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

伊豆　陸†　　中島　伸介†　　田中　克己†

† 京都大学大学院情報学研究科　〒 606–8501 京都府京都市左京区吉田本町
E-mail: †{izu,nakajima,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし　現在，Web を利用した情報検索は，単に個人のための情報収集にとどまるものではない．他のユーザに向けた情報提示のために，Web 探索を行なう要求が高まってきている．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環境においては，他のユーザから推薦された Web コンテンツの選ばれた背景を知る事は有用である．背景とは，ユーザが Web コンテンツの検索目的や，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．背景を明示する事により，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．本研究では，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．

キーワード　Web とインターネット, 情報検索, データの可視化, ユーザインタフェース, プロファイル

Visualization and Sharing of Web History on Group-Based Web Browsing

Atsushi IZU†, Shinsuke NAKAJIMA†, and Katsumi TANAKA†

† Graduate School of Informatics, Kyoto University　Yosidahonmati, Sakyou-ku, Kyoto,606–8501 Japan
E-mail: †{izu,nakajima,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract　Recently, information retrieval on Web browsing is not always for individual use. It is considered that demands for the information presentation to other users has been increasing. It is useful to get to know the background as which the contents recommended by other users were chosen in the environment where a group shares the contents obtained by individual. Because if users know what others think to search , how many similar pages to recommendation they browse or what range they peruse the Web contents ,users can know whether recommendation contents is selected carefully and perused comprehensively. In this paper, we propose a way to share selectiveness and comprehensiveness on Web browsing by graphical visualization.

Key words　Web and Internet, Information Search, Visualization, Profile

1. はじめに

WWW の急激な発達に伴い，Web から情報を検索するという行為は身近なものになりつつある．Web の利用形態においても多様化しており，チャット機能やアノテーション機能を介したコミュニケーションが行なわれるようになってきている．またコミュニケーションツールも広く利用されつつある状況にある．現在，ユーザが Web 上で行う情報探索は単に個人のための情報収集という利用方法だけでなく，グループに向けた情報提示という場で利用する要求が高まってきている．office や教育現現場でのグループワークにおいてや，個人のショッピングや旅行先決定などで，他人から教えられた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と考えられる．本研究ではこの利用法をグループ支援型 Web ブラウジングと呼んでいる．また, 従来の個別の Web ブラウジングでは，ユーザが明確な検索要求を持たない場合や検索キーワードを思いつかない場合，うまく検索を行えないという問題がある．グループ支援型 Web ブラウジングによって共有されたコンテンツを見る事によりユーザはより大きな視野, 情報源を元に情報収集を行なえるものと考えられる．

本研究では, 個人の Web 探索により得たコンテンツをグループで共有するために，グループに Web コンテンツを推薦するという形を想定している．この時の問題点として次の点が挙げられる．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事である．

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事を考えた．グループで共有する情報として，探索結果の Web ページだけではなく, ユーザの探索履歴の情報までを含めて保存し，共有する事から，その時のメンバーの探索目的や探索範囲を把握する事ができるのではないかと考えたからである．またグラフの自動併合, 分割についても考察する．

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のようなものがあるが，Web 探索の背景を知る事を対象としたものはない．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 Web ページ自体を対象としており，その関連分野等は自分で探さなければならないといった問題点がある．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．

本研究では，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．この手法により複数のユーザによる効果的な Web 探索の実現を可能とするシステムを提案する．

2. コンテクストブックマーク

コンテクストブックマーク [3] とは，Web ブックマークを拡張したものである．従来のブックマークは，記憶しようとするページの URL とタイトルのみを保存するものであるが，コンテクストブックマークは，ブックマークをつけるまでに閲覧した Web ページの情報を同時に保存させるものである．保存するメタデータとして以下のものがあげられる．

- ブックマークした Web ページの URL，タイトル，代表キーワード
- 閲覧した Web ページ群の URL，タイトル，代表キーワード
- 精選度（どの程度類似ページを閲覧し，精査したものかを表す指標）

