

選択文字列を用いた Web ページ推薦システムでの ユーザ参加型リンクアンカ付与機能の実証実験による評価

Analysis on Experiment of User-Participatory Creation Link Anchors
in Web Page Recommendation System Based on Marked Strings

松岡 有希*1*2

Yuki Matsuoka

坂本 竜基*3

Ryuuki Sakamoto

伊藤 禎宣*3*4

Sadanori Ito

武田 英明*1*2*3

Hideaki Takeda

小暮 潔*3

Kiyoshi Kogure

*1総合研究大学院大学

The Graduate University for Advanced Studies

*2国立情報学研究所

National Institute of Informatics

*3株式会社国際電気通信基礎技術研究所

Advanced Telecommunications Research Institute International

*4東京農工大学

Tokyo University of Agriculture and Technology

We propose a social navigation system called *Aikuchi* that enables users to link from a marked string in a web page to another web page and share it with others. *Aikuchi* highlights a marked string in the web page. It functions as a link anchor that includes two kinds of links: *footprint links* that users have linked from marked strings in web pages to other web pages and *recommendation links* suggested by the system based on four algorithms. Users can jump from the link anchor on the web page to other web pages. We presented the system at a conference. Based on analysis of user logs, users preferred Web page recommendations based on strings marked by other users to those based on page similarity using TFIDF and collaborative filtering. Users also preferred footprint links to recommendation links. We found that sharing links that users attached from marked strings in web pages to other web pages is useful for social navigation.

1. はじめに

人は Web 上で情報を探るとき、一般的に Google^{*1}や Yahoo!^{*2}のような検索エンジンを利用する。そこで得られる情報は、検索クエリが含まれる Web ページの一覧である。この場合、人は Web ページ一覧の中からほしい情報を探し出さなければならない。一方で、Amazon.com^{*3}のようなソーシャルナビゲーションシステムでは、ユーザが商品のページにアクセスすると、以前他のユーザが購入した履歴に基づいて関連する商品を知ることができる。“ソーシャルナビゲーション”とは、前にユーザが残した足跡（行動履歴）を現在のユーザが利用することである[?]。ソーシャルナビゲーションを利用したシステムを使うと、ユーザは他人の興味や知識に気づいたり、他人の意見に基づいて関連する情報の推薦を取得したり、コミュニケーション手段の一種類として Web ナビゲーションを利用することが可能になる[?]

本研究では、ユーザが残した足跡として、ユーザが Web ページ内で選択した文字列、すなわちマーキングを付与した文字列に着目した。以降、マーキングを利用したソーシャルナビゲーションシステム「合口」について述べ、人工知能学会全国大会 (JSAI2006) においての運用で得られたデータを使って分析した結果について述べる。

2. 合口

我々はソーシャルナビゲーションシステム「合口」において、ユーザが残す足跡としてマーキングを利用した。人々が本などの書籍を読むとき、興味を持った文字列に下線を引くこと、す

なわちマーキングを付与することがある。人はそれぞれ異なる観点を持っているので、一つのページにおいてマーキングを付与する文字列が異なることがある。複数のユーザ間でマーキングが付与された文字列を共有することによって、彼ら自身では見つけることのできなかつたページを見つけてことができるかもしれない。

合口はユーザが発表ページ内の文字列をマウスカーソルで選択した際 (図 1)、4 種類の推薦アルゴリズムに基づいて算出された他のページへの推薦リンク (発表ページのタイトル) が書かれた小窓を表示する (図 2)。ユーザが提示された推薦リンクの中から気に入ったものをクリックをすると、クリック先のページへ遷移すると同時に、合口はユーザが選択した文字列をハイライト表示することで、リンクアンカにする (図 3)。このハイライト表示された文字列のことをマーキング文字列と呼ぶ。発表ページに付与されているマーキング文字列をユーザがマウスカーソルでなぞると、合口は足跡リンクと推薦リンクが書かれた小窓を表示する (図 4)。足跡リンクは、以前ユーザがマーキング文字列から他のページへ遷移したことのあるリンクで、推薦リンクはシステムが 4 種類の推薦アルゴリズムに基づいて推薦したリンクである。

我々は、マーキングされた文字列を共有することがソーシャルナビゲーションに役立つのかどうかを調べるために、4 種類の推薦アルゴリズムを用意した。用意した推薦アルゴリズムは下記のとおりである。

- tfidf を使ったページ間類似度による推薦
- 発表ページに付与されたマーキング文字列の数を使った協調フィルタリングによる推薦
- ユーザが選択した文字列内の語とマーキング文字列内の語とのマッチングによる推薦
- ユーザが選択した文字列内の語と発表ページ内の語とのマッチングによる推薦

連絡先: 松岡 有希, 国立情報学研究所, 〒 101-8430 東京都千代田区一ツ橋 2-1-2, 03-4212-2681, m-yuki@nii.ac.jp

