情報処理学会論文誌　データベース　Vol. 2　No. 3　130–140 (Sep. 2009)

ソーシャルブックマーキングの周期性発見と時期連動型検索ランキングへの適用

山家雄介[†1]　中村聡史[†1]
アダム ヤトフト[†1]　田中克己[†1]

Web コンテンツの中には，1 年のうち特定の時期に頻繁に利用されるような周期性を持つものが存在する．本論文では，そのようなコンテンツが人気となる時期や対応するキーワードをソーシャルブックマークを分析することで，Web 検索結果から時宜を得たページの発見を支援する時期連動型ランキング手法を提案する．実際には，ソーシャルブックマークにおける過去のブックマーク日時や付与されたタグを分析することで人気度の予測を行い，検索が行われた時期に応じた検索結果のランキングを可能にする．まず，コンテンツごとのブックマークの周期性などの特性を調査し，その分析結果に基づいてコンテンツの人気予測モデルを提案する．次に，このモデルに基づいた Web 検索結果のランキング手法を提案し，実データによる評価を行う．

Discovering Periodicity of Social Bookmarking and Its Applications to Time-dependent Ranking

Yusuke Yanbe,[†1] Satoshi Nakamura,[†1]
Adam Jatowt[†1] and Katsumi Tanaka[†1]

Some web contents are used frequently in certain seasons of the year. In this paper, we propose time-dependent page ranking method that helps discovering such periodically used web contents by analyzing time periods when they get popular according to the rates of their social bookmarks. In practice, the method enables time-specific ranking based on social bookmark history utilizing the timestamps of bookmarks and their tags. First of all, we analyze temporal characteristics of Web content such as periodicity of its social bookmarks. Next, based on the results of this analysis, we propose a model of popularity estimation of Web content. Finally, we demonstrate a ranking method based on the proposed model, and make evaluation using real data.

1. まえがき

今日，Web 検索サービスは一種のインフラのようなものになっており，人々の生活において欠かせないものになりつつある．人々は知らない言葉に出会ったときや何かしらの物品を購入するとき，生活において病気や料理などで気になったときに，検索することで，即座に情報を取得することができる．

ここで生活に関する検索では，ある特定の時期に必要となる情報が多数存在する．たとえば，春が近づくと花粉の飛び具合が気になったり年末になると年賀状に適切な画像を探したりするであろう．話を逆転させると，ふだんはあまり必要とされていないがある特定の時期にだけ需要が増加する Web ページがあるといえる．たとえば「卒業論文の書き方の解説」のようなコンテンツは，卒業論文が書かれる時期に特に多く利用される．

しかし，今日の Web 検索エンジンでは，このようなコンテンツ需要の時期による変化に適応して，検索結果のランキングを修正するといったことは行われていない．そこで，このような周期性を持つコンテンツ需要を何らかの方法で発見し，検索結果のランキングに利用することが考えられる．

そこで我々は，近年 CGM として注目を浴びているソーシャルブックマークに着目した．その中でも特にブックマークの際に発生する時間情報に注目しこの情報の分析を行うことで，どの時期に需要が増加するかを予測できるのではというのが我々の研究のベースである．たとえば先述の「卒業論文の書き方」のページ[⋆1]のはてなブックマークにおけるブックマーク数の時間変化は，**図 1** のとおりである．この図からも明らかなように，卒業論文が書かれる時期（2007 年と 2008 年の 1 月）によくブックマークされていることが分かる．

そこで本論文では，検索クエリログのような時刻印を持つ頻度データからバーストが発生している区間を半自動的に抽出する手法を，ソーシャルブックマークのログに適用し，図 1 のような年単位で特定の時期に需要が増えるコンテンツを，ソーシャルブックマークが行われたページから自動的に抽出するための分析を行う．次に，時期に連動して変化する検索クエリに連動して Web 検索結果の再ランキングを行うシステムを提案する．

提案したシステムに対して実データに基づいた評価を行った結果，提案システムの再ランキングで元の検索結果の下位から上位に浮上するコンテンツには，1 年のうち決まった時期

†1 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University
⋆1 http://staff.aist.go.jp/toru-nakata/sotsuron.html

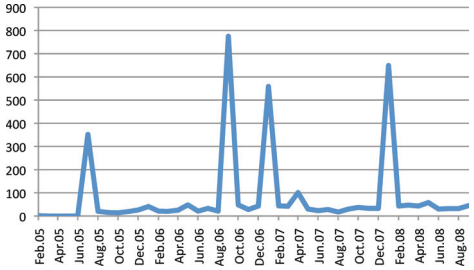

図 1　卒業論文の書き方に関するページのブックマークパターン（出典：はてなブックマーク）
Fig. 1　Bookmark pattern of a page showing how to write bachelor thesis (source: Hatena Bookmark).