精選度を計算するにあたって以下の２つの仮定を挙げている．

（1） ブックマークするまでに多くのページを閲覧していれば，そのブックマークの価値は高い．

（2） ブックマークするまでに閲覧しているページ群と，ブックマークしたページとの類似度が高ければ，そのブックマークの価値は高い．

精選度によりブックマークの持つ意味や重要性を表現できると同時に, 他のユーザがそのブックマークをどのような範囲を閲覧して選んだものか知る事ができる．従来では困難であった探索結果の妥当性を，他のユーザでも評価する事が可能になるものである．本研究では推薦 Web ページを保存するにあたって本コンテクストブックマークの保存手法を基礎とし，この考えを拡張するものである．保存するメタデータとして上記以外にリンクアンカー文字列や検索キーワードを考慮に入れる．

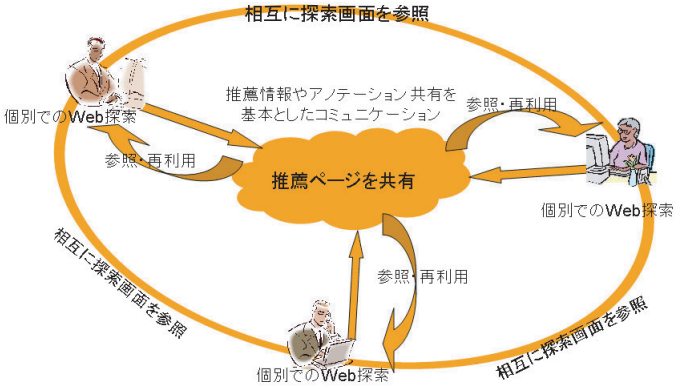

図 1　グループ支援型 Web 探索システム

3. 閲覧履歴の視覚化

3.1 グループ支援型 Web 探索

本研究では個人個人の 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．概念図を図 1 に示す．本研究においてグループは不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を思った情報（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．

3.2 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た場合を考える．スキー場を探しているユーザがその推薦ページを見たとすると, ユーザが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，ユーザがその推薦ページ以外の類似ページを見たいと思うと考えられる．しかし，類似ページにも場所が近い，値段が近いなどいろいろな角度から見た類似性が考えられ，また類似ページがどれくらい調べられたのかもわからないため，どこをさらに調べると効率的か不明である．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．それら Web 探索自体の評価を行ない，その情報を付加する事で推薦情報の価値を明らかにするのが本研究の目的である．

情報の付加にはユーザの閲覧履歴を利用する．ユーザが推薦した Web ページにたどり着くまでに見てきた Web ページと辿ったリンクからどのような傾向のページを閲覧したのか, どのような類似ページをどれだけ閲覧したのかを抽出し，推薦 Web ページに付加するものである．付加情報からわかる推薦 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．

- 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.
- 精選度
  推薦ページはどれくらい他の Web ページと吟味されたのかを表す指標.

図 2 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 ページ全体のうち, どの部分が適合ページで，検索結果や実際に見たページ，推薦ページはどの部分か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の関係を指し，精選度は推薦ページと実際に見たページの関係を指す.

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る．次節ではユーザの目的の推定方法について述べる.

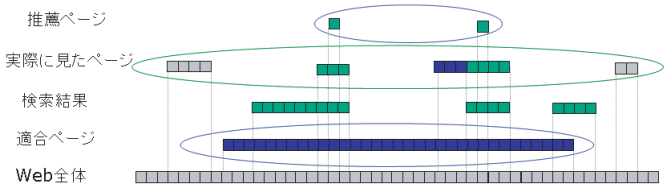

図 2　包 含 関 係

3.3　閲覧履歴のトラッキング

トラッキングとは, ユーザの Web 閲覧履歴に蓄えられているユーザのアクティビティ情報から, ユーザの行動を追跡する事である. ユーザのアクティビティ情報というのは, ユーザの閲覧した Web の URL，検索キーワード，クリックしたアンカー, 推薦した Web ページなど, Web 閲覧においてユーザの行動から生じる情報である. 本節ではユーザの Web 探索履歴のうち, どのようなアクティビティ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.

図 3 にユーザの探索目的の相違について示した．同じ Web ページに別々の Web ページからユーザが来た事を示している．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ページに来ている．三重にあるホテルを見たいという目的でこのページを見に来た事がわかる．他方のユーザは, 良い温泉を紹介するページからターゲットページにやってきており，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，ユーザのそのページに期待する閲覧目的は異なる．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 2 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.

