広い情報を載せた Web ページ、世界での Firefox のシェアに関する Web ページもあれば、拡張機能の実装における注意事項の Web ページを取得していた。このように

表 6 タグペア [Perl, Programming] を用いた類似ページ取得結果 (抜粋)

タイトルおよび URL
[ThinkIT] 第 1 回: Web アプリケーションフレームワークとは (1/4) http://www.thinkit.co.jp/free/tech/9/1/1.html
Perl の MVC フレームワーク Catalyst に入門してみた: NDO::Weblog http://naoya.dyndns.org/~naoya/mt/archives/001682.html
Perl 講座 [Smart] http://www.rfs.jp/sb/perl/index.html
JPEG 画像を SWF に変換する - 酒日記 はてな支店 http://d.hatena.ne.jp/sfujiwara/20070618/1182172990
BBC、「Perl on Rails」開発 - @ IT http://www.atmarkit.co.jp/news/200712/03/bbc.html

表 7 タグペア [カレー, レシピ] を用いた類似ページ取得結果 (抜粋)

タイトルおよび URL
Pumpkin in India ひき肉のカリー (Keema do Pyaza) http://humtypumpkin.blog40.fc2.com/blog-entry-102.html
スパイスたっぷり本格チキンカレーのレシピ - [おもてなし料理] All About http://allabout.co.jp/gm/gc/185352/
うまさ別次元! カレー大革命: ためしてガッテン - NHK http://cgi2.nhk.or.jp/gatten/archive/program.cgi?p_id=P20070131
ためしてガッテンのカレー特集がすごい!: akiyan.com http://www.akiyan.com/blog/archives/2007/02/post_81.html
海上自衛隊: 海上自衛隊レシピページ: 過去のレシピ http://www.mod.go.jp/msdf/formal/family/recipe/archive/currey.html

表 8 タグペア [Firefox, Browser] を用いた類似ページ取得結果 (抜粋)

タイトルおよび URL
SCRAPBOOK :: Firefox Extension http://amb.vis.ne.jp/mozilla/scrapbook/index.php?lang=ja
ITmedia Biz.ID: 第 2 次ブラウザ大戦、始まる (1/3) http://www.itmedia.co.jp/bizid/articles/0611/28/news079.html
Mozilla Links 日本語版: Firefox 3 のロケーションバー 9 個のワザ http://mozlinks-jp.blogspot.com/2008/06/firefox-3-9.html
Firefox の世界シェア、20 パーセントを越える - スラッシュドット・ジャパン http://slashdot.jp/it/08/12/07/0727249.shtml
AutoPagerize http://autopagerize.net/

表 9 タグペア [Emacs, Editor] を用いた類似ページ取得結果 (抜粋)

タイトルおよび URL
[を] Emacs 初心者のための最低限のコマンド表 http://nais.to/~yto/clog/2007-01-11-1.html
新入学生 / 新社会人応援企画 : ある Web プログラマーの作業環境 豪傑の三種の神器【前編】 (2/2) - ITmedia エンタープライズ http://www.itmedia.co.jp/enterprise/articles/0703/07/news017_2.html
プログラムしないけど emacs 使っています - コトリコ http://d.hatena.ne.jp/kotorikotoriko/20070310/1173510573
文書作成に役立つ Emacs Muse - SourceForge.JP Magazine : オープンソースの話題満載 http://opentechpress.jp/developer/article.pl?sid=07/04/11/0221217
まつもとゆきひろのハッカーズライフ : 第 4 回 Emacs 対 vi (1/2) - ITmedia エンタープライズ http://www.itmedia.co.jp/enterprise/articles/0706/26/news003.html

内容にばらつきが出た原因として、Web ページを 2 つのタグから構成されるタグペアに変換したことによる情報の大幅な欠如が原因と考えられる。

4.4 関連ページの取得評価

以下のタグペアを入力して、取得した下位タグをした。その結果を表 10, 11, 12, 13 に示す。また、それぞれの適合率を表 14 に示した。

取得したタグのうち、下位タグとなるもので確信度が最大のタグと、上位タグを用いて関連ページの取得を行った。用いたタグペアと取得件数を表 15 に示す。また、それぞれのタグペアで取得した Web ページの抜粋を表 16, 17, 18, 19 に示す。

