

ソーシャルブックマークを利用した Web 検索結果の対話的再ランキング 手法

山家 雄介[†] 中村 聡史[†] アダムヤトフト[†] 田中 克己[†]

[†] 京都大学情報学研究科 〒 606-8501 京都市左京区吉田本町
E-mail: †{yanbe,nakamura,adam,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし ソーシャルブックマークは、Web 利用者がブラウジング中に興味や関心を持ったページを記録、共有、分類そして発見することを支援する Web サービスである。本論文では、ソーシャルブックマークから各 Web ページのメタデータを抽出し、これを利用することによる Web 検索結果の対話的再ランキング法を提案する。我々はこの手法を実現するプロトタイプを実装し、そのインタフェースを中心に試験的な評価を行った。

キーワード メタサーチエンジン, ソーシャルブックマーク, 情報検索, ランキングアルゴリズム

Interactive Re-ranking of Web Search Results by Social Bookmarks

Yusuke YANBE[†], Satoshi NAKAMURA[†], Adam JATOWT[†], and Katsumi TANAKA[†]

[†] Faculty of Engineering, Kyoto University Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501 Japan
E-mail: †{yanbe,nakamura,adam,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract Social bookmarking services are an emerging type of Web services that let users store, share, annotate, and discover interesting resources. In this paper, we propose an interactive reranking method for web search results that utilizes social bookmarking data such as the number of bookmarks, timestamp, associated tags, sentiment expressions, etc. We demonstrate a prototype system that we have implemented and show some preliminary evaluation done in order to test our approach.

Key words meta search engine, social bookmarks, information retrieval, ranking algorithm

1. ま え が き

近年、Web のリンク構造が、ページの品質を推測する指標としてよく利用されている。PageRank [8] はリンク構造分析を用いたもっとも有名な手法である。PageRank アルゴリズムの背景にあるアイデアは、ページ間の大規模なリンク構造を明らかにすることで、ページの人気度と間接的な品質を捕らえることにある。しかし今日、リンクに基づいたアルゴリズムは多くの欠点があると指摘されている [14]。欠点としては、スパムに弱いこと、平均的なユーザにはリンクを作成すること自体が難しいこと、またリンクにはいろいろな目的や意味があり、必ずしも推薦とはいえないことなどがある。それゆえ、リンクに基づいたランキングアルゴリズムの過去の成功にもかかわらず、現在はこれらの制限に対応する新たなアプローチが必要とされている。

Web2.0 の流れに伴って、ソーシャルブックマークは Web 検索の性能を向上させる可能性をもっていると我々は考える。ソーシャルブックマークでは、あるページの人気度はそのページをブックマークしたユーザの人数ではかられる。我々はこのソ

シャルブックマークの特性に注目し、PageRank と SBRank (ソーシャルブックマーク数により導出される値) を融合した検索エンジンのランキングに関する検討を行っている [11]。

PageRank と SBRank は、どちらも一種のページの有用性を測るための尺度といえるが、評価のされ方に大きな違いがある。直感的には、我々が Web のユーザを大まかにコンテンツ作成者とコンテンツ消費者に分けたとき、PageRank は「コンテンツ作成者によるコンテンツへの評価」といえる。一方で、SBRank は「コンテンツ消費者によるコンテンツへの評価」と見なすことが出来るだろう。一般に、SBRank は PageRank よりも時間的な変動が大きい、Golder らの報告 [4] によると、ソーシャルブックマークのひとつである del.icio.us において、最初にブックマークされてから 10 日以内に最大の人気度に達するページは全体の 67% にのぼる。新しく作成されたページに対する応答性の高さに加えて、ソーシャルブックマークの優位な点はリソースに関連付けられたメタデータである。複数のユーザがページにタグ付けすることで生まれるタグのパターンは、ページの話題とその特性を表している。この情報は一般に複数のコンテンツ消費者によって作成されることから、コンテ

ンツ作成者によって与えられたメタデータよりも信頼が置けるものだと考えられる。

現時点でソーシャルブックマークは Web の検索結果すべてをカバーできるほどの規模はないが、ソーシャルブックマークにおいてブックマークされているページの数は日々増加しており、リンク構造に基づいた尺度とブックマークベースのページランキング尺度の組み合わせは、現在のところ最適な方針であると我々は考える。そこで本論文ではまず PageRank と SBRank の比較的分析に取り組む。この調査の目的は、現状の Web 検索エンジンにおけるページの品質推定とソーシャルブックマークにおけるブックマーク数などを使用した、統合型 Web 検索手法の可能性を分析することにある。我々は、ソーシャルブックマークから取得したデータに基づいていくつかの分析を行い、提案した手法の実現可能性と有効性を確認した。

