複数人でのWeb協調探索のための探索履歴可視化手法の提案

枝　隼也† 島田　諭† 関　洋平† 神門　典子†† 佐藤　哲司†

† 筑波大学大学院図書館情報メディア研究科　〒305–8550 茨城県つくば市春日 1-2
†† 国立情報学研究所　〒101-8430 東京都千代田区一ツ橋 2-1-2
E-mail: †{eda,sat,yohei,satoh}@slis.tsukuba.ac.jp, ††kando@nii.ac.jp

あらまし　大きな課題について，ユーザがさまざまなページを訪問し情報間の関係を調べながら進める探索的検索は，近年その重要性の認識が高まっている．本論文では，単独のユーザが閲覧したページの探索的検索履歴，更には複数人による協調探索の過程を，キーワード平面上に可視化する手法を提案する．提案法では，探索範囲となる話題範囲を，キーワードの意味的関係を距離とする探索空間マップを作成し，ユーザが閲覧したページをマッピングしていく．ある課題を探索した複数人の履歴のデータを対象に，提案法を適用した結果，各々のユーザには異なった探索的空間があることが明らかとなり，探索空間マップ内に占める探索済み空間の比率を示すマップ充足率を用いて，それぞれのユーザの主観的な探索達成度合いを反映できていることも確認された．さらに，複数ユーザの探索空間を組み合わせるせることによって，より有効な探索を実施できる可能性が示唆されたので報告する．

キーワード　探索的検索，可視化，閲覧履歴，自己組織化マップ

Browsing History Visualization Method for Web Collaborative searching

Junya EDA†, Satoshi SHIMADA†, Youhei SEKI†, Noriko KANDO††, and Tetsuji SATOH†

† Graduate School of Library, Information and Media Studies, University of Tsukuba
1-2 Kasuga, Tsukuba, Ibaraki, 305–8550 Japan
†† National Institute of Informatics
2–1–2 Hitotsubashi, Chiyoda, Tokyo, 101–8430 Japan
E-mail: †{eda,sat,yohei,satoh}@slis.tsukuba.ac.jp, ††kando@nii.ac.jp

1. はじめに

Webでの情報収集では，情報要求が曖昧な場合や調査対象が多岐に渡る場合も多い．このため，様々なページを閲覧し情報間の関係性等を考えながら検索と閲覧を繰り返す，探索的検索（Exploratory Search）の重要性が高まっている．探索的検索では，ユーザが過去に閲覧したページの履歴を保存した閲覧履歴を解析することが，ユーザの探索行動を理解する事は重要な手段となっている．しかし，その保存方法はページタイトルとURL等のリストを時系列順に蓄積していくものがほとんどで，訪問したページに記述された内容や，探索過程における話題の遷移を知るためには新たな履歴の保存と解析手法が必要とされていた．また，調査範囲が広い場合や調査対象が複雑な場合は，複数人で分担して行う協調探索が有効とされており，探索済みの範囲を他者に伝える，あるいは他者の結果を的確に把握する手法が必要とされていきた．

そこで，本論文では，単独のユーザが閲覧したページの履歴，更には，複数人による協調探索の過程を，キーワード平面上に可視化する手法を提案する．提案法では，探索範囲となる話題範囲を，キーワードの意味的関係を距離とする探索空間マップを作成し，ユーザが閲覧したページをマッピングしていく．さらに，閲覧ページ内に出現する語と一致する探索空間マップ上のキーワードに閲覧の程度を重畳することで，閲覧済の話題空間を可視化する手法を考案し，協調探索の課題である他ユーザとの探索空間共有を実現する．

このようなユーザ毎に異なる探索行動を探索空間マップ上で可視化することで，そのマップの充足率から各ユーザのWeb探索の達成度合いを示すことができる．さらに，時系列に沿ってマップの充足する様子から，ユーザの探索過程の閲覧した話題の，順序や範囲の把握を容易に行うことができるようになる．また，複数人のユーザに対して，それぞれのユーザが探索した範囲を空間的に可視化することによって，Webの協調探索にお