に行われる行事などに関する Web ページが多く，傾向としてシーズン時は取引型の検索，それ以外の時期は調査型の検索をサポートすることが分かった．

本論文は次のように構成される．2 章では関連研究について述べる．3 章ではソーシャルブックマークの周期性分析を行い，どのようなページに周期性があるのか，また，周期性が発生している部分が他の部分とどのような関係があるのか調査する．4 章では 3 章の結果に基づいた Web 検索結果の再ランキング手法を提案する．5 章では提案手法を実データに適用して検索ランキングを生成したうえで評価を行い，6 章でその考察を行う．最後に本論文のまとめを行い，今後の課題について触れる．

2. 関 連 研 究

Broder [4] は Web 検索エンジンのクエリログを分析した結果，ユーザの情報要求の種類に基づいて以下の 3 種類に分類した．

- navigational query（誘導型の検索）
  ある事柄に関する代表的なページにたどり着くための検索
- informational query（調査型の検索）
  ある話題に関する参考になる情報を発見するための検索
- transactional query（取引型の検索）
  ソフトウェアのダウンロードやオンラインショッピングなど，Web サイトを通して何らかの実体を得ることを目的とした検索

これに対して Evans らは，近年のソーシャル化した Web の動向に対応・拡張された，社会的な検索モデル[6] を提唱している．

Web ページや検索エンジンにまつわる時間情報を利用した研究は，過去よりいくつか行われている．Amitay らは，Web のリンク構造分析でオーソリティを抽出する際に，ページ間にリンクが張られた時期という時間要素を導入することで，より時宜を得たオーソリティを抽出する試み[1] を行っている．この研究はコンテンツを作成する側を対象にしているのに対して，本論文はソーシャルブックマークのユーザというコンテンツ閲覧側の注目度に着目している点が異なる．

また Vlachos らは，Web 検索エンジンのクエリログを頻度の変化の周期性や，周期のパターンが類似するクエリ発見する一連の手法[12] を提案している．近年，Web 検索エンジンの膨大な検索クエリログが利用できるようになったことによって，クエリログからのトレンドの自動抽出や，トレンド間の関係の抽出，検索エンジン評価などを行うクエリログマイニングの研究[2),5),9),16)] は，活発に行われている．

しかしながら，クエリログを用いたアプローチによって発見できるのは，検索クエリ，つまり特定のキーワードの人気度の周期性であり，ブックマーク履歴による特定のコンテンツの周期性とは性質が異なるものである．

近年はソーシャルブックマークの特性についていくつかの研究グループによって分析が行われている．Golder らはソーシャルブックマークにおけるユーザ行動やタグがどのように使われるか，またページのブックマーク数の変化にどのようなパターンがあるかといった分析[7] を行った．彼らは実験により，ページのブックマーク数の増加率や，ユーザが使用するタグの種類の増加率に関するいくつかの典型的なパターンを明らかにした．毛受らは，ブックマークの時間情報を利用して，ページの注目度を予測する手法[15] を提案した．また，最近では，Heymann らがソーシャルブックマークの情報を Web 検索エンジンに利用するにあたって有用であろういくつかの分析[8] を行っている．Mei らは検索クエリログの情報をソーシャルブックマークのタグの情報で補完するアプローチによって，Web 検索エンジンのユーザの振舞いを調査したり，Web 検索エンジンの精度を向上させたりする手法[10] を提案した．

このようにソーシャルブックマークに関する研究は近年さかんになってきているものの，ソーシャルブックマークにおける時間情報の分析結果を，検索結果のランキングに活用するといったアプローチをとっている研究はまだあまり例がない．

我々はこれまでに，ソーシャルブックマークの際に付加されるタグや時間情報を用いた，Web 検索結果の対話的な再ランキング手法[14] を提案した．この提案の中でブックマークの

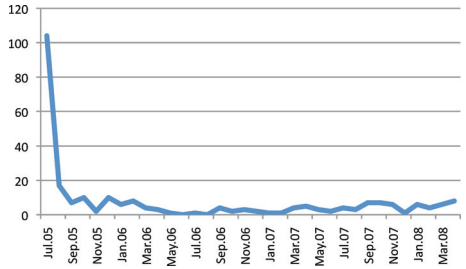

図 2　一過性のブックマークのパターン
Fig. 2　Peaky pattern of bookmark.

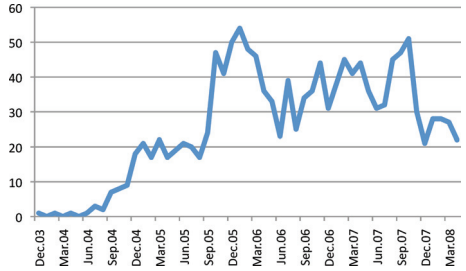

図 3　継続的なブックマークのパターン
Fig. 3　Continuous pattern of bookmark.

継続性やページの特性の一部とし，それを検索クエリのパラメータとして用いるという提案を行っている．具体的には，**図 2** のようなごく短い期間のみブックマークされる一過性のページと，**図 3** のような継続してブックマークされるページがあることに着目し，それらの特性を Web 検索におけるのクエリの一部として応用したものである．なおこれらのグラフは横軸がブックマークが行われた月，縦軸がその月に行われたブックマークの件数を示している．

本論文はこれをさらに発展させ，ブックマークの周期性に着目する．もしある特定のタイミングで有用なコンテンツがある場合，それを Web 検索エンジンのランキングに反映させることを提案するものである．

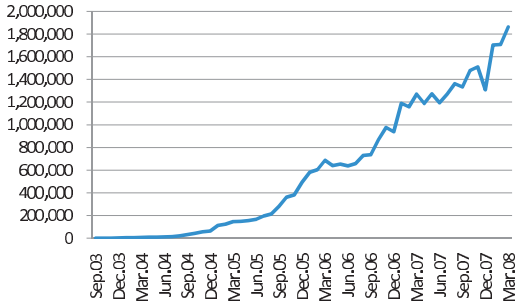

図 4　データセット全体の月ごとブックマーク数
Fig. 4　Number of bookmarks per month in dataset.