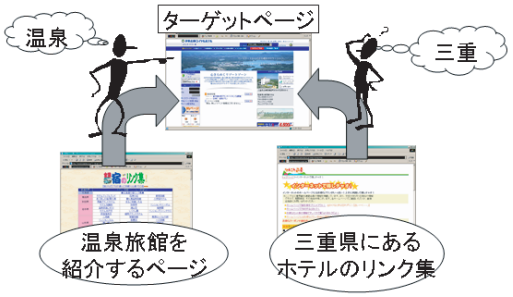

図 3　目的の相違

3.3.1　閲覧履歴全体

抽出する履歴としては，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の履歴を対象としている．まず前提として，Web 閲覧で用いられる検索キーワードはユーザの探索目的を表していると考えられる．またユーザは Web ページ内のリンクから次のページに移る場合, 主に今見ている Web ページの情報を頼りに最も自分が見たいと思うアンカーをクリックするものと考えられる．つまりユーザは選べる中で自分の探索目的に最も適合すると思われる最善のリンクを選ぶと考えられる．そこで Web 探索全体におけるユーザの目的はユーザがクリックしたアンカーの文字列と検索エンジンに入力したキーワードからある程度抽出する事ができると考えられる.

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 [9] 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，ユーザが推薦を行なうまでの一連の探索における目的を表すものと考える．また，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.

3.3.2　閲覧 Web ページ

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，その単語と出現頻度を要素とした特徴ベクトル，”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と，前節述べたページ特徴ベクトルから, 各 Web ページに対するユーザの目的を含んだ，動的な特徴を現す特徴ベクトルを定義する．アンカーの期待特徴ベクトル $\vec{a}$ とリンク先の Web ページのページ特徴ベクトル $\vec{p}$ を足し合わせることに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 $\vec{Be}$ はアンカー特

徴ベクトル $\vec{Aa}$ とページ特徴ベクトル $\vec{Bp}$ から $\vec{Be} = \vec{Aa} + \vec{Bp}$ と表せる．また，ブラウザの戻るを利用したり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 4，5 参照）

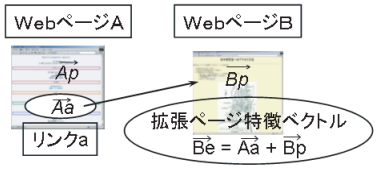

図 4　拡張ページ特徴ベクトル

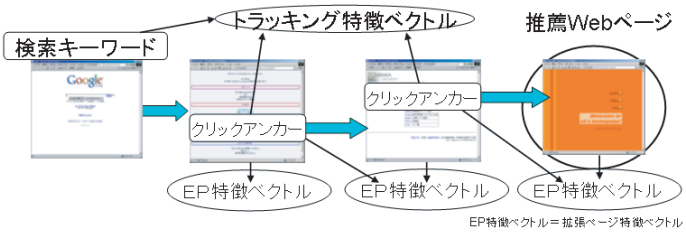

図 5　特徴ベクトル

すべての特徴ベクトルの次元をそろえる為に閲覧 Web ページに出現するすべての単語数 V を元に特徴ベクトルの次元を V 次元とし，ページに存在しない単語の要素は 0 としてすべての特徴ベクトルを V 次元に作り変えておく．

3.4　トラッキンググラフ

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研究では考える．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．

3.4.1　網　羅　度

3.3 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．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に，網羅度は高いと考えられる．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事とする．類似度の高い Web ページを多く見ている Web 探索が網羅度が高いといえる．よって網羅度 (Exhausitivity) 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 の計算方法については後述する．

$$Exhausitivity = \sum^{n} Sim(T, P) \tag{1}$$

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，うまさが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．よって類似度で考えた時に，達成すべき目的と関係ない Web ページまで目的を網羅したものとされ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高く，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．本研究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 $v = (v1, v2, ...vN)$ とすると類似度計算時は $v = (v1^2, v2^2, ...vN^2)$ としている．