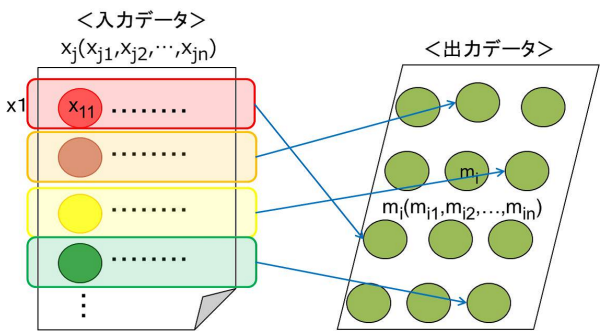

図 1　自己組織化マップの仕組み

ける，ユーザごとの探索した話題の範囲の違いの把握と共有を容易することができる．

2. 関 連 研 究

また，Honkela ら [1] は，特定分野の出版物や研究論文等が保管されているデータベースを対象に，自己組織化マップを用いて文書とともに研究分野や研究者などの属性をも可視化する media map を提案している．本研究では，可視化の対象は特定の分野のデータベースではないが，ユーザに与えられた検索課題に沿った特定の話題を表す文書集合から探査空間マップを作成するという点で類似している．

村上ら [2] は，ユーザの時系列の興味空間上に外化記憶の想起を支援する，興味空間ブラウザを提案している．興味空間ブラウザは，過去に閲覧したページの閲覧日時，クリックしたアンカーテキストを保存し，数量化 3 類を用いてキーワード群の相関を計算し 2 次元空間上に表示している．本研究とは，時系列上にユーザの過去に閲覧したページの情報からキーワードを抽出し 2 次元空間に可視化する点で類似しているが，その計算法と更にページのカバーする話題の範囲に着目するという点で本研究と異なっている．

さらに，探索的検索の支援の分析に関する研究として，種市ら [3] は，ユーザの Web 探索行動における情報評価に着目し，短期大学生を対象として Web 探索行動の実験調査を行っている．その中でユーザは，Web ページの情報を必要なものであるかどうか評価する際，タイトルやページ構成見やすさなどで評価することも多く，コンテンツの詳細な内容や複数の情報源を比較検証したりする本来重要なコンテンツの質的な評価は欠落する傾向が示されたとしている．本研究では，閲覧したページの表す内容が探索課題に対してどの程度マッチしているのかを提示するため，コンテンツの質的な評価の支援になるのではないかと考えている．

3. 自己組織化マップ:Self-Organizing Map

提案手法では，自己組織化マップにより探索空間マップを構成するため，ここで SOM の概要について説明する．自己組織化マップ (SOM: Self-Organizing Map) は、T. Kohonen により提案された教師なしのニューラルネットワークアルゴリズムで、高次元データを 2 次元平面上へ非線形写像するデータ解析方法である．自己組織化マップは、入力層と出力層により構成された 2 層のニューラルネットワークである．今，入力層には分析対象となる個体 j の特徴ベクトルを $xj(xj1, xj2, \cdots, xjn)$，出力層には 12 個のノードがあるとする．図 1 で示すように，出力層における任意の 1 つのノードは，入力層における特徴ベクトルとリンクしている．初期段階では乱数により各変数との間に重み $mi(mi1, mi2, \cdots, min)$ が付けられている．