3. ブックマークの周期性分析

本章では，ブックマークの周期性と発見手法，および周期性が発生している期間の特性について分析を行う．実験を行うにあたって，ソーシャルブックマークサービスのうち，比較的早い時期（2003 年 9 月）にサービスの提供を開始した[⋆1]Del.icio.us のデータを利用した．これは，ソーシャルブックマークサービスが世に現れてから比較的日が浅い Web サービスであるため，その中でも早い時期にサービスを開始した Del.icio.us が，ページに対するブックマークの年単位の周期性を検出できる可能性が高く，都合が良いからである．

Del.icio.us では，サービス内で最近ブックマークされたページ一覧[⋆2]が公開され，随時更新されている．この一覧を定期的に確認することで，ページのブックマーク情報を 2008 年 4 月 22 日から同年 5 月 26 日にかけて収集した．この方法によるデータの収集は，Heymann らによるソーシャルブックマークの分析[8] の際にも採用されているものである．この収集方法の利点は，データ収集の期間中に Del.icio.us に投稿されるすべてのページを収集の対象にするため，サービス内で頻繁に使用されるタグで検索して得られた検索結果からページを収集するよりも，偏りを小さくおさえられることである．**図 4** は得られたデータセットのすべてのブックマークを，ブックマークが行われた月を軸にして集計したものである．

図 4 を見ると，近年の Del.icio.us のユーザ数の急速な増加を考慮すれば，2004 年から

⋆1 http://en.wikipedia.org/wiki/Delicious_(website)
⋆2 http://del.icio.us/recent?min=10

2005 年に行われたブックマークもある程度取得できていることが分かる．

この収集の結果得られたデータセットの規模は 338,018 ページ，35,406,051 ブックマークであった．ただし，収集時点でページのブックマークが 10 件に満たなかった場合は，個々のブックマークの時間情報を確かめなくても年単位の周期性を持たないのが明らかなので，あらかじめ除外してある．

3.1　ブックマークの周期性の発見

Vlachos らによって，Web 検索エンジンのクエリログにおいて，移動平均を利用して検索クエリのバーストを検出する手法[12] が提案されている．ブックマークの周期性を発見するにあたって，この手法でバーストを検出する手順はは大まかに以下のとおりである．

(1)　系列 $t = (t_1, \cdots t_n)$ に対する長さ w の移動平均 MA_tw を計算する．

(2)　閾値 $cutoff = mean(MA_tw) + x * std(MA_tw)$ を設定する．

(3)　以下の式を満たす区間をバーストと定義する．

$$Bursts = \{t | MA_tw(i) > cutoff\} \quad (1)$$

ここで，系列 i は任意のページに対するページにソーシャルブックマークが n 件あったとき，それぞれのブックマークが行われた時間を特定の粒度で集計したものである．今回の実験では使用したデータセットの制約から，ブックマークの集計の粒度に “月” を使用した．そのため，系列の要素 t_1 から t_n の実際の値はページの月別ブックマーク数である．関数 $mean()$ および $std()$ はそれぞれ系列の値の平均値と標準偏差を表している．Vlachos らの手法[12] は　バースト検出の閾値 $cutoff$ の安定化のために，その閾値として系列全体の平均値に加えて標準偏差も用いているが，x はこの標準偏差に対する重みづけパラメータである．同様に，w は移動平均のウインドウサイズであり，バースト検出の安定のため実際の系列の値ではなく異動平均した値が用いられる．MA_tw はあるページの月ごとのブックマーク数の系列 t の移動平均である．なお，今回の実験では系列 t がすでに 1 カ月の粒度で平滑化されているため，w の値は 1 を採用した．

表 1 は，この手法をデータセット中の各ページの過去のブックマークに適用したときの，検出されたバーストの回数の分布である，このとき，重みづけパラメータ x については，いくつかの値を試したところ $x = 2.5$ を使用したときに比較的多くのバーストを適切に検出できたので，この値を使用した．

なお，バースト回数が 1 回のページについては，2 つの種類がある．まずは最初のバーストから 1 年を超える日数がすでに経過したページ（1-β 型）である．これは図 2 に示した一過性のブックマークのパターンが相当する．もう 1 つは，実験時期が最初にバーストが検出されてから 1 年以内であるために，年周期のバーストがまだ存在しないページ（1-α 型）である．後者は一過性のブックマークパターンを持つページなのか，それとも前述のように毎年バーストが発生するページなのかがまだ確定していない，比較的新しいページである点で，前者と性質が異なる．少なくともこの 1-α 型のページについては，ある時点でバースト回数が 0 回のページよりは，年周期を持つバーストが発生する可能性が比較的高いと考えられる．

表 1　検出されたバースト回数とそのページ数
Table 1　Number of bursts and pages.