また、ソーシャルブックマークにおいてブックマークされたページに付加されているタグに注目し、タグから印象などを抽出して検索結果を再ランキングする手法を提案する。このタグデータは、ページの内部に含まれていないある種の補完情報(メタデータ)として利用できる。例えば、ソーシャルブックマークではページの内容や分野などのタグを使用して分類されることが一般的だが、“便利”や“面白い”といったページの印象に関するタグを使用するユーザも存在する。このような印象情報はページを分析するだけでは取得できない。さらに、ソーシャルブックマークでは通常、個々のブックマークが行われた日時が利用可能なので、これを分析することで時間的なクエリの発行も可能になる。

次に我々の手法の有用性を模索するために実装したプロトタイプシステムについて述べる。これは PageRank と SBRank の両尺度を合成することに加え、ユーザが作成した注釈、およびページに対するユーザの振る舞いに関する一般的な分析も統合したものである。結果として、我々のシステムではユーザはページを内容、関連付けられたメタデータ、ソーシャルブックマークにおける時間的な特徴、ユーザの印象やその他の特徴に基づいて検索できる。これについて我々は提案した検索形式の有用性を示す試験的な実験の結果を示す。

2. 関連研究

すでにいくつかの議論や分析が行われている [4] [6] [12] ものの、ソーシャルブックマークはそれ自身が比較的新しい取り組みであることもあり、まだ十分に研究されているとは言えない。今までに行われた研究は主に、folksonomy という、情報を整理するための新しい分類方法に着目している [9] [10]。我々の研究と密接に関係するのは、Golder と Huberman による、ソーシャルブックマークにおけるユーザ行動、タグの使用頻度などの規則性に関する報告 [4] である。彼らはユーザが使用するタグの種類に着目し、タグを機能に応じて 7 つのカテゴリに分類した。この分析は del.icio.us のシステムに対して行われた。また、最近では Marlow らが、従来の階層型分類に対するタグによる分類の利点についての議論 [6] をしている。Wu らは、ソーシャルブックマークを例に、アノテーションが付加された Web

上のリソースに対する意味的な検索モデル [9] を提案した。しかしこれまでの研究では、リンク構造とソーシャルブックマークの尺度の比較分析も、それらを組み合わせるといった検討もなされていない。

メタサーチエンジン [3] [5] [7] は、我々の取り組みと密接に関係している。Web 上では今までに多くのメタサーチエンジンが実用化されてきた。メタサーチエンジンは、複数の情報ソースを利用することによって Web 全体に対する再現率を向上させるとともに、より新しい情報を提供可能とするものである。しかし、リンク構造とソーシャルブックマークからなる統合ランキング尺度に関してはまだ取り組まれていない。これは Web のリンク解析によるランキングとソーシャルブックマークの情報が異なる特性をもつことと、それらの比較的分析がなされていないことが理由だと思われる。本稿は、ソーシャルブックマークが持つ特性を綿密に調査し、隔たりを埋め、Web 検索に応用することを検討するものである。

3. 統計的な分析

本節では、SBRank と PageRank の両尺度を比較し、さらにソーシャルブックマークを用いて Web 検索を強化するというアプローチの有用性について議論する。

3.1 データセット

我々はデータソースとして、はてなブックマークを選択した。はてなブックマークは 2007 年 6 月現在日本において最もユーザ数が多いソーシャルブックマークサービスである。我々は SBRank と PageRank の相関分析をするため、データセット A を、印象を表すタグの分析を行うためにデータセット B を取得した。それぞれについて下記に延べる。データセット A は 2007 年 2 月 16 日にはてなブックマークから取得した。データセットを取得するために、我々はまず 742 種類のタグをタグリスト^(注1)より取得した。次に、取得したタグを用いて最近ブックマークされることが多い 18,377 個 URL のセットを取得した。いくつかの URL は複数のタグから取得されるので、重複を取り除いたところ、8,029 の URL が抽出された。それぞれのページは Tag, URL, firstDate, SBRank, PageRank というデータをもつ。ここで、firstDate はあるページがソーシャルブックマークにおいて初めてブックマークされた日付を示す。