(a) 入力層 と出力層におけるすべてのユニットの中から，最も類似しているユニット mc を探し出し，そのユニットを勝者とする．類似度は式 (1) によって計算される．

$$||xj - mc|| = min||xj - mi|| \tag{1}$$

(b) 勝者のユニットおよびその近傍のユニットの重みベクトル mi を更新する．更新は下記の式 (2) によって行う．

$$mi(t+1) = \begin{cases} mi(t) + hci(t)\lfloor xj(t) - mi(t)\rfloor & i \in Nc \\ mi(t) & i \notin Nc \end{cases} \tag{2}$$

$$hci(t) = \alpha(t)exp(\frac{-||rc - ri||^2}{2\sigma^2(t)}) \tag{3}$$

式 (3) の中の $hci(t)$ は近傍関数で，ユニット c とその近傍のユニット i の近さによって xj の影響を調整する．式 $hci(t)$ の中の $\alpha(t)$ は学習率係数で，rc と ri はユニット c と i の 2 次元上の座標ベクトルである．$\sigma^2(t)$ はユニット c の近傍領域 Nc の半径を調整する関数である．$\alpha(t)$,$\sigma^2(t)$ は学習回数 (あるいは時間) を変数とする単調減少関数である．学習回数を変数とする最も簡単な単調減少関数は単調減少関数である．この t は学習回数，T は事前に設定した学習の総回数である．

(c) すべての入力特徴ベクトル ($j = 1, 2, \cdots, m$) に対して (a)〜(c) を繰り返す

SOM の結果は，出力層の画面に図示される．個体の出力画面のユニットは，格子状 (正方形)，蜂の巣状 (六辺形) などが提案されているが，蜂の巣状が多く用いられている．蜂の巣状というのは，文字どおり蜂の巣のように正六角形のユニットを並べ，出力層の画面を構成する．出力層の画面は、上述のアルゴリズムにより、似ているもの同士を同じユニット、あるいはその近辺のユニットに配置する．

4. Web 協調探索のための探索履歴可視化手法

提案する探索履歴可視化手法の流れを図 2 に示す．提案法は，その処理を大きく分けて

I. 探索空間マップの作成

II. 探索的検索履歴マッピング

の 2 段階からなる．I では探索課題に対応する話題空間を構成する文書を解析し，SOM を用いてキーワードを意味的な関係性を距離で表した 2 次元の平面に表現する．II では作製したへ探索空間マップへユーザの閲覧履歴をマップのベクトルとキーワードベクトルを比較することでマッピングする．以下この手順に従って，各処理を説明する．

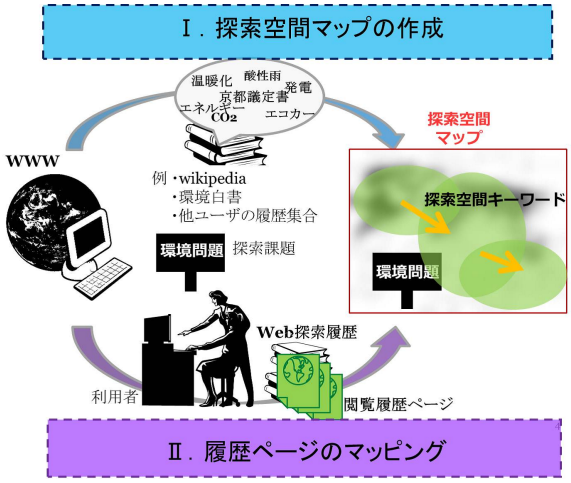

図 2　Web 協調探索のための探索履歴可視化手法

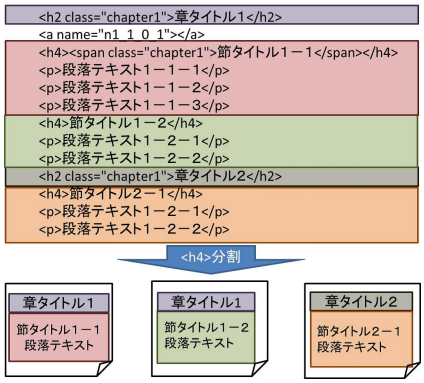

図 3　h4　分　割

4.1　探索空間マップの作成

探索空間マップは探索課題内の話題に関するキーワード同士の関連を SOM の計算によって距離を関連度とする 2 次元のマップで表現する．そのため探索課題に対応する文書からキーワードの関連性を示したベクトルを作成する必要がある．そこで本手法では html 文書内を，適切な粒度で分割し，その分割部分ごとに出現するすべてのキーワードの TFIDF 値を計算する．