バースト回数	ページ数	割合 (%)
1	80,207	82.56
(1-α)	(39,186)	(40.34)
(1-β)	(41,021)	(42.22)
2	15,110	15.55
3	1,709	1.76
4	116	0.12
5	5	0.01
合計	97,147	100.00

さらに分析を進めたところ，バーストが 2 回以上検出された 16,940 ページのうち，1,297 ページは毎年同じ月にバーストが発生していることが判明した．このようなページの具体例を，次項以降で分類し解説する．

3.1.1　1 年のうち特定の時期に行われる活動に関するページ

図 5 は大学の卒業研究に関するページ[⋆1] のブックマーク数を月ごとに集計したものである．検出されたバーストの時期を見ると，米国の大学で卒業研究が佳境にさしかかる毎年 5 月にブックマークが集中していることが分かる．このような，1 年のうち特定の期間に行われる活動に起因する周期的ブックマークの例をいくつかあげる．**図 6** の，ガーデニングに関するポータルサイト[⋆2]の月別ブックマーク数の推移を示したものである．図 6 を見ると，北半球の多くの地域で，種まきなどのガーデニングの作業が発生する季節である春，つまり 3 月から 5 月に特にブックマークが集中していることが分かる．類似した背景を持つページとして Snowfrakes and Snow Crystals というページ[⋆3]があげられる．

⋆1 http://instruct1.cit.cornell.edu/courses/ee476/FinalProjects/
⋆2 http://ww.yougrowgirl.com/
⋆3 http://www.its.caltech.edu/ atomic/snowcrystals/

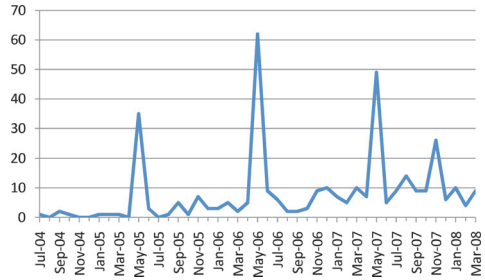

図 5　米国の卒業研究に関するページの月別ブックマーク数
Fig. 5　Number of monthly bookmarks of a page about graduation project.

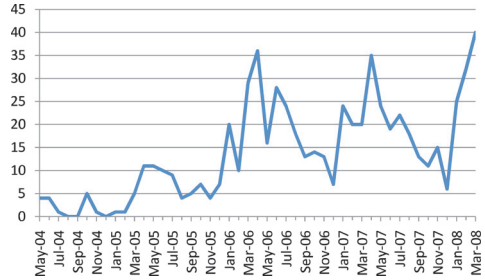

図 6　ガーデニングに関するページの月別ブックマーク数
Fig. 6　Number of monthly bookmarks of a page about gardening.

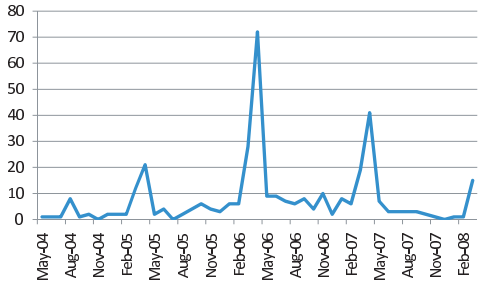

図 7　エイプリルフールに関するページの月別ブックマーク数
Fig. 7　Number of monthly bookmarks of a page about April Fool.

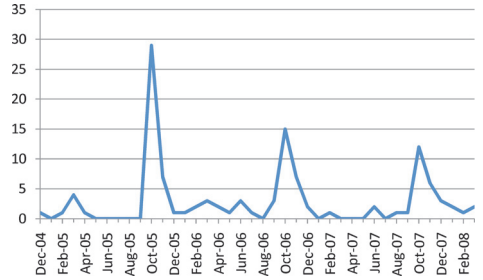

図 8　カボチャ細工に関するページの月別ブックマーク数
Fig. 8　Number of monthly bookmarks of a page about pumpkin carving.

3.1.2　一年のうち特定の日に行われる伝統行事に関するページ

図 7 は，過去にいろいろなメディアで行われたエイプリルフールのジョークをまとめたページ[⋆1]の月別ブックマーク数の推移を示したものである．図 7 を見ると，毎年エイプリルフールが話題になる 4 月に特にブックマークが集中していることが分かる．

3.1.3　伝統行事に関連する内容を扱ったページ

図 8 は，カボチャを使った彫刻に関するページ[⋆2]の月別ブックマーク数の遷移を示したものである．米国の年間行事の 1 つであるハロウィンと関係が深いといえる．また図 8 を見ると，ハロウィンが催される毎年 10 月にブックマークが集中していることが分かる．また，**図 9** は，とあるショッピングサイト[⋆3]の月別ブックマーク数の推移を示したものである．図 9 は毎年 11 月にブックマークが集中している．このページは一見，ブックマークのバーストが発生する要素がないように思えるが，実際にこのショッピングサイトを訪れると，Black Friday といわれるセールを毎年 11 月に開催しているという記述がある．類似した背景を持つページとして，Santa Claus and Christmas at Nothpole というページ[⋆4]があげられる．

3.1.4　その他の特定のイベントを扱ったページ

図 10 は，毎年同じ時期に開催される Web デザインに関するカンファレンスの Web サ

⋆1 http://www.museumofhoaxes.com/hoax/aprilfool/
⋆2 http://www.pumpkingutter.com/
⋆3 http://gottadeal.com/
⋆4 http://www.northpole.com/

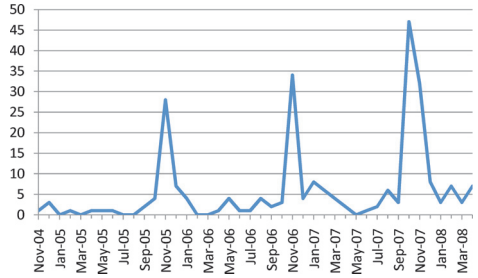

図 9　ショッピングサイトに関するページの月別ブックマーク数
Fig. 9　Number of monthly bookmarks of a page about shopping.

