

# グループ支援型 Web 閲覧における 閲覧履歴の視覚化と共有 Visualization and Sharing of Web History on Group-Based Web Browsing

伊豆 陸<sup>▼</sup> 中島 伸介<sup>♦</sup>  
田中 克己<sup>▲</sup>

Atsushi IZU Shinsuke NAKAJIMA  
Katsumi TANAKA

Web 情報検索は、単に個人のための情報収集にとどまるものではない。個人が Web 閲覧をする事で得た Web コンテンツをグループで共有する Web 利用法が求められている。他のユーザから推薦された Web コンテンツを共有するような状況では、このコンテンツが選ばれた背景を知る事は重要である。この背景を理解することができれば、ユーザの閲覧範囲がいかに目的に対して網羅的であるか、また、推薦されたコンテンツがいかに閲覧した中から精選されているかについて知る事ができるからである。ここで、この選ばれた背景とは、ユーザの Web 検索目的や、閲覧履歴における類似 Web ページの閲覧数、Web コンテンツの閲覧範囲等の情報である。本論文では、Web ページが推薦されるまでの履歴をグラフによって可視化する事により Web 探索の網羅度や精選度を表現し、共有する手法を提案する。

Recently, information retrieval on Web browsing is not only for single user but also for multiple users. Then, it is required methods for sharing Web contents that a user obtains through the Web navigation. It is important to understand the background (navigation history) until finding out an objective content in the case of sharing contents that users recommend. Because they can understand whether recommendation contents are selected carefully and investigated comprehensively if they understand the background. The background corresponds to user's purpose of Web information retrieval, a number of Web pages the user navigates, and information of the scope of user's Web navigation. In this paper, we propose a way to share selectiveness and exhaustiveness on Web exploration by using graphical visualization.

## 1. はじめに

WWW の急激な発達に伴い、Web での情報検索は身近なものになりつつある。職場や教育現場でのグループワークにおいて、また Web ショッピングや旅行先決定などで、他人から教

えられた Web 情報を参考にする場面は多い。つまり、個人が Web 探索により得たコンテンツをグループで共有するという Web 利用法が求められている。

本論文では、ユーザが Web 探索により得たコンテンツをグループ内に推薦するモデルを想定している。これを現状の方法で行おうとすると、ユーザが推薦を行うまでの Web 探索範囲、選択理由が明確でなく、推薦情報の価値がわかりにくいという問題点がある。そこで、我々はユーザの Web 閲覧履歴を用いる事でその問題を解決する。本論文では、探索履歴を利用する事により、ユーザの探索目的や探索範囲を把握することを目的とする。

協調作業を支援する従来のシステムとしては、いわゆるグループウェアがあるが、Web 探索を対象としたものではない。また、各ユーザのブックマークを共有するような Web サイトも存在するが、タイトルと URL のみを保持するブックマークの共有ではグループを支援する事は難しい。また、ユーザへのコンテンツ提案手法として、ソーシャルフィルタリング [1] や様々な推薦システム [2] もある。しかし推薦される Web ページの関連分野は自分で探さなければならないという問題点がある。また、Web ページ自体を評価するものであって Web 探索行為そのものを評価したものではない。

本論文では、探索結果のみではなく、探索結果に至るプロセスやその行為を含めて共有する手法を用い、ユーザの Web 探索行為自体を評価し、視覚化、共有する。この手法により、他のユーザの推薦コンテンツを見る事でより効果的な Web 探索を行えるものと考えられる。

## 2. 閲覧履歴の視覚化

### 2.1 グループ支援型 Web 探索

本論文では個人個人の Web 探索による情報がグループ内で共有されている状況をグループ支援型 Web 探索 [4] と呼んでいる。本論文においてグループは特定多数を対象としている。本システムはクライアント端末を持つユーザグループと情報を蓄積、共有するサーバマシンからなる。各ユーザは Web 閲覧を従来のように行い、他のユーザにも薦めたいと思った情報 (Web ページ) をサーバに送信する。また、その時同時に、履歴情報やクリックしたアンカー情報など、解析に必要な情報もサーバに送る。解析に必要な情報は後述する。サーバは送信された Web ページを保存し、グループ内で利用できる環境にするものである。

