

閲覧者によるハイパーリンクの追加を実現する 後付け型リンクアンカ設定システムの提案

A web-based system allowing viewers to add hyperlink anchors on web pages

坂本 竜基^{*1} 小暮 潔^{*1} 伊藤 禎宣^{*1*2} 武田 英明^{*1*3*4}
Ryuuki Sakamoto Kiyoshi Kogure Sadanori Ito Hideaki Takeda

^{*1}ATR 知識科学研究所 ^{*2}東京農工大学
ATR Knowledge Science Laboratories Tokyo University of Agriculture and Technology
^{*3}総合研究大学院大学 ^{*4}国立情報学研究所
The Graduate University for Advanced Studies National Institute of Informatics

This paper describes a system which allows web surfers to set hyperlink anchors while viewing web pages. Hyperlinks connecting web pages shows an importance of each web page. This information is used by Web search services. Current hyperlinks on Web pages, however, are set by their authors. It means that the importance of hyperlink includes only their author's knowledge or intention which is relationship among web pages. Under this system, not only authors but also visitors can set the hyperlinks through web browsers, and thus the hyperlink will include the audiences' knowledge much like Folksonomy. In this paper, we explain the system's architecture and show a brief result of an experiment on JSAI2006.

1. はじめに

通常の Web における HTML による文章コンテンツは、著者によって作成、サーバ上に設置され、不特定多数の閲覧者がこれを閲覧するため、著者と閲覧者が分離していると言える。これに対して、掲示板や Wiki、Folksonomy といったサービスでは、主たる著者が規定できず、不特定多数の閲覧者がそのまま共著者の一部としてコンテンツ作成に参画する形態をとっている。一般的に著者よりも閲覧者のほうが圧倒的多数であるため、その閲覧者群の創造性や生産性がコンテンツに反映されれば、情報の量や多様性は充実しやすい。膨大な情報の分類をソーシャルブックマークサービスをはじめとする Folksonomy は、この大量の人間がおこなった結果のもつ量と多様性を利用したサービスといえる。ここで人間がおこなう作業は、タグと呼ばれる二次データの入力であり、一次データの処理結果を主とする検索エンジンでは得ることが難しい情報の取得を実現している。

一方、検索エンジンも単に一次データである Web ページ（以後、ページと呼ぶ）に含まれる字句を解析するだけでなく、他の情報も利用してその精度向上を図っている。この代表的なものは、コンテンツ同士の関係情報であり、PageRank アルゴリズムが有名である。これは、各ページにおけるリンクアンカは著者がなんらかの意味をもって記述したものであり、A タグによる他の HTML 文章へのハイパーリンク（以下、リンクとする）先のページへの一種の投票と捉え、全体的なリンクの入出次数による統計的情報を検索結果に反映させるものである。

本稿では、リンクアンカを閲覧しているページの任意部分に誰でも設定可能とするシステムを提案する。これは、ソーシャルブックマークサービスにおいて不特定多数の人物がそのページを示すタグを入力するように、あるページに訪れる不特定多数の閲覧者が PageRank においてそのページからの投票とみなされるリンクアンカを設定可能とする環境であり、タグの代わりにリンクを用い、なおかつ対象ページ上に直接埋め込むことが可能な Folksonomy システムと換言できる。本稿では、

連絡先: 坂本竜基, ATR 知識科学研究所, 京都府相楽郡精華町光台 2-2-2, +81 774 95 1401

このシステムの概要説明と JSAI2006 で簡易的運用をおこなった結果を紹介する。

2. 後付け型リンクアンカ設定システム

本システムの目的は、閲覧者によるリンクアンカの設定を実現することにある。この閲覧者によって設定されたリンクアンカのことを、通常の著者によって設定されたリンクアンカを区別するため「後付リンク」と呼ぶ。閲覧者は、後付リンクのリンク先として、任意の URL を入力することも、システムが推薦する関連ページからリンク先を選択することも可能である。後付リンクのリンクアンカは直接 HTML を書き換えて A タグを静的に埋め込む破壊的設定と、元の HTML は変化させず JavaScript によって動的にリンクを設定する、非破壊的設定が考えられる。現在のところ、後者の非破壊的設定が実装されているが、検索エンジンのクローリングに対応するため、破壊的設定機能も実装する予定である。

まず、本システム利用の準備として、ユーザによる後付け型リンクアンカの設定を許可するページの管理者は、対象 HTML 内に JavaScript コードを埋め込む必要がある。このコードは、HTML 内に直接記述してもよいし、JSONP(JSON with padding) によってサーバから間接的に読み込みをおこなってもよい。この場合は、実際にその HTML へ記述するのはコードではなくコードの URL となる。ユーザは、このコードが記述・実行されるページにおいて、マウス操作によって新たなリンクアンカを設定することが可能となる。コードは、ユーザのマウス操作による選択文字列部分の認識とリンク先登録用のフォームの表示、及びクライアント-サーバ間の通信を主に担当する。この通信は Ajax (Asynchronous JavaScript + XML) によっておこなわれる。

ユーザは後付リンクを設定したいページ上の文字列をマウスのドラッグ操作によって指定する。ここで、選択された文字列部分は XPath[Clark 99] と共に直ちにサーバ側に送信され、そのページのリンク先として推薦される複数のページが導出される。この推薦ページはクライアント側に戻されてブラウザ上に後付リンク先の候補としてタイトルが一覧表示される。

ユーザが、このうちから一つを選択すると、その文字列の箇所が A タグによってリンクアンカとされ、後にそのページを閲覧するすべてのユーザに提示されることとなる。推薦ページ以外のページをリンク先として設定したい場合は、入力フォームにその URL を入力することで設定可能となる。