なお、PageRank は Google ツールバーを用いて、各ページの PageRank 値を取得した。Google ツールバーはブラウザのツールバーであり、アクセスしたページの PageRank 値を表示する機能をもつ。PageRank 値は 0 から 10 までの範囲でページに対して与えられる。このとき 0 がもっともページの評価が低い値となる。

また、はてなブックマークでは「ホットエントリ」という、1 日の間にブックマークが多かったページのセットが毎日 50 件ずつアーカイブされている。我々は 2006 年 12 月 23 日にこれを過去 100 日分取得した。複数の日に渡ってホットエントリに含まれるページもあるので、取得したページから重複を取り除

(注1) : <http://b.hatena.ne.jp/t>

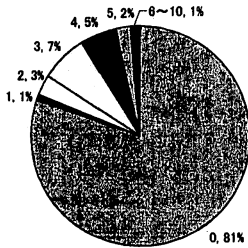


図1 PageRank 値の分布
Fig. 1 Distribution of PageRank

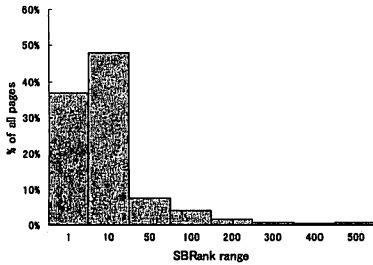


図2 SBRank 値の分布
Fig. 2 Distribution of SBRank

いたところ、3,663のページが残った。さらに、各ページでブックマークの際に付加されたタグを収集し、これをデータセットBとした。

3.2 PageRank 値と SBRank 値の分布

図1はデータセットA中のPageRank値の割合を示している。興味深いことに、多くのページのPageRank値が低く、特に、81%のページはPageRank値が0であることが判明した。このようなページを検索エンジンを用いて探し出すことは比較的困難である。低いPageRank値をもつページは、クエリに高い関連性を持っていたとしても、検索結果の上位に表示されることは難しい。

また図2はSBRank値の割合を示している。SBRank値は10以上50未満がボリュームゾーンであり、数百以上のブックマークを集めるページは少数であることが分かる。

図3はのSBRank値とPageRank値の相関を示す散布図である。ここで、SBRank値とPageRank値との間の相関係数は非常に低い値($r=0.10$)を示した。ここで、もしこの相関係数が非常に高ければ、SBRank値がPageRank値に追従していることになり、ページの品質を計るのにはPageRank値単体で十分ということになるが、実験結果がある一定の範囲に収まったことから、両尺度を組み合わせることでページの品質を推定するという枠組の可能性が確保されたことになる。

3.3 時間的な分析

本節では、ソーシャルブックマークの時間的な側面に着目する。図4は各ページが最初にブックマークされた月の分布である。この図から約9割のページが、実験日から3ヶ月以内にブックマークされたページであることが分かる。このことはソーシャルブックマーク動的な側面を表しており、ソーシャルブ

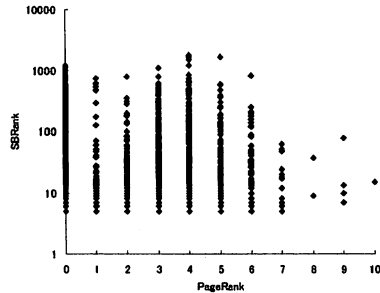


図3 PageRank 値と SBRank 値の散布図
Fig. 3 Scatter plot between PageRank value and SBRank

クマークにおけるユーザの興味深い振舞いとして見る事ができる。

図5はPageRank値が0であるページのみを対象とした分布である。前述の通り、このようなページはデータセット中の大半を占める。一般にこれらのページは他のページよりも後になってソーシャルブックマークに持ち込まれたものだと考えられる。図4および図5の結果は、GolderとHuberman[4]による、「67%のページにおいて、ブックマーク数の増加は最初にブックマークされてから10日以内にピークに達する」という報告と一致する。