### 2.2 探索履歴の評価

まず具体例として、北海道のスキー場のページが推薦されている場合を考える。スキー場を探しているユーザが、その推薦ページを見た時に、その推薦ページだけで行きたいスキー場を決定するという事は想定し難い。推薦されたスキー場が、他所に比べてどのように良いのかが明記されていない限り判別しにくいからである。結果、ユーザはその推薦ページ以外のページも見たいと思うであろうが、どこをさらに調べれば効率的なのかは不明であり、ユーザは結局、ページを推薦したユーザと同じような Web 探索を行う事になる。従ってそれら Web 探索自体の評価情報を付加する事で推薦情報の価値を明らかにし、Web 探索を効率化するのが本論文の目的である。そこで、推薦 Web ページの価値として、ユーザの Web 探索における網羅度、精選度を以下に定義する。

- (1) 網羅度: 検索目的から考えられる Web ページ集合に対して、どの程度の範囲のページを閲覧したかを表す指標。
- (2) 精選度: 推薦ページはどれくらい他の Web ページと吟味

▼ 学生会員 京都大学大学院情報学研究科修士課程

[izu@dl.kuis.kyoto-u.ac.jp](mailto:izu@dl.kuis.kyoto-u.ac.jp)

♦ 正会員 独立行政法人 情報通信研究機構 (NICT)

メディアインタラクシオン G 専攻研究員

[snakajima@dbl-lab.is.ocha.ac.jp](mailto:snakajima@dbl-lab.is.ocha.ac.jp)

▲ 正会員 京都大学大学院情報学研究科

[tanaka@dl.kuis.kyoto-u.ac.jp](mailto:tanaka@dl.kuis.kyoto-u.ac.jp)

されて選ばれたのかを表す指標。

図1にユーザのWeb探索におけるWebページの包含関係についての概要を示した。Webページ全体と適合ページ、検索結果、実際に見たページ、推薦ページとの関係を模式的に示したものである。この図で網羅度は適合ページと実際に見たページとの関係を指し、精選度は推薦ページと実際に見たページとの関係を指す。

ここで、網羅度と精選度を算出するに当たって適合ページを知る必要があることがわかる。しかし、広大で、動的に変化しているWWWにおいて、ユーザの欲するWebページ数を知る事は不可能である。そこでユーザの閲覧目的を推定する事により、近似的に適合ページを把握する事を考えなければならない。

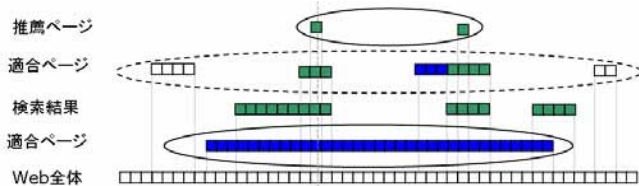


図1 包含関係

Fig.1 Relation of Inclusion

### 2.3 閲覧目的の推定

本節ではユーザのWeb探索履歴のうち、どのような情報を用いて解析し、ユーザの探索目的を抽出するかについて述べる。

同じWebページ(ターゲット)に別々のWebページからユーザが来た場合を考える。例えば、一方のユーザは、三重県にあるホテルを探しており、三重県のホテルのポータルサイトからターゲットに来たとすると、三重にあるホテルを見たいという目的でこのページに来た事がわかる。他方、良い温泉を紹介するページからターゲットにやってきたユーザは温泉を目的として見に来た事がわかる。

このように、同じWebページを取ってみても、ユーザのページに期待する閲覧目的は異なる。そこで、ユーザのこれまでたどってきたWebページや選んだリンクアンカーから目的を推定することを考える。次の節で閲覧履歴全体のユーザの目的抽出法と閲覧Webページごとの目的抽出法についてそれぞれ述べる。

#### 2.3.1 閲覧履歴全体

対象とする履歴としては、ユーザがWeb閲覧を始めてから、あるページを推薦するまでとしている。Web閲覧で用いられる検索キーワードとユーザがクリックしたアンカーの文字列のキーワードからユーザの閲覧目的を抽出する。