*1 <http://www.google.com>

*2 <http://www.yahoo.com>

*3 <http://www.amazon.com>

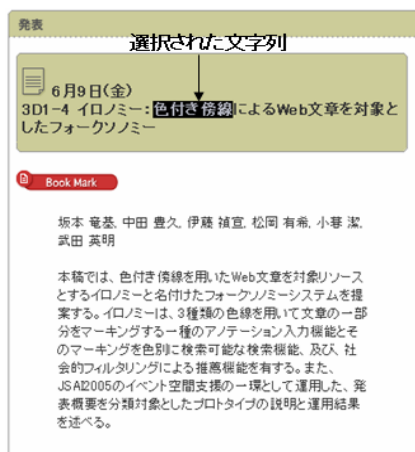


図 1: ユーザが Web ページ内の文字列を選択する

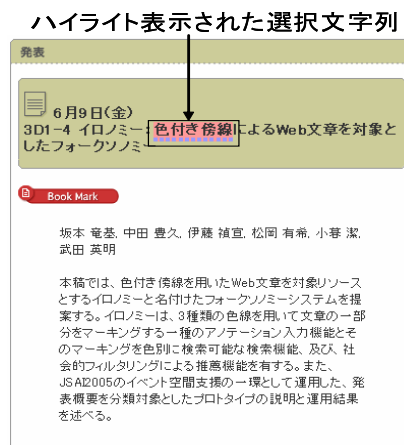


図 3: 選択文字列をハイライト表示する

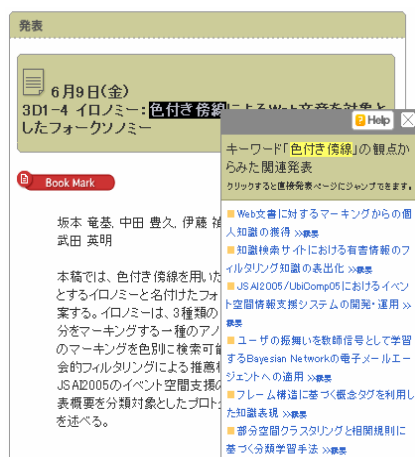
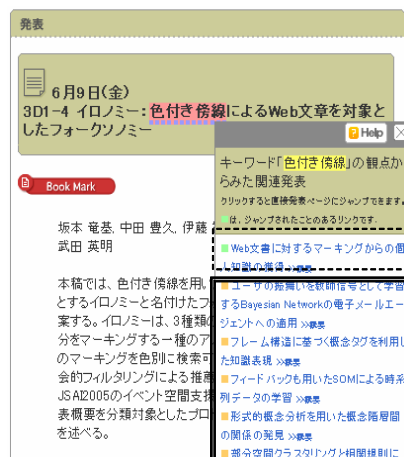


図 2: 推薦リンクが書かれた小窓を表示する



----- 足跡リンク
 ——— 推薦リンク

図 4: 足跡リンクと推薦リンクが書かれた小窓を表示する

アルゴリズム A では、ユーザがマウスカーソルで文字列を選択した発表ページに対して、tfidf[?] を用いた類似度が高い他の発表ページを推薦する。[?] より、tfidf 値の高い語はユーザがマーキングする語である可能性が高い語ということが分かった。従って、ユーザは tfidf を使ったページ間類似度の高い発表ページを好む可能性がある。ページ間の類似度は、発表ページ内の文章を茶筌 [?] を使って形態素解析し、得られた語の中から名詞の tfidf 値を求めた後、発表ページ間の類似度をコサイン類似度を使って計算した。ここでは、ユーザが選択した文字列とは関係なく、ユーザが選択した発表ページに対して、類似度の高い発表ページを推薦する。

アルゴリズム B では、協調フィルタリング [?] を用いて、ユーザにとって類似度の高い発表ページを推薦する。ユーザがマーキング文字列を付与した発表ページのうち、同じ発表ページにマーキングを付与したことがあるユーザ同士は似た嗜好を持っている可能性が高い。そこで、ユーザによる発表ページへの評価値を、発表ページ内にユーザが付与したマーキング文字列の数とし、協調フィルタリングによる計算を行った。

アルゴリズム C では、ユーザが発表ページ内で選択した文字列内の語と他の発表ページに付与されているマーキング文字列内の語とのマッチングを行ってマッチしたページを推薦す

る。他の発表ページに付与されているマーキング文字列は全ユーザによるものである。自分が付与したマーキング文字列だけでなく、他人が付与したマーキング文字列もマッチングの対象にすることで、マーキングの共有がページナビゲーションに役立つのかを試している。