具体的には，探索課題に関する html 文書の分割として html タグの $\langle h4 \rangle$ によって行う．html 文書を $\langle h4 \rangle$ で分割するイメージについて図 3 で示す．html 文書中には，様々なタグが存在しているが $\langle h4 \rangle$ で区切られる文章の多くは，意味のまとまりとなっておりその区切りで分割することによって，キーワード間の意味的関連を判別することができると考えたためである．

つまり文書を適切な粒度で分割しそこでのキーワード TFIDF で作成したベクトルから，SOM を用いることで意味的関連を距離に反映した，指定する縦 s_h 個 × 横 s_w 個のユニット，合計 $C_s(s = s_h \times s_w)$ のユニットを持つキーワードのマップを作成することが出来る．探索空間マップは，それら分割した html 文書すべてに出現するキーワードのうち SOM の計算によって指定するそれぞれのユニットに 1 つずつ割り振られた，キーワード C_s 個で構成されている．

4.2　探索的検索履歴マッピング

次に，作成した探索空間マップ上にユーザの探索的検索履歴をマッピングしていく．マッピングしたイメージを図 2 に示す．ここでは，ユーザの閲覧したページからキーワード出現数をもとにそのページのキーワードベクトルを計算し，探索空間マップのユニットそれぞれのベクトルと履歴文書のベクトルを比べ一番距離の近いユニットをその文書の位置として，履歴のマッピングを行う．ここで，閲覧ページ $P_i(i = 1, 2, \cdots, l)$ に出現するキーワードを $w_{ij}(j = 1, 2, \cdots, m)$ とし，ページ P_i が配置された SOM 空間上のユニットを C_i とする．探索空間マップは，閲覧ページを充分にマッピングできるだけのキーワード N 個 $w_k(k = 1, 2, \cdots, N)$ で構成されている．したがって，閲覧ページに出現するキーワード w_{ij} は，探索空間マップを構成する複数のユニットのいずれかに配置されたキーワードと一致することが，高い確率で期待される．そこで，あるページ P_i を閲覧したことで得られるキーワード w_k の**閲覧度** θ_{ik} を以下の式で算出する．

$$\theta(i,k) = \begin{cases} \frac{\eta_{ij}}{|C_i - C_k| + 1} & (w_k = w_{ij}\text{のとき}) \\ 0 & (w_k \neq w_{ij}\text{のとき}) \end{cases} \tag{4}$$

ここで，η_{ij} は閲覧ページ P_i に出現するキーワード w_{ij} の重要度を表すスコアである．C_k はキーワード w_k が配置された SOM 上のユニットであり，$|C_i - C_k|$ はユニット C_i と C_k の間の距離とする．また，

$$\sum_{j=1}^{m} \frac{\eta_{ij}}{|C_i - C_k| + 1} = 1$$

として，閲覧した個々のページの閲覧度合いの総計を正規化しておく．

l 個のページを閲覧した際に得られる閲覧履歴は，SOM 上のに配置された個々のキーワード w_k に対して，

$$\sum_{i=1}^{l} \theta(i,k) \tag{5}$$

で与えられ，ページの閲覧が進むにつれて，各キーワードの閲覧度合いが累積されていく．このように，閲覧したページのスコア（ここでは，閲覧度合いという）を，土台となる SOM 上に配置した個々のキーワードに累積することで，幅広い話題を扱ったページを閲覧した場合は，SOM 空間の幅広い範囲のキーワードにスコアが分散し，また，特定の狭い話題を扱ったページを閲覧した場合は，SOM 空間の局所のキーワードにスコアが重点的に配分されることになる．