リンク先として推薦されるページの導出には、様々な方法が考えられる。次章で紹介する実験では、対象となるページ群の規定とユーザ認証が可能であったため、tf・idf による重み付けをおこなったキーワードマッチングによるページ内文章の関連度からの導出や、ソーシャルフィルタリング [Resnick 94] によるユーザ間の関連度による導出をおこなった。

設定された後付リンクは、サーバ上にリンク元とリンク先のページの URL、選択文字列のセットで保存される。ユーザは、サーバ上に設けられた検索用インタフェースから、あるページへリンクされている/しているページを検索できる。また、選択文字列を Folksonomy のタグのように検索する機能も提供している。

3. 実験

提案システムの検索機能を省略したプロトタイプシステム [Matsuoka 06] を第 20 回人工知能学会全国大会 (JSAI2006) において運用した結果を示す。JSAI2006 は発表件数が 200 件を超える学術会議であり、大会のホームページでは会議参加に対する利便性向上のためのサービスを各種提供していた。このうちのひとつが、各発表のタイトル、概要、著者などの情報を提示する発表情報を示すページの提供であり、各発表一つに対して一つの紹介ページが存在していた。この各紹介ページ内に提案システムの JavaScript コードの埋め込みをおこない、ユーザがタイトルか概要などといったページ内の文字列を選択することで推薦ページが提示される環境を作った。

推薦ページの導出は、ページ間の静的な類似性、ユーザ間の類似性、後付リンクにおけるリンクアンカの語の共起性などによる導出をおこない、推薦ページの一覧として表示した。また、過去に後付リンクとして設定された文字列はページが読み込まれたタイミングで下線を表示し、そこにマウスカーソルを移動させると、過去にどの推薦ページが選択されたのかが判別できるインタフェースを提供した。なお、任意ページへのリンク設定機能は省略したため、JSAI2006 以外のページへのリンク設定は事実上不可能であった。

システムは、大会が始まる数週間まえから運用が開始され、大会が終わるまでの間のデータを取得した。結果、この間に設定された後付リンクの件数は 156 件であった^{*1}。一ページあたりの後付リンクの件数のヒストグラムを表 1 に示す。また、後付リンクによって設定されたリンク先のページの一ページあたりの被リンク数を表 2 に示す。なお、0 件となっている件数は、後付リンクアンカが、被リンクのどちらかが存在しているページの件数であり、どちらでもないページは除外されている。

後付リンクのアンカとして指定された文字列は「クラスタリング」、「マイクロコミュニティ」といった単語が多かったが、通常のリンクでは少ない「情報の理解とそれへのアクセスを支援」、「脳波の変化と Web ページデザインの関係」等といった長い表現への設定も散見された。これは、本実験では選択文字列は形態素解析エンジンによって解析された単語とのマッチングに基づく推薦エンジンが多かったため、多くの文字列を選択

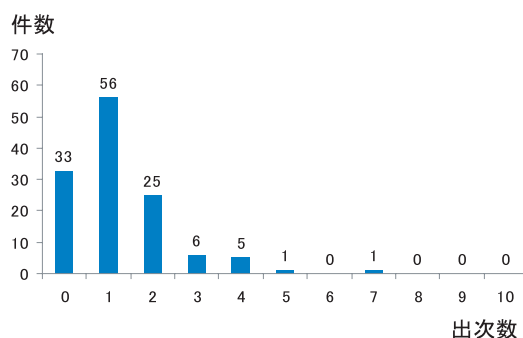


図 1: 後付リンクが設定されたページのリンク数

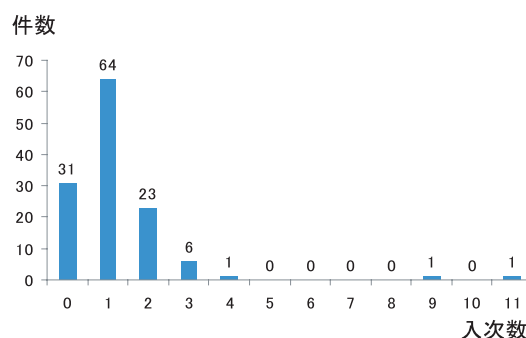


図 2: リンク先として指定されたページの被リンク数

すると、それだけ多くのページが推薦される結果となったことに起因すると考えられる。

4. まとめ

本稿は、ページの閲覧者によるリンクアンカの設定を実現する後付型リンク設定システムを提案した。このシステムにより、Web ページは、本来著者だけが設定可能であった他の Web ページへのハイパーリンクを、不特定多数の閲覧者も設定可能にする環境となり、閲覧者の意図がリンクアンカとして残る。今後は、システム運用による Web ページの関係性の変化を分析する予定である。

謝辞 本研究の一部は情報通信研究機構の委託研究により実施したものである。

参考文献

- [Clark 99] Clark, J, and DeRose, S. : XML Path Language (XPath) Version 1.0, <http://www.w3.org/TR/1999/REC-xpath-19991116> (1999).
- [Resnick 94] Resnick, P, Iacovou, N, Suchak, M, Bergstrom, P, and Riedl, J. : GroupLens: An Open Architecture for Collaborative Filtering of Netnews, Proc. ACM Conf. on Computer Supported Cooperative Work, pp.175-186 (1994).
- [Matsuoka 06] Matsuoka, Y, Sakamoto, R, Ito, S, Takeda, H, and Kogure, K. : Social tagging using marked strings in web pages, SAAW2006 poster (2006).

*1 開発者らの利用件数は除外