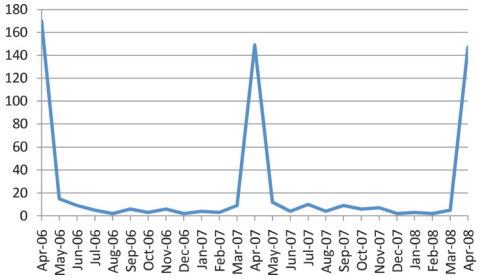

図 10　Web デザインのカンファレンスに関するページの月別ブックマーク数
Fig. 10　Number of monthly bookmarks of a page about web design conference.

イト[⋆1]の月別ブックマーク数の推移を示したものである．図 10 を見ると，カンファレンスが開催される毎年 4 月に特にブックマークが集中していることが分かる．また，**図 11** は毎年 4 月に開催される．ゴルフ大会のページ[⋆2] である．大会が開催される毎年 4 月にブックマークが集中していることが分かる．類似する背景を持つページとして，National Novel Writing Month という Web サイト[⋆3]があげられる．このように，伝統行事とまではいかないものの，毎年開かれる個別のイベントに対しても，周期的なブックマークが発生している．

⋆1 http://naked.dustindiaz.com/
⋆2 http://www.masters.org/en_US/index.html
⋆3 http://www.nanowrimo.org/

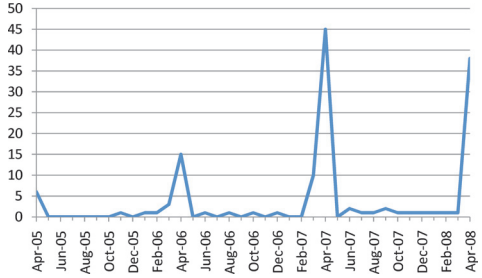

図 11　ゴルフの大会に関するページの月別ブックマーク数
Fig. 11　Number of monthly bookmarks of a page about golf tournament.

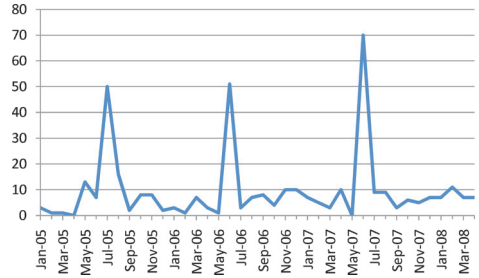

図 12　デジタルライブラリに関するページの月別ブックマーク数
Fig. 12　Number of monthly bookmarks of a page about digial library.

3.1.5　周期性が発生する背景が明らかでないページ

図 12 は，デジタルライブラリに関するサイト[⋆4]の月別ブックマーク数の推移を示したものである．図 9 は毎年 6 月または 7 月にブックマークが集中しているが，なぜこのようなブックマーク傾向を持つのかは明らかではない，同様のページとして，ペーパークラフトに関するページ[⋆5]があげられる．

以上の調査結果において，伝統行事など，暦に関係する内容を扱ったページが，年単位のブックマークの周期性が見られるのは直感的に理解しやすいといえる．しかし一方で，実際に年単位の周期性が観測されるページの中には，その背景が明らかでないものも存在する．

⋆4 http://fantastic.library.cornell.edu/index.php
⋆5 http://ravensblight.com/papertys.html

3.2　タグの周期性の発見

前節で Vlachos らのバースト検出手法を用いてページに行った調査を，本節ではデータセット全体のタグに対して行う．ユーザによるページの分類結果であるタグの周期性を発見することによって，隠れた周期性を発見したり，まだ現れていない周期性を予測したりするのに役立てることが本調査のねらいである．

この実験の結果，以下のタグが特定の月に 2 回以上バーストが発生していることが分かった．

- 伝統行事に関するタグ
  - blackfriday - 11 月
  - halloween - 10 月
  - valentine - 2 月
- 伝統行事に関係があるタグ
  - egg - 4 月（イースター）
  - pumpking - 10 月（ハロウィン）
  - santa - 12 月（クリスマス）
- 毎年同じ時期に開催されるイベントに関するタグ
  - discworld[⋆1] - 9 月から 10 月
  - nanowrimo[⋆2] - 10 月から 11 月
  - sxsw[⋆3] - 2 月から 3 月
- ある季節特有の活動や事象，旬などに関係があるタグ
  - airconditioning - 6 月
  - hitchhikers - 5 月
  - salmon - 1 月および 3 月
  - snowfrake - 12 月
- 日付に関するタグ
  - january - 1 月
  - june - 6 月
  - october - 10 月
- 背景が明らかでないタグ
  - finalist - 3 月
  - gasoline - 5 月
  - kettle - 1 月
  - pi - 3 月から 4 月

⋆1 http://en.wikipedia.org/wiki/Discworld
⋆2 http://www.nanowrimo.org/
⋆3 http://sxsw.com/

表 2　ページ “Driver Packs for XP” に付加されたタグ
Table 2　Top used tags for page “Driver Packs for XP”.