アルゴリズム D では、ユーザが発表ページ内で選択した文字列内の語と他の発表ページ内の語とのマッチングを行ってマッチしたページを推薦する。ここでは、ユーザが発表ページ内で選択した文字列内の語を検索クエリとし、他の発表ページ内に含まれているかどうかを調べている。一般的にユーザが Web ページを探すのに最も利用するのは検索エンジンであるため、検索エンジンで行われることと同じ手法を推薦に取り入れた。

各推薦アルゴリズムにおいて、選択文字列 (ユーザがマウスカーソルで選択した文字列) や他の発表ページに付与されたマーキング文字列、発表ページ内の文字列を使用するかどうかを表 1 にまとめた。

本研究では、ユーザがこれらの推薦アルゴリズムのうち、ど

表 1: 各推薦アルゴリズムにおいて使用する文字列の比較

推薦アルゴリズム	選択文字列	マーキング文字列	発表ページ内の文字列
A	×	×	
B	×		×
C			×
D		×	

のアルゴリズムを好むのかについて調べた。アルゴリズム A のページ間の類似度はシステムの運用前にあらかじめ計算しておき、その他の推薦アルゴリズムに関しては合口の運用中に動的に計算した。合口はユーザが発表ページ内の文字列をマウスカーソルで選択すると、各アルゴリズムにつき最大 2 つのページを推薦し、表示順位はランダムに並べた。ユーザにはこれらの推薦アルゴリズムや表示方法については知らせていない。ユーザがどの推薦アルゴリズムを好んだのかについては、ユーザが合口によって推薦された発表ページへのリンクをクリックした時に、そのリンクを推薦するのに用いたアルゴリズムを好んだと判断した。

2.1 運用結果

合口が対象とした Web ページは論文のタイトルや発表者や概要を含む発表ページで、全部で 276 ページあった。合口は JSAI2006 の開催前から運用しており、分析対象としたデータは、2006 年 5 月 22 日～6 月 9 日までの運用によって得られたデータである。運用の結果、45 人のユーザが 1 回は発表ページ内の文字列を選択し、そのうち 28 人が提示されたリンクをクリックした。また、88 人のユーザが 1 回は発表ページ内のマーキング文字列をマウスカーソルでなぞり、そのうち 33 人が提示されたリンクをクリックした。

2.2 推薦アルゴリズムの選好結果

学会会期前と会期中でユーザが発表ページ内の文字列をマウスカーソルで選択したときに、各推薦アルゴリズムによって推薦された発表ページの数を表 2 に、各推薦アルゴリズムによって推薦された発表ページのうちユーザが選択した発表ページ数を表 3 に示している。これによると、会期前に推薦された発表ページのうち、最も使われた推薦アルゴリズムは A で、次に多いのは推薦アルゴリズム D である。一方で、推薦アルゴリズム B と C は推薦に使われた回数が少ない。これは、合口運用の初期段階のため、ユーザによって付与されたマーキング文字列の数が少ないからである。図 5 は、各推薦アルゴリズムによって推薦された発表ページのうちユーザが選択した割合 (ユーザが選択した発表ページ数/システムが推薦した発表ページ数 × 100) を示している。これによると、会期前のユーザは推薦アルゴリズム A によって推薦された発表ページを一番好んでいるが、他の推薦アルゴリズムによるものとほとんど差はない。

一方で会期中になると、ユーザは推薦アルゴリズム C と B によって推薦された発表ページを好むようになった。推薦アルゴリズム B によって推薦された発表ページが選ばれなくなったのは、各ユーザが付与したマーキング文字列のデータが少ないため、同じページばかり推薦されてしまうようになったからである。図 5 より、ユーザは tfidf を用いたページ間類似度や協調フィルタリングを用いた推薦よりも、ユーザが選択した文字列を用いた推薦 (表 1) を好むことを示している。中でも、ユーザは推薦アルゴリズム C によって推薦された発表ページ