一方で、リンク関係に基づくページのランキングのアプローチは新鮮な情報の検索という意味では効果的とはいえない。この理由として新しく作成されたページが他のページから被リンクを獲得するのに長い時間を要することが挙げられる[1][9]。結果として、作成されて間もないページは、仮に品質やクエリに対する適合度が高かったとしても通常の検索エンジンでは見つけるのが困難である。このような、新しいページに対する負のバイアスはBaeza-Yatesらによって確認されている[1]。これを支持する材料として、図6はfirstDateとPageRank値の散布図がある程度の負の相関($r=-0.51$)を示していることが挙げられる。

3.4 印象分析

ソーシャルブックマークの興味深い特徴のひとつとして、ユーザがブックマークした対象に対して表現された印象がしばしば含まれるということが挙げられる。このことより、ユーザがWebページに対して抱いた印象を生かした検索を可能になる。印象に基づくWebページの品質の推定の利点は多くのユーザがあるページが「面白い」「便利だ」「感動した」などの印象に基づく特性が多く回数繰り返されることによりユーザ同士の暗黙の合意の基で行われる点にある。

印象タグの個数と種類を測るために、我々はデータセットBを使用した。このデータセットに含まれるタグはGolderとHubermanによって定義されたタグの分類[4]のうち、2つのグループに大別できる。

- ページが何、または誰に関するものかを識別するタグ
- ページの品質や印象を表すタグ

以降は、これらのタグをそれぞれ内容タグおよび印象タグと記述する。これらを検出するために、データセットBについて

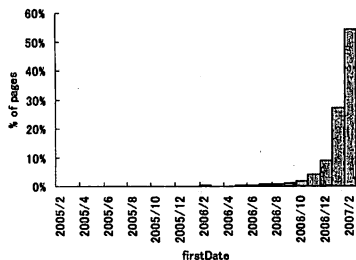


図4 PageRank値とSBRank値の散布図

Fig.4 Scatter plot between PageRank value and SBRank

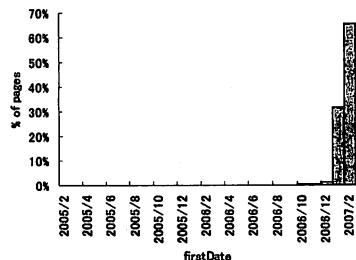


図5 PageRank値とSBRank値の散布図 (PageRank値が0のページのみ)

Fig.5 Scatter plot between PageRank value and SBRank (Pages which has PageRank equal to 0)

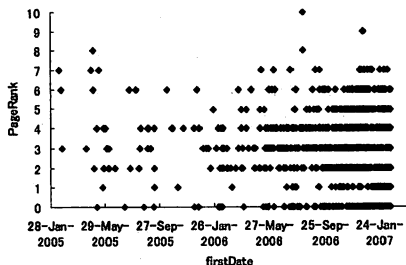


図6 firstDateとPageRank値の散布図

Fig.6 Scatter plot between firstDate and PageRank

出現頻度で上位1,100件のタグをデータセットから手で集計した。表1はこのうち上位10件の内容タグと感情タグである。この結果、内容タグは平均して感情タグよりも多く使用されていることが判明した。また、高頻度で使用されているタグの上位30件について、内容タグと感情タグの割合は約10:1であることがわかった。図7はタグの使用頻度の分布であり、上位4件のタグは非常に多く用いられているものの、それ以降はあまり使用されていないことがわかる。同義語を含めると、最もよく使用されている印象タグは「便利」、「すごい」および「ひどい」の3種類であることがわかる。

4. ソーシャルブックマークを利用した再ランキング

これまでの分析結果に基づいて、本章では我々が提案するソーシャルブックマークを用いた対話的な再ランキング法について解説する。図9は我々が実装したプロトタイプの検索結果

タグ	N	タグ	N
Web	16633	便利	5381
google	15674	これはすごい	5046
ネタ	14453	これはひどい	4123
javascript	11840	お役立ち	3041
youtube	10858	興味深い	638
tips	10784	おもしろ	616
css	9411	これは便利	544
design	8423	おもしろい	419
2ch	8381	役立ち	377
社会	7412	なるほど	365