通常時	頻度	バースト時	TF·IMF 値
drivers	85	useful	5.00
windows	80	windows	1.33
unattended	42	drivers	1.17
installation	38	ubcd	1.00
software	37	computerstuff	1.00

ここで，タグの周期性の分析は，関連研究でも行われている特定のクエリログの周期性の分析と類似していると考えられるが，タグは，それをブックマークに使用したページに対して陽に結び付けられている点でクエリログと異なる．そのため，ソーシャルブックマークサービス内で特定の時期に多く使用されるタグは，最初にブックマークされてから 1 年以内，つまり表 1 における 1-α 型の，ブックマークの周期性を検出できないページに対して，周期性の発生の有無を予測する手がかりとなる可能性がある．

3.3　ブックマークの周期性とタグの使用傾向の関係分析

バーストが検出されたページにおいて，どのようなタグがバースト期間中にのみ現れているのかを調査した．**表 2** は，Windows XP のドライバを集めたページ[⋆4]に付加されたタグのうち，バースト期間，およびそれ以外の期間における上位 5 件のタグである．なお，バーストが発生した月に限り特に使用されているタグを明らかにするため，バースト時のタグについては TF·IDF のドキュメントを月に置き換えたものを TF·IMF 値としたときに，この値が高い上位 5 件を記載した．

このページのほかにもいくつかのページにおいて，バースト時には useful のようなページの印象に関するタグが用いられる場合があることが分かった．このような現象が起こる原因についてはまだ不明な点が多いが，興味深い特性を持っている可能性があり，今後詳しく調査を行う予定である．

以上の分析結果から，ページのブックマークの周期性は，Web 検索サービスの検索結果

⋆4 http://driverpacks.net/

を実社会の伝統行事などのタイミングでランキングに連動させることで，時宜を得た有用なものになる可能性が高いことが見てとれる．次章からは，この結果をふまえたブックマークの周期性による Web 検索結果のランキング手法を提案する．

4. ブックマークの周期性に基づく Web 検索結果のランキング手法

本章では，ブックマークの周期性に基づいた Web 検索結果のランキング手法を提案する．

ここで，今までの議論をもとにした検索が行われる時期と，ページのブックマークのバーストの周期性に基づいた時期連動型ランキングを行う式を定義する．

$$score(j) = N(\{m(t_j)|m(t_j) = m(t_q)\}) \quad (2)$$

ただし，$m(t_j)$ はページ j のブックマークにおいて，3.1 節で定義した式 (1) によってバースト検出された月，$m(t_q)$ は検索クエリが発行された月である．$N(\cdots)$ はページ j のバースト検出月（$m(t_j)$）のうち，検索クエリが発行された月 $m(t_q)$ と等しいものの個数を表す．

直感的には，たとえば検索を 5 月に行った場合は，ランキングのスコアはそれより過去数年間において 5 月にバーストが発生していた回数によって決定される．検索結果のすべてのページに対して上記の式を適用したうえで，$score(j)$ の値が高いページほど，検索が行われた時期において関連性が高いと判断され上位にランキングされる．これはたとえば，本論文の冒頭で述べたような状況で，過去の同じ時期に頻繁にブックマークされていることから，その時期に特に有用である可能性が高いページが上位にランキングされることを想定している．次章ではこのランキング手法の評価を行う．

なお，今回提案するランキング手法においてスコアが等しいページが複数あった場合は，元の検索結果のランキングを優先するものとする．今後このようなページに対しても陽にランキングを行う方法としては，検索結果中のより上位にランキングされたページのスニペットとの類似度を比較し，類似度が高いページ順にランキングする，擬似適合性フィードバック[11] のような手法を適用することが考えられる．

5. 評 価 実 験

本章では，前章で提案した検索結果のランキング手法に対する定性的な評価を行う．ブックマークに周期性があるページを含む検索結果において，提案手法によるランキングが既存のランキングに対してどのように変化させるか調査する．

今回はデータセットに英語のページが多く含まれていることを考慮し，検索エンジンは Google.com（英語版）を使用し，検索クエリは “black friday”，“halloween” および “world science question” を使用した．なお Black Friday とは米国で 11 月の収穫祭の次の金曜日にいろいろなお店でバーゲンセールを行うイベントで，内容的には日本における正月の初売りに近いといえる．

表 3　5 月におけるクエリ “black friday” のランキング
Table 3　Ranking of query “black friday” in May.

スコア	元順位	URL
0	1	http://en.wikipedia.org/wiki/Black_Friday_(shopping)
0	2	http://en.wikipedia.org/wiki/Black_Friday
0	3	http://www.blackfriday.info/

表 4　11 月におけるクエリ “black friday” のランキング
Table 4　Ranking of query “black friday” in November.

スコア	元順位	URL
2	3	http://www.blackfriday.info/
2	5	http://bfads.net/
0	1	http://en.wikipedia.org/wiki/Black_Friday_(shopping)

まず検索結果として，各クエリにおいて Google SOAP Search API[⋆1] を用いて検索結果の上位 200 件を取得した．次に検索結果の各ページに対して式 (4) を適用し，データセット中のブックマーク情報を用いてランキングを行った，クエリとその実行時期ごとのランキングの上位 3 件を**表 3**，**表 4**，**表 5**，**表 6**，**表 7**，**表 8** に示す．

クエリ “black friday” の 5 月時点の検索結果（表 3）では，Wikipedia など Black Friday の一般的な定義について書かれたページがトップにランキングされた．一方で 11 月のランキング（表 4）は，Black Friday の広告など，実際にショッピングに役に立つページがトップにランキングされている．

クエリ “halloween” の 10 月時点の検索結果（表 5）では，トップにランキングされたのはハロウィンの衣装を購入できるインターネットショッピングサイトであり，シーズン中に特に有用である．一方でシーズンから外れた 5 月のランキング（表 6）では，Wikipedia のハロウィンのページがトップにランキングされ，ユーザはここからハロウィンに関する一般的な情報を得ることができる．

クエリ “black friday” および “halloween” のランキング結果で共通しているのは，行事

⋆1 http://code.google.com/apis/soapsearch/reference.html

表 5　10 月におけるクエリ “halloween” のランキング
Table 5　Ranking of query “halloween” in October.