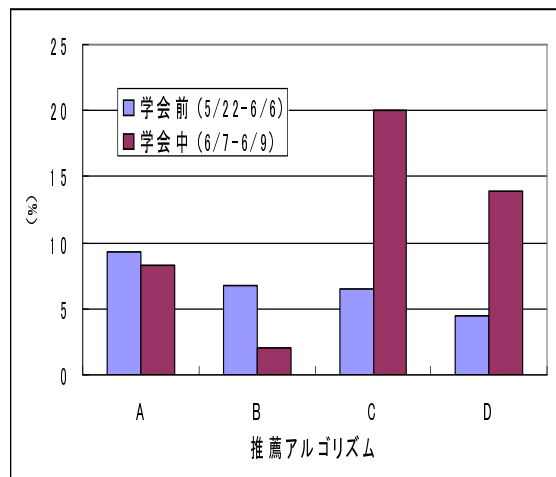


図 5: 各推薦アルゴリズムによって推薦された発表ページのうちユーザが選択した割合

を一番選んでいることから、マーキングが付与された文字列を利用したページナビゲーションが好まれたということが言える。

表 2: 各推薦アルゴリズムによって推薦されたページ数

	A	B	C	D
学会前 (5/22-6/6)	288	132	107	222
学会中 (6/7-6/9)	84	50	55	79

表 3: 各推薦アルゴリズムによって推薦された発表ページのうちユーザが選択したページ数

	A	B	C	D
学会前 (5/22-6/6)	27	9	7	10
学会中 (6/7-6/9)	7	1	11	11

2.3 ユーザが生成したリンクとシステムが推薦したリンクの選考結果

合口では、ユーザがマーキング文字列をマウスカーソルでなぞると、足跡リンクと推薦リンクが書かれた小窓を表示した (図 4)。このとき、ユーザは足跡リンクと推薦リンクのどちらを好んだのかについて調べた。表??は合口が推薦リンクや足跡リンクとして推薦したページ数を、表??は推薦リンクや足跡リンクとして推薦された発表ページのうちユーザが選択したページ数を示している。推薦リンクや足跡リンクとして推薦された発表ページのうちユーザが選択した割合を示した図??によると、ユーザは足跡リンクにより推薦されたページを好んでいることが分かる。ユーザはシステムが推薦したリンクよりも他のユーザによって付与されたリンクを信頼すると言える。従って、Web ページ内でマーキングされた文字列から他ページへ張られたリンクをユーザ間で共有することはソーシャルナビゲーションにとって役立つことが分かった。

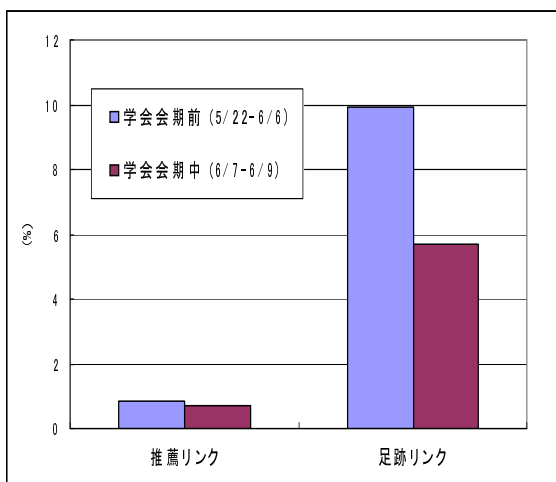


図 6: 推薦リンクや足跡リンクとして推薦された発表ページのうちユーザが選択した割合

表 4: 推薦リンクや足跡リンクとして推薦したページ数

	推薦リンク	足跡リンク
学会前 (5/22-6/6)	2302	534
学会中 (6/7-6/9)	1316	298

表 5: 推薦リンクや足跡リンクとして推薦された発表ページのうちユーザが選択したページ数

	推薦リンク	足跡リンク
学会前 (5/22-6/6)	20	53
学会中 (6/7-6/9)	9	17

3. まとめ

本研究では、ソーシャルナビゲーションにおいてユーザが残す足跡として、ユーザが Web ページ内でマーキングを付与する行為を利用した。JSAI2006 における合口の運用結果の分析により、tfidf を用いたページ間類似度や協調フィルタリングによるページ推薦よりも、マーキングされた文字列内の語のマッチングによるページ推薦が好まれることが確認できた。また、他のユーザがページ内に残したマーキングによるリンクがシステムによって推薦されたリンクよりも好まれることも確認できた。すなわち、マーキングが付与された文字列はユーザ間で共有可能であるため、ソーシャルナビゲーションにおいて有効に働くことが分かった。

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

- [Dieberger 00] A. Dieberger, et al.: Social navigation: techniques for building more usable systems, interactions, Vol. 7, No. 6, pp. 36-45, 2000.
- [Barra 02] Maria Barra and Paul Maglio and Alberto Negro and Vittorio Scarano, GAS: Group Adaptive System, Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, 2002
- [Matsuoka 06] 松岡有希, 坂本竜基, 中田豊久, 伊藤禎宣, 武田英明: 論文概要に対する色付きアンダーライン付与システムの運用・分析, DEWS2006 論文集 (2006).
- [Matsumoto 03] 松本裕治他, 「形態素解析システム『茶釜』 version 2.3.3 使用説明書」, 奈良先端科学技術大学院大学, 2003.8
- [Salton 91] Salton, G.: Developments in automatic text retrieval, Science, Vol. 253, pp. 974-980 (1991).
- [Rensnick 94] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, " GroupLens: An Open Architecture for Collaborative Filtering of Netnews, " Proc. ACM Conf. on Computer Supported Cooperative Work, pp.175-186, Chapel Hill, North Carolina, U.S.A, Oct 1994.