3.4.2　精　選　度

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と考えられる．そこでこれらの客観的な類似度をみるために，推薦 Web ページのページ特徴ベクトルと各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 (Selectivity) を表す式は推薦ページのページ特徴ベクトル rP 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．

$$Selectivity = \sum^{n} Sim(rP, P) \tag{2}$$

各類似度計算にはコサイン類似度を用いる．特徴ベクトル F,Q の類似度 sim(F,Q) は以下の通りである．

$$sim(F, Q) = \cos(F, Q) \tag{3}$$

$$= \frac{F \cdot Q}{\| F \| \| Q \|} \tag{4}$$

3.4.3　網羅度と精選度による評価

3.4.1，3.4.2 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，履歴 Web ページの類似度をグラフ上に配置し，表す事を考える．図 6 に探索履歴の精選度と網羅度についてあらわす E-S グラフを示した．

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 探索を推定する事ができる．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なページを見ている．
- 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，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．

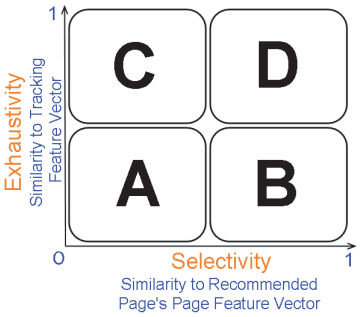

図 6　S-E グラフ

3.4.4　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．

- 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．
- 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ることで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 V 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 $\vec{0}$ 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．

図 7 に示すように，中央のトラッキンググラフに新たなリーフノード L2 が入力されたとする．今，R はルート，L はリーフノード，N は中間ノードである．L2 とリーフノードとの類似度がどれも閾値を超えなかった場合左のグラフのようになる．L2 と L1 の類似度が閾値を超えた場合右のグラフのようになる．この時，新たにできたノード N の特徴ベクトル $\vec{n}$ は $L1 = \vec{l1}, L2 = \vec{l2}$ を用いて $\vec{n} = \vec{l1} + \vec{l2}$ となる．

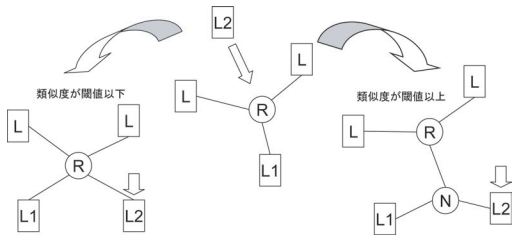

図 7　トラッキンググラフ

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に持つ．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，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．また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．

また，このグラフの特徴として，ページ内の単語だけを使った静的なグラフではない事が挙げられる，ユーザがどのリンクから飛んできたかを考慮しており，ユーザの閲覧にしたがって動的に変化させる事ができる．

3.4.5　トラッキンググラフの作成アルゴリズム

実際のトラッキンググラフ作成アルゴリズムは以下の通りである．

入力:ユーザが閲覧した Web ページの拡張ページ特徴ベクトル
出力:トラッキンググラフ

（1） 初期ノードを設定する
まず何の枝も張っていない初期ノードである *Root* を作成する．

（2） 最初に閲覧した Web ページ
閲覧 *Web* ページの拡張ページ特徴ベクトル *L1* を作成する．
Root から枝を張った先にリーフノード *L1* として繋ぐ．
Root の特徴ベクトルとして *L1* を設定する．

（3） 2 番目に閲覧した Web ページ
閲覧 *Web* ページの拡張ページ特徴ベクトル *L2* を作成する．
L1 の特徴ベクトルとの類似度を計算する．
類似度が閾値未満である時
Root から枝を張った先にリーフノード *L2* としてつなぐ．
Root の特徴ベクトルに *L2* を足し合わせる．
類似度が閾値以上である時
L1 と *L2* の特徴ベクトルを足し合わせ, 新たにできた特徴ベクトルを中間ノード *N1* とし,*Root* からの枝につなぐ．
L1 と *Root* の枝を切り,*L1* を *N1* に繋ぐ．
N1 にリーフノード *L2* として繋ぐ．