スコア	元順位	URL
2	94	http://www.buycostumes.com/
0	1	http://www.halloween.com/
0	2	http://en.wikipedia.org/wiki/Halloween

表 6　5 月におけるクエリ “halloween” のランキング
Table 6　Ranking of query “halloween” in May.

スコア	元順位	URL
0	1	http://www.halloween.com/
0	2	http://en.wikipedia.org/wiki/Halloween
0	3	http://en.wikipedia.org/wiki/Halloween_(1978_film)

表 7　1 月におけるクエリ “world science question” のランキング
Table 7　Ranking of query “world science question” in January.

スコア	元順位	URL
3	9	http://www.edge.org/
0	1	http://www.newscientist.com/lastword.ns
0	2	http://www.skytopia.com/project/science/science.html

表 8　5 月におけるクエリ “world science question” のランキング
Table 8　Ranking of query “world science question” in May.

スコア	元順位	URL
0	1	http://www.newscientist.com/lastword.ns
0	2	http://www.skytopia.com/project/science/science.html
0	3	http://www.gsfc.nasa.gov/scienceques2002/20021025.htm

に関連したページがトップにランキングされたことである．

一方でクエリ “world science question” で 1 月にトップとなるページは，**図 13** に示したとおり，その時期にブックマークのバーストが発生しているものの，ランキングされるページの内容は特に季節や伝統行事に関係するものではない．

6. 考　　察

クエリ “black friday” および “halloween” でトップに表示されたのは，行事に関連した

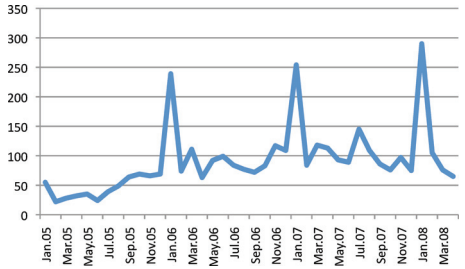

図 13　実社会の行事とは無関係なページのバーストが検出される例
Fig. 13　Example of detected annual burst which is not related to real-world event.

商品の販売や購入ができるページであった．実際，このようなページはシーズン前に特に有用であるものの，それ以外の期間の重要性は相対的に低いといえるので，手法が有効に働いたケースといえる．関連研究の冒頭でも述べたとおり，これらの検索結果は，シーズン直前では Broder による分類[4] によると取引型の検索に相当する場合が多くなると予想される．そのため提案手法は，検索が行われる時期に応じて，検索対象が注目される時期は取引型の検索に，それ以外の時期は調査型の検索に最適化していると解釈できる．

クエリ “world science question” で 1 月にトップに表示されるページについては．調査の結果，サイトの管理者がこのページ上で The World Question Center という特集を毎年 1 月に行っているためであることが判明した．このことから，表 7 のランキングは，毎年同じ時期に Web 上で行われるイベントに近い時期にユーザが関係する検索クエリを入力したとき，その存在を Web 検索エンジンの検索結果を通じて知ることができる例といえる．

ほかにも，毎年同じ時期にバーストするページの中には，なぜそのような現象が起きたのか原因が不明なものも存在した．ソーシャルブックマークはユーザのブックマーク行動によって複雑な社会的相互作用が発生するシステムであるので，このような現象がなぜ発生するのかについて，今後の調査や分析が望まれる．なお，このような問題に対する解決手段の 1 つとしては，バースト期間に集中して利用されたタグを利用することが考えられる．3.2 節では周期的に多く利用されるタグを分析し，3.3 節では，バーストが発生している期間とそれ以外の期間によって使用されるタグの違いについて調査した．これらの調査によって検出された特徴的なタグが，ページのブックマークにバーストが発生した背景を把握する手がかりになる可能性がある．

7. まとめと今後の課題

本論文では，ソーシャルブックマークの周期性発見に基づいた，時期連動型検索ランキング手法を提案し，実データを使用した試験的な評価を行った．その結果，1 年のうち決まった時期に行われる行事に関係がある，取引型の検索のランキングを最適化するのに有用である例が確認できた．一方で，ページに対するブックマークに周期性が発生する理由が明らかでないケースもあり，この点を明らかにすることは今後の課題の 1 つである．また，今回は月単位のブックマーク件数をもとに，年単位のページの需要の周期性を検出したが，今後は異なる粒度，たとえば曜日や時間帯といった単位でページの周期性の検出なども含めてより一般的な方法でバーストを検出できるように手法を改善する予定である．また，今回はブックマーク数がバーストしている時期とそれ以外の時期のタグの使用傾向の差について試験的な分析は行ったものの，提案手法である検索ランキングへの適用は行わなかった．そこで，これを検索結果のランキングへの応用することもあわせて検討していく予定である．