図7 使用頻度が上位10件の内容タグと印象タグ

Fig.7 Top 10 frequent content tags and sentiment tags

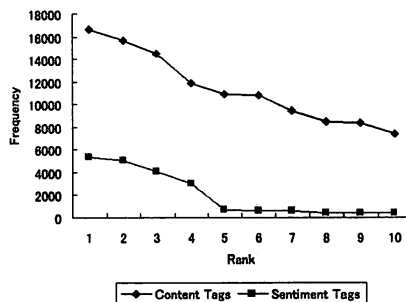


図8 内容タグおよび印象タグの使用頻度の分布

Fig.8 Distributions of content tags and sentiment tags

画面である。ユーザはスライダーを用いて各再ランキングに用いるパラメータを入力する。実際の再ランキングは図10のアルゴリズムに基づいて行われる。

4.1 SBRank値による再ランキング

SBRank値とPageRank値の重みづけによるページの人気の修正[11]は、本プロトタイプの最も基本的な機能である。パラメータ α および関数 B によって、検索結果のリランキングが行われる。これにより、PageRank値がもつ新しいページに対する負のバイアスを解消し、閲覧者の視点による評価に基づいた再ランキングが可能になる。

4.2 ブックマークの時間特性による再ランキング

ページの日付ごとのブックマーク数の分散を計算すると、短期間で多くのブックマークを集めるようなニュース的なページや、長期間に渡って継続的にブックマークされる資料的なページは特徴的な値を示す。パラメータ γ および関数 V によって、このようなページを抽出することが可能である。

4.3 印象タグによる再ランキング

印象タグで検索結果を再ランキングすることで、ユーザが特定の印象を抱いたページを抽出可能となる。先の実験結果から、実際に広く用いられている印象タグは3種類であるので、これらを類似の表現も含めて特徴ベクトル化し、ページのタグと類似度を計算することで実現している。実際のクエリは3種類の印象タグに対応するラジオボタンを選択することで行う。

5. 評価

本節では、提案手法に対する試験的な評価を行う。プロト

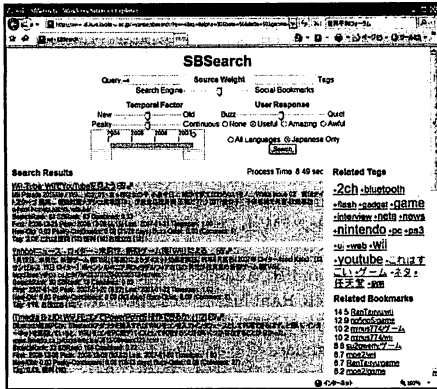


図9 プロトタイプのスクリーンショット
Fig.9 Screenshot of prototype system

$$Rank(p_i) = (1 + B(p_i)) \cdot (1 + F(p_i)) \cdot (1 + V(p_i)) \cdot (1 + C(p_i)) \cdot (1 + T(p_i, q)) \cdot (1 + T^{emo}(p_i, q)) \cdot (1 + S(p_i, t_{buz}, t_{emo}))$$

where:

$$B(p_i) = \alpha \cdot SBRank(p_i) + (1 - \alpha) \cdot SearchRank(p_i)$$

$$F(p_i) = \beta \cdot \frac{FirstDate(p_i) - \min_{1 \leq j \leq n} (FirstDate(p_j))}{\max_{1 \leq j \leq n} (LastDate(p_j)) - \min_{1 \leq j \leq n} (FirstDate(p_j))}$$

$$V(p_i) = \gamma \cdot \frac{Var(p_i, FirstDate(p_i), LastDate(p_i))}{\max_{1 \leq j \leq n} (Var(p_j, FirstDate(p_j), LastDate(p_j)))}$$

$$C(p_i) = \delta \cdot \frac{N(comment_i)}{N(comment)}$$

$$T(p_i, q) = \alpha \cdot sim(tag_i, tag_q)$$

$$T^{emo}(p_i, q) = \alpha \cdot sim(tag_i^{emo}, tag_q)$$

$$S(p_i, t_{buz}, t_{emo}) = \frac{AddBook(p_i, t_{buz}, t_{emo})}{AddBook(p_i, FirstDate(p_i), LastDate(p_i))}$$

$SearchRank(p_i)$ is a rank of p_i

$SBRank(p_i)$ is the number of bookmarks of p_i

$FirstDate(p_i)$ is the first date when a bookmark was made to p_i

$LastDate(p_i)$ is the last date when a bookmark was made to p_i

tag_i is a tag vector of p_i

tag_i^{emo} is a sentiment tag vector of p_i

tag_q is a tag vector of q

$AddBook(p_i, t_{buz}, t_{emo})$ is the number of bookmarks made to p_i in (t_{buz}, t_{emo})