（4） n 番目に閲覧した Web ページ
閲覧 *Web* ページの拡張ページ特徴ベクトル *Ln* を作成する．

Root に直に繋がっているすべての中間ノードの特徴ベクトルと *Ln* との類似度をそれぞれ計算する.
どの類似度も閾値未満である時
Root に対してリーフノード *Ln* との接点の計算（＊ノードへの接続）を実行する.
類似度が閾値以上のものがあった時
閾値以上であった中間ノード群を *Nk* とする.
Nk に子として繋がっているすべての中間ノードの特徴ベクトルとの類似度を計算する.
どの類似度も閾値未満である時
Nk に対してリーフノード *Ln* との接点の計算（＊ノードへの接続）を実行する.
類似度が閾値以上のものがあった時
閾値以上であった中間ノード群を *Nl* とする.
Nl に対して *Nk* に行った事と同じ事を行い, 以下ノードがなくなるまでそれを繰り返す.
（*end*）
（＊ノードへの接続）
ノード *N* にリーフノード *L* を接続する場合の計算方法は以下の通りである.
L の特徴ベクトルと *N* に繋がるすべてのリーフノードとの類似度をそれぞれ計算する.
類似度が閾値未満である時
N から枝を張った先にリーフノード *L* としてつなぐ.
ノード *N* に対して特徴ベクトルの再計算（†ノードの更新）を実行する.
類似度が閾値以上のものがあった時
閾値以上であったリーフノード群を *Lk* とする.
ノード *N* に対して特徴ベクトルの再計算（†ノードの更新）を実行する.
Lk と *L* の特徴ベクトルを足し合わせ, 新たにできた特徴ベクトルを中間ノード *Nk* とし,*N* からの枝につなぐ.
Lk と *N* の枝を切り,*Lk* を *Nk* に繋ぐ.
Nk にリーフノード *L* として繋ぐ.
（*end*）
（†ノードの更新）
ノード N にリーフノード L を接続する時の中間ノードの特徴ベクトルの再計算方法は以下の通りである.
N の特徴ベクトルに L の特徴ベクトルを足し合わせる.
N が繋がる先の中間ノード N’ の特徴ベクトルにも L の特徴ベクトルを足し合わせる.
以下 N’ から Root にいたるまで繋がる中間ノードすべての特徴ベクトルに L の特徴ベクトルを足し合わせる.
（end）

4. 閲覧履歴の共有

4.1 トラッキンググラフの表示

前節でトラッキンググラフの作成方法について述べたが，トラッキンググラフをグループ内で共有した時に, ノードに何を表示するのかを定義していない．ここではノードを代表する特徴語をそれぞれの中間ノード，リーフノードに与え，表示する事について述べる.
（1） Root について
保持する特徴ベクトルの最もその要素が大きい単語を表示する.
（2） 中間ノード/リーフノードについて
すべての中間ノード/リーフノード間での tf/idf 値の最も高い数単語を表示する.

- tf/idf 値 (term frequency/inverse document frequency) とは任意の文章中におけるタームの出現頻度を文章に現われるタームの数で正規化した tf 値とタームの現れる文書数で正規化したの逆数 idf を掛け合わせた重みである.

$$tf_{ij} = freq(i,j) \tag{5}$$

$$df_j = dfreq(j) \tag{6}$$

freq(i,j) は文章 Di におけるターム Tj の出現頻度を表す. Dfreq(i) は文書群におけるターム Tj の出現文書数を表す. よって tf/idf 値は

$$tfidf_{ij} = freq(i,j) * log(N/dfreq(j)); \tag{7}$$

と表せる．N は総文書数を表す．本システムでは特徴ベクトル一つを一つの文書とみなして計算している.

この手法によりリーフノードには各 Web ページのユーザの Web 探索全体に対する特徴語が提示される．またルート，中間ノードにはユーザの Web 閲覧の目的を反映させる事ができていると考えられる．具体的なトラッキンググラフの概要図を図 8 で示した．北海道にある温泉のついたスキー場を調べた例である．例であるので実際の結果とは違うが，この例で，ユーザはスキー場を重点的に調べており主に各地のスキー場を調べている事がわかる，しかしその他いろいろなページもでており北海道にある温泉のついたスキー場に関してはこの他にはあまりなさそうであるとわかる.