また，ソーシャルブックマーク以外の情報，たとえば Web 検索エンジンのクエリログなどとあわせて分析を行うことで，ブックマークのログととクエリログでお互いの情報を補完し，Web 検索エンジンとソーシャルブックマーク双方の有用性の向上に寄与することも考えられる．

謝辞　本研究の一部は，京都大学グローバル COE プログラム「知識循環社会のための情報学教育研究拠点」，および，文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」，計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」（研究代表者：田中克己，A01-00-02，課題番号 18049041），計画研究「情報爆発に対応する新 IT 基盤研究支援プラットフォームの構築」（研究代表者：安達淳，Y00-01，課題番号：18049073），ならびに，NICT 委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」（研究代表者：田中克己）によるものです．ここに記して謝意を表します．

参 考 文 献

1) Amitay, E., Carmel, D., Herscovici, M., Lempel, R. and Soffer, A.: Trend Detection Through Temporal Link Analysis, *Journal of the American Society for Information Science and Technology*, Vol.55, No.14, pp.1270–1281 (2004).
2) Arikan, I., Bedathur, S.J. and Berberich, K.: Time Will Tell: Leveraging Temporal Expressions in IR, *WSDM* (Late Breaking-Results) (2009).
3) Beitzel, S.M., Jensen, E.C., Chowdhury, A., Grossman, D.A. and Frieder, O.: Hourly analysis of a very large topically categorized web query log, *SIGIR 2004*, pp.321–328 (2004).
4) Broder, A.: A taxonomy of web search, *SIGIR Forum*, Vol.36, No.2, pp.3–10 (2002).
5) Chien, S. and Immorlica, N.: Semantic similarity between search engine queries using temporal correlation, *WWW 2005*, pp.2–11 (2005).
6) Evans, B.M. and Chi, Ed.H.: Twards a Model of Understanding Social Search, *Proc. ACM 2008 conference on Computer Supported Cooperative Work* (2008).
7) Golder, S. and Huberman, B.A.: Usage Patterns of Collaborative Tagging Systems, *Journal of Information Science*, Vol.32, No.2, pp.198–208 (2006).
8) Heymann, P., Koutrika, G. and Garcia-Molina, H.: Can Social Bookmarking Improve Web Search?, *Proc. 1st ACM International Conference on Web Search and Data Mining* (*WSDM'08*) (2008).
9) Li, X. and Croft, W.B.: Time-based language models, *CIKM 2003*, pp.469–475 (2003).
10) Mei, Q., Jiang, J., Su, H. and Zhai, C.X.: Searching and Tagging: Two Sides of the Same Coin?, UUIC Technical Report No.UIUCDCS-R-2007-2919 (2007).
11) Baeza-Yates, R. and Ribeiro-Neto, B.: *Modern Information Retrieval*, p.308, Addison Wesley Longman (1999).
12) Vlachos, M., Meek, C., Vagena, Z. and Gunopulos, D., Identifying similarities, periodicities and bursts for online search queries, *Proc. 2004 ACM SIGMOD international conference on Management of data*, Paris, France (2004).
13) Page, L., Brin, S., Motwani, R. and Winograd, T.: The pagerank citation ranking: Bringing order to the web, Technical report, Stanford Digital Library Technologies Project (1998).
14) Yanbe, Y., Jatowt, A., Nakamura, S. and Tanaka, K.: Can Social Bookmarking Enhance Search in the Web?, *Proc. 7th ACM/IEEE-CS Joint Conference on Digital Libraries* (*JCDL2007*), Vancouver, Canada (June 2007).
15) 毛受　崇，吉川正俊：ブックマークの時系列情報を利用したソーシャルブックマークにおける注目度予測，第 19 回データ工学ワークショップ（DEWS2008）論文集，B9-5 (2008).
16) Zhao, Q., S.C.H. Hoi, Liu, T.-Y., Bhowmick, S.S., Lyu, M.R. and Ma, W.-Y.: Time-dependent semantic similarity measure of queries using historical click-through data, *WWW 2006*, pp.543–552 (2006).

(平成 21 年 3 月 20 日受付)
(平成 21 年 7 月 10 日採録)

(担当編集委員　天笠 俊之)

山家　雄介（学生会員）

京都大学大学院情報学研究科博士後期課程在学中．2006 年宮城大学事業構想学部デザイン情報学科卒業．ソーシャルブックマークとメタサーチに関する研究・開発に従事．電子情報通信学会，日本データベース学会各学生会員．

中村　聡史（正会員）

京都大学大学院情報学研究科社会情報学専攻特任講師．2004 年大阪大学大学院情報学研究科博士後期課程修了．博士（工学）．主にヒューマンコンピュータインタラクション，ウェブ検索の研究に従事，日本データベース学会会員．

アダム ヤトフト（正会員）

京都大学大学院情報学研究科社会情報学専攻特任助教．2005 年東京大学大学院情報理工学系研究科電子情報学博士課程修了．博士（情報学）．主にウェブ検索，ウェブアーカイブマイニングの研究に従事．ACM 会員．

田中　克己（正会員）

京都大学大学院情報学研究科社会情報学専攻教授．1976 年京都大学大学院博士前期課程修了．博士（工学）．主にデータベース，マルチメディアコンテンツ処理，ウェブ検索の研究に従事．IEEE Computer Society，ACM，人工知能学会，日本ソフトウェア科学会，日本データベース学会各会員．