ユーザがクリックし、ナビゲーションが行なわれたリンクアンカーの文章から形態素解析[5]により単語を抽出し、その単語と出現頻度を要素として取り出し、すべてのクリックアンカーについて収集する。検索キーワードに出現頻度としての重みをつけ単語集合に含める。これらの単語集合から単語と出現頻度を要素とした特徴ベクトルを作成する。この特徴ベクトルを"トラッキング特徴ベクトル"と呼び、ユーザが推薦を行うまでの一連の探索における目的を表すものとする。

#### 2.3.2 各閲覧Webページ

まず、トラッキング特徴ベクトルと同様に各閲覧Webページの文章から単語を抽出し、その単語と出現頻度を要素として"ページ特徴ベクトル"を作成する。ページ特徴ベクトルはユーザの閲覧順序に依らないページの静的な特徴を表すも

のである。

次にユーザがクリックし、ナビゲーションが行なわれたリンクアンカーの文字列から単語を抽出し、"アンカー特徴ベクトル"を作成する。アンカー特徴ベクトル  $a$  とリンク先のWebページのページ特徴ベクトル  $p$  を足し合わせることでより一つの特徴ベクトルを作り、閲覧するWebページのユーザの目的を含んだ特徴ベクトルとする。この特徴ベクトルを"拡張ページ特徴ベクトル"とし、閲覧ページごとに決定する。

具体的には、WebページAのリンク  $a$  からページBに飛んだ場合、ページBの拡張ページ特徴ベクトル  $Be$  は、アンカー特徴ベクトル  $Aa$  と、ページ特徴ベクトル  $Bp$  から  $Be=Aa+Bp$  と表せる。また、"戻る"や、ブックマークを利用してページを開いた場合は拡張ページ特徴ベクトルとしてそのページ特徴語だけを保存する。(図2参照)

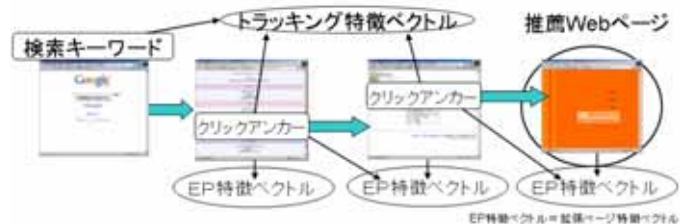


図2 各種特徴ベクトル  
Fig.2 Feature Vector

### 3. トラッキンググラフ

#### 3.1 トラッキンググラフの導入

各ページ特徴ベクトルを一つのノードに持たせ、網羅度と精選度を表すグラフを作成する事で可視化し、ユーザに提示する事を本論文では考える。そのグラフをトラッキンググラフと呼び、今節ではその生成方法について述べる。

##### 3.1.1 網羅度

2.3.節でユーザが推薦Webページを探し出すまでの目的の推定方法を述べた。また、2.2.節の定義から、ユーザが目的に沿ったWebページをより多く見ている場合、その探索の網羅度は高いとわかる。本論文では、Webページがユーザの目的に沿っているかを近似的に表すために、トラッキング特徴ベクトルと、各ページ特徴ベクトルの類似度を用いる。つまり、この類似度の高いWebページを多く見ているWeb探索は網羅度が高いといえる。よって網羅度を表す式は、トラッキング特徴ベクトル  $T$  と各閲覧Webページのページ特徴ベクトル  $P$  を用いて次のようになる。類似度  $Sim$  の計算方法にはコサイン類似度を用いている。

$$Exhaustivity = \sum_n Sim(T, P)$$

ここで、この方法で網羅度を測るに当たって、ユーザのWeb探索への熱心さ、うまさに関係すると考えられる。Web閲覧途中で、当初の検索目的に無いリンク先を頻繁に見に行くユーザのトラッキング特徴ベクトルは、目的と関係のない単語の要素の値も高いものとなる。よって類似度を考えた場合、達成すべき目的と関係のないWebページに対して目的を網羅したとみなしてしまうからである。