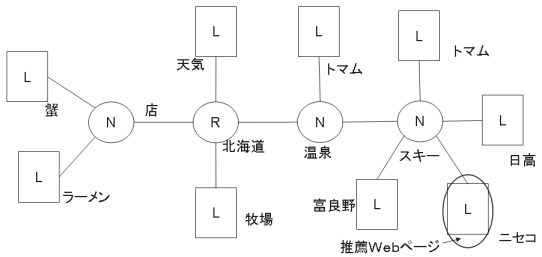

図 8 トラッキンググラフ

しかし閲覧 Web ページが大量になってくると，トラッキンググラフが膨張し画面上に表示するには現実的ではなくなってくると考えられる．また，複数ユーザ間でトラッキンググラフを共有する際に併合，分解の問題が出てくると思われる．そこでトラッキンググラフを自動的に併合, 分割, 再構成し，ユーザに提示するモデルについて次節で述べる.

4.2 トラッキンググラフの動的な併合・分割・再構成

多くの Web ページを閲覧するにつれて拡大するトラッキン

ググラフを画面内におさめるために, 自動的に分割，または縮小させる必要がある．また，推薦コンテンツが重複した時，複数ユーザのトラッキンググラフを併合させる必要がある．今節ではその方法について述べる．

- トラッキンググラフの自動縮小，分割

トラッキンググラフにおいて中間ノードはそこに繋がるリーフノードを代表する特徴ベクトルを持つ．そこでトラッキンググラフが大きくなるにつれ余分なリーフノードを省略する事が考えられる．中間ノードに繋がるリーフノードの数が少ないものほど, そのリーフノードを省くものとする．その中間ノードの特徴に関連する Web ページは, 検索目的から外れていると考えられるからである．また，ルートからリーフノードまでの中間ノードの数が多いものほど重要であると考えられるため，中間ノードの少ない部分の枝葉を省くものとする．

また，各ノード間の類似度の大きさにより閾値を設け，自動分割を行なう．これによりグラフを小さく出来ると共にユーザの網羅度による履歴の分類を行なう事が可能であると考えられる．

- トラッキンググラフの自動併合

複数ユーザが同じページを参照した場合にグラフの自動併合を行なう．事により他ユーザの閲覧した範囲も含めた閲覧範囲の関係を視覚化する事が可能である．

5. 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，各ユーザの閲覧履歴からトラッキング特徴ベクトル，ページ特徴ベクトルを計算し，網羅度，精選度を計算，グラフに表す部分を実装した．今節ではシステムの概要とそのシステムを使った時の評価について考察する．

5.1 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．ユーザが Web ページの推薦を行なう．(図 9 参照) 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．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．(図 10 参照)

図 9 Browser

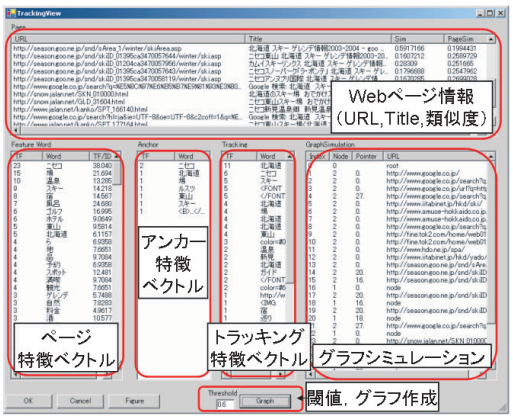

図 10 TrackingViewer

（2） E-S グラフを表示する．

3.4.1 節〜3.4.3 節で示した手法により，グラフを表示する．この時，各点をクリックすることで，その点に対応するページの単語と TF 値を確認できるものとした．(図 11 参照)

（3） トラッキンググラフを表示する．

3.4.4 節 3.4.5 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ようなグラフが作成されるよう設定するため，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．図 12 で丸いノードはルートもしくは中間ノードを表し，四角いノードはリーフノードを表している．ノードに表示している数字はユーザの閲覧した順序，ノード付近に書いてある文字はそのノードの tf/idf 値の上位 3 単語である．グラフの表示方法としては単純に，接続された枝の数で 360 °を割り枝の長さを一定として配置した後，ノードがかさならないようずらしたものである．現在は見づらいものとなっているが，既存のグラフ表示アルゴリズム等を適応し改善する予定である．