$sim(tag_i^{emo}, tag_q)$ is similarity between tag vector of p_i and tag vector of q

図10 再ランキングアルゴリズム

Fig.10 Reranking algorithm

イブに対して我々の提案に含まれるいくつかの複雑なクエリを適用し、そのランキングの効果を示す。表2は各クエリに対する上位3件の検索結果およびGoogleにおける元の順位を表している。

“search:digital library SBRank:0.5 lang:all”というクエリに対する上位3件の検索結果は“Internet Archive”, “CiteCeer”および“The Online Book Page”というタイトルのページであった。最初のものは過去のWebページのデジタルアーカイブである。科学論文の電子ライブラリである。3番目はオンラインで閲覧可能な20,000以上の無料の書籍を紹介している。これらの3件の検索結果は、どれも“digital library”に関連した、

検索クエリ	上位3件	Google 順位
search:digital library lang:all	Internet Archive CiteSeer: The NEC Research Institute Scientific Literature Digital Library The Online Books Page	77 19 2
search:バンクーバー time: new lang: ja	世界平和フォーラム2006@バンクーバー 報告ブログ Mapletown バンクーバー 情報 ツアリズム・バンクーバー	98 9 1
search:wii emouseful lang: ja	Wii-Tube: WiiでYouTubeを見よう Yahoo News: Wii 米男性、新型ゲーム機「Wii」によるダイエット効果を実証 ITmedia Biz.ID: WiiリモコンでPowerPointを操作できるか(1/2)	63 80 33
search:iphone lang: ja from:20040101 to: 20061201	ITmedia News: AppleのiPhoneが半年以内に登場？ Gizmodo Japan: iPhoneの偽CMは出来が良すぎ Appleの「iPhone」のデザイン予想いろいろ - GIGAZINE	4 19 6
search:格差社会 res: buzz lang: ja	カーストという身分制度のあるインドから見た日本の格差社会: 為替王 狂童日報 - 三つの格差社会 格差社会を容認する類の弱者が強む	23 5 42

図11 検索クエリとプロトタイプによる検索結果

Fig.11 Example query and search result by prototype

参考になるページであるといえる。

次に“search Vancouver SBRank: 0.5 time:new lang:ja”というクエリに対しては、“世界平和フォーラム 2006 @バンクーバー 報告ブログ”, “Mapletown バンクーバー情報” “ツアリズム・バンクーバー”というページが上位にランキングされた。トップのページは、2006年6月にバンクーバーで開かれた、世界平和フォーラムというイベントに関するWebサイトである。このサイトにはバンクーバーに関連した比較的最近のイベント情報が含まれている。このページは検索結果はもともとGoogleで98番目に表示されていたものであり、Googleから見つけ出すのは難しいと考えられる。残り2つのページはバンクーバーの旅行ガイドといった内容であった。

3番目のクエリとして“search:wii SBRank 0.5 emo:useful lang:ja”をシステムに与えたところ、上位3件は次のとおりであった。“Wii-Tube: WiiでYouTubeを見よう”, “Yahoo News: Wii 米男性、新型ゲーム機「Wii」によるダイエット効果を実証”, “ITmedia Biz.ID: WiiリモコンでPowerPointを操作できるか(1/2)”。これらのタイトルから分かるように、検索結果はNintendo Wiiというゲーム機をユニークな方法で使用するのに便利な情報だといえる。

“search:iphone SBRank:0.5 from:20040101 to:20061201 lang:ja”というクエリに対しては“ITmedia News: AppleのiPhoneが半年以内に登場?”, “Gizmodo Japan: iPhoneの偽CMは出来が良すぎ” and “Appleの「iPhone」のデザイン予想いろいろ - GIGAZINE”という検索結果を得た。クエリの時間枠はiPhoneが正式発表される前に設定した。そのため、検索結果の上位はiPhoneに対する予想が記述されたページとなっている。これは本システムの時間による検索の一面である。

また、 “search:格差社会 SBRank:0.5 res:buzz lang:ja”というクエリに対してシステムは“カーストという身分制度のあるインドから見た日本の格差社会: 為替王”, “狂童日報 - 三つの

格差社会” and “格差社会を容認する類の弱者が望むもの”。というページを返した。これらのページには格差社会に抗議する、といった内容のもので、ソーシャルブックマークのユーザの多くはブックマークの際に何らかのコメントを残している。