この問題はトラッキング特徴ベクトルの各要素の値の偏りによって解決する。トラッキング特徴ベクトルのある要素の値だけが大きく、他の要素の値は低いWeb探索は一貫性があり、より正しい網羅性を表現していると考えられるからである。本論文ではトラッキング特徴ベクトルの値を二乗する事

により要素の出現頻度による差を広げ、この一貫性を考慮するものとした。つまりトラッキング特徴ベクトルを  $v = (v_1, v_2, \dots, v_N)$  とすると類似度計算時は  $v = (v_1^2, v_2^2, \dots, v_N^2)$  としている。

### 3.1.2 精選度

2.2.節の定義から、ユーザが推薦する Web ページを探し出すまでに、類似したページを多く見てから選んだ場合に精選度は高い。そこでこれらの客観的な類似度をみるために、推薦 Web ページとその他の閲覧 Web ページの拡張ページ特徴ベクトルの類似度を求める。この類似度により精選度を明らかにする。よって精選度を表す式は推薦ページの拡張ページ特徴ベクトル  $rP$  と各閲覧 Web ページの拡張ページ特徴ベクトル  $P$  を用いて次のようになる。

$$Selectivity = \sum_n Sim(rP, P)$$

### 3.2 網羅度と精選度による評価

前節から求めた網羅度、精選度を用いてユーザの Web 探索を評価する指針について述べる。網羅度、精選度の定義から、これらの値は上限が決まらないため正規化できない。そこでユーザが推薦ページを決めるまでの Web 探索に対して、各 Web ページをグラフ上に配置し、探索を評価する事を考える。図3に探索履歴の精選度と網羅度についてあらわす S-E グラフを示した。

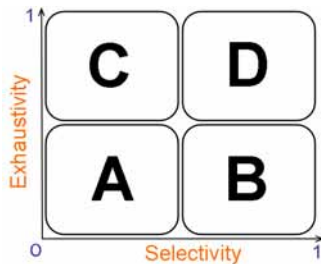


図3 S-E グラフ  
Fig.3 S-E Graph

このグラフの横軸は推薦ページと各 Web ページの類似度、縦軸は Web 探索におけるトラッキング特徴ベクトルと各ページの特徴ベクトルの類似度としている。ユーザの Web 探索の結果がこのグラフの何処に偏っているかにより、ある程度ユーザの Web 探索を推定することができる。グラフの A, B, C, D の各部分に偏った時の意味合いは次の通りである。

- A : ばらばらにいろいろページを見ている。
- B : 推薦ページの関連ページは多く見ているが実際の目的にはあまり沿っていないページ群である。
- C : 目的に沿って見ているが推薦ページとはあまり関連の無いページを多く見ている。
- D : 目的に沿ったページであり推薦ページとも関係あるページが多い。

B, D に多くページが分布している Web 探索は精選度が高く、C, D に多くページが分布している Web 探索は網羅度が高いと言える。S-E グラフを作成する事によりユーザの Web 閲覧探索を網羅度、精選度の面から評価する事ができる。

しかし、このグラフでは各々のページの関係や、実際にどういったページを閲覧したのかは判りづらい。また、閲覧履歴の評価をユーザ間で共有するために、ページ間の関係を表す必要がある。そこでページ間の関係を示したトラッキンググラフの作成法について次節で述べる。

### 3.3 トラッキンググラフの生成法

トラッキンググラフを作るに当たっての指針として以下の2点が挙げられる。

- (a) 拡張ページ特徴ベクトルの類似度が高い Web ページ同士ほど近づける。
- (b) 閾値以上の類似ページを集め、それらを代表するノードを作成することによりクラスタリングを行なう。

具体的なトラッキンググラフの作成法を以下に示した。グラフには閲覧履歴の時系列にそって過去の Web ページから拡張ページ特徴ベクトルを一つずつ入力し、反復的に作るものとする。