下位タグの取得結果について述べる。結果を見るとタグペア [Perl, Programming] や [Firefox, Browser] の結果 (表 10 および表 12) の適合率は他の結果よりも良い。これらは共に情報系に関係するタグであることから、情報系のブックマークはブックマーク単体でタグが階層構造を成した状態で SBM に登録され易いと考えられる。また SBM にて多用されているため、多くの下位タグを取得できたと考えられる。

しかし表 13 から、“vi”や“vim”といった正解となるタグが確認できるが、その他は関係が見受けられないタグであった。また、表 11 のように確信度が高くとも下位タグとならないタグも含まれている。現時点では確信度の高い上位数件 (今回は 15 件) を正解としたが、絶対数ではなく確信度がしきい値以上のタグを正解とするなどの変更を考える必要がある。

また取得した汎化ページは (表 16, 17, 18, 19) それぞれのタグペアに適した Web ページといえる。しかし、類似ページ取得の考察でも述べたように、基点ページの情報が欠如するため、基点ページと同様の論点を持つ Web ページの選別が困難である。したがってタグ

表 10 タグペア [Perl, Programming] の下位タグの取得結果

下位タグ候補	確信度	正解
C 言語	0.24	
Algorithm	0.23	×
Language	0.22	×
C	0.21	
source	0.21	×
Code	0.19	×
C++	0.19	
coding	0.19	×
designPattern	0.19	×
Haskell	0.18	
Development	0.18	×
Matz	0.18	×
Lisp	0.16	
oop	0.16	×
computer	0.14	×

表 11 タグペア [カレー, レシピ] の下位タグの取得結果

下位タグ候補	確信度	正解
米粉	0.78	×
食べ方	0.57	×
パン	0.47	
マシン	0.47	×
炭水化物	0.39	×
おやつ	0.35	×
材料	0.34	×
お取り寄せ	0.28	×
ホームベーカリー	0.27	×
炊飯器	0.26	×
パスタ	0.26	
カロリー	0.26	×
スープ	0.24	
基本	0.20	×
作り方	0.20	×

表 12 タグベア [Firefox, Browser] の下位タグの取得結果

下位タグ候補	確信度	正解
chrome	0.15	
webkit	0.13	×
Sleipnir	0.13	
safari	0.12	
ie	0.12	
Opera	0.10	
ie6	0.10	
IE7	0.09	
mozilla	0.08	×
Bug	0.08	×
plugin	0.07	×
ブラウザ	0.05	×
software	0.03	×
summarySite	0.03	×
Debug	0.03	×

表 13 タグベア [Emacs, Editor] の下位タグの取得結果

下位タグ候補	確信度	正解
Wysiwyg	0.30	×
vi	0.08	
Vim	0.02	
HTML	0.01	×
OSX	0.01	×
Tools	0.01	×
Software	0.01	×
image	0.01	×
tool	0.01	×
windows	0.01	×
programming	0.01	×
mac	0.01	×
javascript	0.00	×
greasemonkey	0.00	×
ajax	0.00	×

表 14 下位タグの適合率

タグベア	適合率
[Perl, Programming]	0.67
[カレー, レシピ]	0.20
[Firefox, Browser]	0.47
[Emacs, Editor]	0.13

表 15 関連ページの取得結果

タグベア	取得件数
[C 言語, Programming]	25
[パン, レシピ]	14
[chrome, Browser]	101
[vi, Editor]	23

表 16 タグベア [C 言語, Programming] を用いた関連ページ取得結果 (抜粋)

タイトルおよび URL
C プログラミング診断室 http://www.pro.or.jp/~fujii/mybooks/cdiag/ UML モデルをどうやって C 言語に落とし込むか (1/3) - @ IT MONOist http://www.atmarkit.co.jp/fembedded/robocon/etrobo06/road03/road03a.html memccpy() で文字列をコピーする - bk ブログ http://0xcc.net/blog/archives/000122.html WisdomSoft http://wisdom.sakura.ne.jp/ C 言語入門、東京大学情報科学科の場合 - スラッシュドット・ジャパン http://slashdot.jp/articles/08/01/25/0318220.shtml