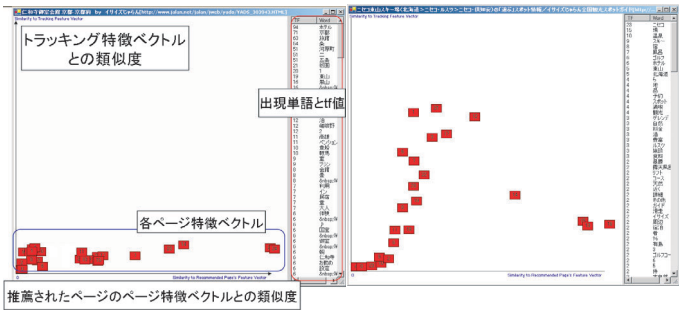

図 11 E-S グラフ

5.2 考　　察

図 11 の右の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 E-S グラフで表したものである．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}

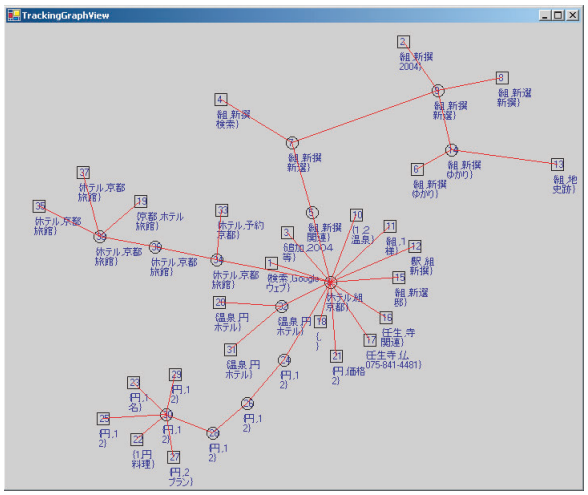

図 12　TrackingGraphView

となっている．3.4.3 節の手法で，E-S 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．E-S 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．

一方，図 11 の左のグラフは京都の新撰組のゆかりの地を巡る旅行を計画し，ホテルを探した例である．トラッキング特徴ベクトルの上位単語は {旅館，新撰組，四条大宮，京都} であり，推薦ページは四条にあるホテルの旅行プランである．実際の Web 探索としては前半に新撰組について調べ，その中から行きたい地名をピックアップし，後半に地名からホテルを調べている．前半で新撰組についていろいろ調べている部分と後半のホテルを探す部分に関連性が見出せていないためいろいろなページを見に行った事と同じ事になり網羅度が低くなっていると思われる．この問題を解決するためには，トラッキング特徴ベクトルの単語に関連性などを持たす，といった事を考えなければならないと思われる．

図 12 は，図 11 の左の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，京都のホテルを重点的に調べている事がわかる．左下の {円，１，２} 等と表示されているページはホテルの値段をあらわすページであると思われる．ページの内容によってある程度クラスタリングする事ができていると考えられる．また，ルート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，無駄なページ群が多かったことも表すことができていると考えられる．

6. おわりに

本研究では個人の 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．他のユーザはこのグラフを見る事により推薦情報の価値を効果的に推定でき，新たな Web 探索の一助とするものが出来ると考えられる．

今後の課題としては，以下の４点が挙げられる．

- ほぼ同じような目的を表すトラッキングベクトルを持つ探索履歴を二つ重ね合わせる事で，適合ページの数を統計的に推定する事ができると考えられる．その手法について実装，実験を行ないたい．
- 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や，単語間の関連性を考慮した抽出方法を検討している．
- 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．
- 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を提示したり，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，分割について実験を行う事を検討している．

謝　辞

本研究の一部は，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] Henry Lieberman, Neil Van Dyke, and Adriana Vivacqua:

[6] 中島伸介，上田正明，田中克己: 検索内容・アクティビティの共有と視覚化に基づくグループ型情報探索システム，Proc. of DBWeb2002, 情報処理学会シンポジウムシリーズ Vol.2002, No.19, pp.129-136 (2002)

[7] Henry Lieberman, Neil Van Dyke, and Adriana Vivacqua: Let's Browse: A Collaborative Web Browsing Agent
International Conference on IUI 1999.
http://lieber.www.media.mit.edu/people/lieber/Lieberary/Lets-Browse/Lets-Browse.html

[8] Federico Bergenti,Agostino Poggi,Matteo Somacher: A collaborative platform for fixed and mobile networks
ACM Press ,NY,November 2002,pp.39-44 .

[9] 奈良先端科学技術大学松本研究室 茶筅ホームページ：
http://chasen.aist-nara.ac.jp/index.html