- (1) トラッキンググラフはルート、中間ノード、リーフノードからなる。すべてのノードは同次元の特徴ベクトルを持つ。リーフノードは特徴ベクトルとして拡張ページ特徴ベクトルを持ち、Web ページを表す。
- (2) まず初期ノードであるルートを作成する。はじめはルートの特徴ベクトルは 0 である。入力された Web ページのページ特徴ベクトルをリーフノードとしてルートに繋いでいく。
- (3) 新たなリーフノードが来る度に、ルートから近い中間ノードから順に類似度を計算し、閾値を超えた中間ノードにリーフノードを接続する。接続する際は、同じ中間ノードに繋がるリーフノードとの類似度を計算し、類似度が閾値以上のノードがあればそれらすべてと新たな中間ノードを作成し、閾値以下であれば中間ノードにそのままつなぐ。
- (4) ルートと中間ノードの特徴ベクトルはその子ノードの特徴ベクトルをすべて足し合わせる。

以上の作成法により、リーフノードは類似度の高い一つ以上の中間ノードを親を持つ。ルートはすべてのリーフノードの特徴ベクトルを足し合わせたノードになる。

- この手法により、グラフは次の特徴を持つ。
- 目的に合致した類似ページをたくさん見ているユーザのトラッキンググラフは特徴ごとの中間ノードに固まり、枝が多く大きなグラフとなる。
- ネットサーフィン時のようにばらばらの Web ページを見ているユーザのトラッキンググラフはルート付近に固まった枝の少ないものとなる。
- ページ内の単語だけを使った静的なグラフではないため、動的に変化させる事ができる。

### 4. プロトタイプ

前節までのシステムのプロトタイプを実装した。グループ内で共有する前段階として、各特徴ベクトルから網羅度、精選度を計算し、グラフに表す部分である。

#### 4.1 プロトタイプシステムの概要

プロトタイプシステムの処理の流れは以下の通りである。

- (1) ブラウザにてユーザは通常通り Web 探索を行う。システムは閲覧 Web ページの単語を解析し、ユーザの閲覧行動に沿って、各ページのページ特徴ベクトル、拡張ページ特徴ベクトルを作成する。ユーザが Web ページの推薦を行なう。システムは閲覧全体からトラッキング特徴ベクトルを作成する。この時、ユーザに閲覧した各ページの URL、タイトル、各特徴キーワードとその頻度を提示する。
- (2) S-E グラフを表示する。

3.1 節で示した手法により、グラフを表示する。この時、各 Web ページを表すグラフ上のノードをクリックすることで、そのノードに対応するページの単語と TF 値を確認できるも

のとしている。(図4参照)

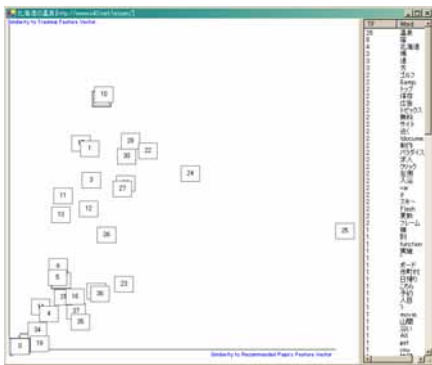


図4 S-Eグラフ(プロトタイプ)  
Fig.4 S-E Graph (Prototype)

(3)トラッキンググラフを表示する。

3.3節で示した手法でグラフ化し、提示する。閾値は最も閲覧の特徴を表す事が出来るグラフが作成されるよう設定する。経験的にデフォルトを0.6としたが、ユーザ入力により変更できるものとした。丸いノードでルートもしくは中間ノードを表し、四角いノードでリーフノードを表している。ノードに表示する数字でユーザの閲覧した順序、ノード付近に書いてある文字でそのノードの特徴語をあらわしている。(図5参照)



図5 トラッキンググラフ(プロトタイプ)  
Fig.5 Tracking Graph (Prototype)