表 17 タグベア [パン, レシピ] を用いた関連ページ取得結果 (抜粋)

タイトルおよび URL
Home Bakery まりの部屋 (ホームベーカリーでパンづくり) http://mari2.net/ 【2ch】ニュー速クオリティ:【レシピ】簡単で旨いパンの食し方 http://news4vip.livedoor.biz/archives/51149286.html 白いきんぴら http://www.oajbnfu.com/16/ 新潟製粉の米粉 http://www.dhbneu.com/16/ 米粉販売 http://www.xtson3.com/6/

表 18 タグベア [chrome,Browser] を用いた関連ページ取得結果 (抜粋)

タイトルおよび URL
Google が独自ブラウザ「Google Chrome」を 9 月 2 日にリリース http://internet.watch.impress.co.jp/cda/news/2008/09/02/20727.html
Analytics 日本版 公式ブログ: Google Chrome と Analytics について http://analytics-ja.blogspot.com/2008/09/google-chrome-and-analytics.html
Chrome のプライバシー機能は穴だらけ? - ITmedia アンカーデスク http://www.itmedia.co.jp/anchordesk/articles/0809/04/news071.html
まだ使ったことがないなら使うべきレベルに到達した 体感速度爆速ブラウザ「Google Chrome」の真の実力 - GIGAZINE http://gigazine.net/index.php?news/comments/20090207_google_chrome/document.writeの中に埋め込んだscriptタグ-微熱everyday-subtech
http://subtech.g.hatena.ne.jp/secondlife/20090606/1244254492

表 19 タグベア [vi,Editor] を用いた関連ページ取得結果 (抜粋)

タイトルおよび URL
Peace Pipe: 効率的なテキスト編集の 7 つの習慣 [vim] http://peace-pipe.blogspot.com/2007/02/7-vim.html
(PHP) プログラマのための VIM (まとめ) - おぎろくはてな http://d.hatena.ne.jp/i_ogi/20070715/1184480249
Vim マスターへの道 - SourceForge.JP Magazine : オープンソースの話題満載 http://opentechpress.jp/special/article.pl?sid=07/07/18/0849244
Vim を使った Ruby On Rails 開発環境の構築 - ナレッジエース http://blog.blueblack.net/item_133
初めて vi エディタを使いたい人へ - Unix 的なアレ http://d.hatena.ne.jp/wadap/20080107/1199716341

ベアのモデルを改良する必要といえる。

5. ま と め

本稿ではユーザが興味に基づいたブラウジングを散策ブラウジングとした上で、閲覧した Web ページを基準に様々な観点から Web ページを推薦する手法について述べた。今後の課題としてはタグベアの判定手法の改良や、Web ページの取得手法の改良、そして取得した Web ページの評価方法の検討、実証が挙げられる。

謝 辞

本研究の一部は、平成 22 年度科研費基盤研究 (B)(2)「ユーザの潜在的意図を用いたレス・コンシャス情報検索基盤の構築」(課題番号: 20300039) によるものです。ここに記して謝意を表すものとします。

参 考 文 献

- 1) 是津 耕司, 田中 浩也, 池田 新平, 金星ヨシ, 田中克己: Web 上での散策行動を支援する周辺情報提示機構, 情報処理学会研究報告. データベース・システム研究会報告, No.71, pp. 343-349 (2003).
- 2) 佐々木 祥, 宮田 高道, 稲積 泰宏, 小林 亜樹, 酒井善則: Folksonomy におけるコンテンツ推薦のためのメタデータ成長モデルの提案 (情報抽出), 電子情報通信学会技術研究報告. DE, データ工学, Vol. 106, No. 150, pp. 67-72 (2006).
- 3) 丹羽 智史, 土肥 拓生, 本位田真一: Folksonomy マイニングに基づく Web ページ推薦システム (エージェント応用システム, <特集> マルチエージェントの理論と応用), 情報処理学会論文誌, Vol.47, No.5, pp. 1382-1392 (2006).
- 4) livedoor クリップ, <http://clip.livedoor.com/>.