6. ま と め

ソーシャルブックマークは Web 2.0 の基盤となるものである。ソーシャルブックマークで人気のあるページは、人間によってチェックされたページの一種の推薦情報として用いることができる。これらのユーザは通常のリンクに基づいたコンテンツ作成者というよりはコンテンツ利用者という側面が強い。ソーシャルブックマークにおいて利用できるデータは、動的で高い、付加されたメタデータ、時間的また感情的な情報が利用可能であるといういくつかの利点がある。

また、我々はリンクに基づいたページランキングの手法と、ソーシャルブックマークの両方を組み合わせることで機能強化された検索モデルを調査した。このような検索モデルは、現状の検索エンジンを強化する様々な手法を提供する。まず、ページの品質をはかる尺度は、ソーシャルブックマークにおける人気度と組み合わせることで向上させることができた。ユーザから与えられたタグをページのメタデータとして用いることによって、クエリに対する再ランキングの結果を向上させることができた。個々のブックマークは通常、ブックマークが行われた時刻が記録されているので、これを利用して時間を考慮したページの人気度とこれに基づく時間的なクエリを構築した。最後に、ソーシャルブックマークのタグを用いることによって、ページに対してユーザが抱いた印象や感情、特性、話題性といったもので検索結果のランキングを修正することができた。

本論文ではまず、実際のソーシャルブックマークに対して複数の視点から統計的な調査を行い、特性を洗い出した。我々は、現在のリンクに基づいたランキングアルゴリズムの制限や、ソーシャルブックマークが示す Web における社会的な相互作用の可用性を考えると、我々が採ったアプローチに類する検索システムは将来的に一般的になってくると予想する。今後、我々は今回とは異なるデータセットでより大規模な実験を行い、より正式な評価を行う予定である。これに加えて、複数のソーシャルブックマークを統合したメタサーチの手法を調査する予定である。

7. 謝 辞

本研究の一部は、文部科学省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」、異メディア・アーカイブの横断的検索・統合ソフトウェア開発（研究代表者：田中克己）、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）、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）および、文部科学省科学研究費補助金若手研究 (B) No.18700111, No.18700129 によるものです。ここに記して謝意を表するものとします。

文 献

- [1] R. Baeza-Yates, C. Castillo and F. Saint-Jean. Web Dynamics, Structure and Page Quality. In M. Levene and A. Poulouvassilis (eds.) "Web Dynamics", Springer, pp. 93-109, 2004.
- [2] J. Cho, S. Roy, and R. Adams, "Page Quality: In Search of an Unbiased Web Ranking," In Proceedings of SIGMOD Conference 2005, pp. 551-562
- [3] C. Dwork, R. Kumar, N. Naor and D.Sivakumar. Rank Aggregation Methods for the Web. Proceedings of 10th World Wide Conference 2001
- [4] S.A. Golder and B.A. Huberman, The Structure of Collaborative Tagging Systems, Journal of Information Science, 2006
- [5] Y. Lu, W. Meng, L. Shu, C. Yu and K. Liu. Evaluation of Result Merging Strategies for Metasearch Engines. Proceedings of VLDB' 05, 2006, pp.141-150.
- [6] C. Marlow, M. Naaman, D. Boyd and M. Davis, "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read", Proceedings of ACM HyperText 2006 Conference, 2006
- [7] W. Meng, C. Yu, and K.-L. Liu, "Building efficient and effective metasearch engines," ACM Computing Surveys, 34(1), 2002, pp. 48-84.
- [8] 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
- [9] X. Wu, L. Zhang and Y. Yu. Exploring Social Annotations for the Semantic Web. World Wide Web Conference 2006, 2006.
- [10] H. Wu, M. Zubair and K. Maly. Harvesting Social Knowledge from Folksonomies. Proceedings of ACM HyperText 2006 Conference, 2006.
- [11] 山家 雄介, 中村 聡史, Adam Jatowt, 田中 克己, ソーシャルブックマークの特性を利用した web 検索のランキング精度向上, 日本データベース学会 Letters(DBSJ Letters) vol. 6, No. 1, 2007 年 6 月
- [12] L. Zhang, X. Wu and Y. Yu. Emergent Semantics from Folksonomies: A Quantitative Study Journal on Data Semantics VI, LNCS 4090, 2006, pp.168-186.