## 4.2 考察

図4のS-Eグラフは北海道のスキー場に近い温泉旅館を探し、ニセコの旅館を推薦した履歴例を表したものである。実際、ユーザは北海道の温泉とスキー場についてよく調べている。しかし、推薦したニセコの旅館のページを比較するようなサイトはあまり見て回っていない。この履歴のトラッキング特徴ベクトル上位単語は{北海道, 温泉, 宿, スキー}となっている。3.2節の手法で、S-Eグラフを考察すると、この履歴は目的に沿った検索を行えているが、推薦ページに似た周辺ページがあまり見られていないのではないかと推定する事ができ、この推定はおおよそ当たっている。S-Eグラフからある程度ユーザの動向が測る事ができていると考える事ができる。

図5は京都のホテルを検索した際の履歴のトラッキンググラフである。新撰組に関する京都のホテルを重点的に調べている事がわかる。ページの内容によってクラスタリングする事ができている。また、ルート0に接続されているリーフノードは推薦ページや目的とあまり関連のないページを現している、無駄なページ群もたくさん見ていることも表す事ができていると考えられる。

## 5. おわりに

本論文では個人のWeb探索情報をグループで共有する際の視覚化の手法について述べた。新たに取り入れたトラッキンググラフという情報抽出グラフによりユーザのWeb閲覧の網羅度や精選度を示した。今後の課題として以下を考えている。

- (1) 目的を導くためにトラッキング特徴ベクトルとしてリンクアンカーの文字を抽出しているのみであるが、Webページの前後関係の文脈によるWebページの印象の違い、単語間の関連性を考慮した抽出方法を検討している。
- (2) 今は閾値を任意に決めているが、ノードの数などからWeb探索ごとに最もその探索をあらわすグラフを作成する閾値を求める方法を検討している。
- (3) ユーザ間のトラッキンググラフの統合環境を実装する事で、各ユーザの閲覧の関係の提示や、ユーザごとの得意不得意分野を明らかにしたい。また、統合時のグラフの自動併合、分割についての実験を検討している。

## 【謝辞】

本論文の一部は、平成15年度文部科学省科学研究費特定領域研究(2)「Webの意味構造に基づく新しいWeb検索サービス方式に関する研究」(課題番号:15017249)、および京都大学21世紀COEプログラム「知識社会基盤構築のための情報学拠点形成」による。ここに記して謝意を表します。

## 【文献】

- [1] U. Shadanand, et al. Social Filtering: Algorithms for Automating 'Word of Mouth', CHI'95, pp.210-217, ACM Press.
- [2] P. Resnick ed. Recommender Systems, CACM Vol.40, No.3, pp.56-89, March 1997. Let's Browse: A Collaborative Web Browsing Agent International Conference on IUI (1999).
- [3] Shinsuke Nakajima, Satoshi Oyama, Kazutoshi Sumiya and Katsumi Tanaka: "Context-Dependent Web Bookmarks and Their Usage as Queries". Proc. of the 3rd International Conference on Web Information Systems Engineering. WISE2002, pp.333-340 (2002).
- [4] 伊豆 陸, 中島 伸介, 小山 聡, 角谷 和俊, 田中 克己: グループ型Web閲覧による探索アクティビティ情報の共有と利用. 第14回データ工学ワークショップ (DEWS2003), 2003年3月.
- [5] 奈良先端科学技術大学松本論文室 茶筌ホームページ: <http://chasan.aist-nara.ac.jp/index.html>

## 伊豆 陸 Atsushi IZU

京都大学大学院情報学研究科修士課程在学中。2003年京都大学工学部情報学科卒業。日本データベース学会学生会員

## 中島 伸介 Shinsuke NAKAJIMA

独立行政法人情報通信研究機構勤務。2004年京都大学大学院情報学研究科博士後期課程修了、博士(情報学)。日本データベース学会、情報処理学会、人工知能学会、環境システム計測制御学会各会員。

## 田中 克己 Katsumi TANAKA

京都大学大学院情報学研究科社会情報学教授。1976年京都大学大学院博士前期課程修了、工学博士。主にデータベース、マルチメディアコンテンツの処理の研究に従事。IEEE Computer Society, ACM, 人工知能学会, 日本ソフトウェア科学会, 情報処理学会, 日本データベース学会等各会員。