更に，話題収集の広さの度合いを表す指標として，探索が進むにつれて探索空間マップ上の閲覧度が一定値を超えたキーワードの全体に対する割合を**マップ充足率**と定義し提案する．探索空間マップ上のキーワード C_s 個のうち，今回は閲覧度 θ_{ik} が 0 より大きいキーワードの割合をマップ充足率とした．閲覧度 θ_{ik} が 0 より大きいキーワードとは，探索履歴中の全ページ中で 1 度でも出現したキーワードということを示している．

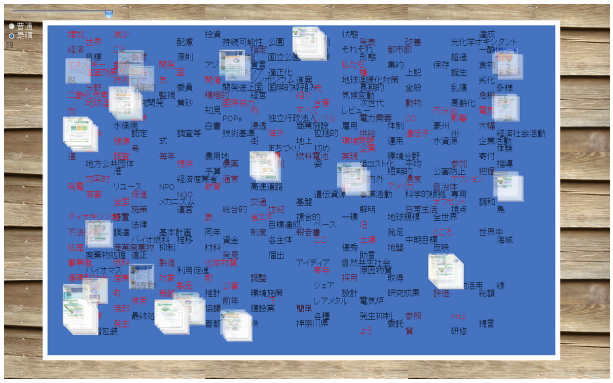

図 4　空間的把握手法を実装したシステム

5. 実装システム

前項で説明した提案法を実装したシステムについて各処理の実装を示す．I. 探索空間マップの作成段階では，Web 閲覧履歴を話題の遷移として扱うため，各ページをキーワードの集合で表す必要がある．キーワード抽出のための形態素解析には MeCab(注1)を用いる．文書から特徴的なキーワードを抽出する手法として形態素が名詞と判定されたものが連続する場合，つなげて一つのキーワードと判定する手法を実装した．また，キーワードベクトルの作成方法については，探索課題に関する話題の Web 文書の分割として仮に html タグの $\langle h4 \rangle$ によって行い，キーワードの TFIDF を計算し作成する．Web 文書を $\langle h4 \rangle$ で分割するイメージについて図 3 で示す．html 文書中には，様々なタグが存在しているが $\langle h4 \rangle$ で区切られる文章の多くは，意味のまとまりとなっておりその区切りで分割することによって，キーワード間の意味的関連を判別することができると考えたためである．さらに，SOM の計算には数値計算ソフトウェア MATLAB(注2)を使用し，Neural Network Toolbox を用いた．その計算結果より，JavaScript を用いて単語を空間的に配置することで，探索空間キーワードとする．II. 探索的検索履歴マッピングする処理においては，ユーザーの履歴ページから同様に形態素解析とキーワードの抽出を行い，各ページを 1 文書としすべてのキーワードを用いた TFIDF 値によるキーワードベクトルを作成する．そのベクトルと，探索空間マップ内のユニットのベクトル間のユークリッド距離を計算し最も値の近いユニットにページをマッピングしていく．更に，各ページの持つキーワードの閲覧度を，マップ上の探索空間キーワードに付与する．そしてその値の大きさに応じて，黒から赤へキーワードのカラーを変化させることで視覚的に表現する．図 4 に実際のシステム実行画面を示す．画面中央に広がる青い部分が探索空間マップであり，その中には空間内の話題の位置を示す，探索空間キーワードが配置されている．画面上左上には，時系列スライドバーを配置し，ユーザの履歴ページを時系列順でマッピングしていく際に使用する．また，ラジオボタンではページの表示方法を，時系列を追う際にすべてのページを累積して表示していくのか，その時系列のタイミングで閲覧していたページのみ表示するのかを選ぶことができるようになっている．履歴ページは，その URL をもとにトップページのサムネイル画像によって，探索空間マップ上に配置される．

(注1)：http://mecab.sourceforge.net/
(注2)：http://www.mathworks.co.jp/

6. 評　　価

6.1 評価に用いたデータ

実際の探索的検索の履歴データについて，Web 上の利用者の情報探索行動や認知プロセスに着目した研究プロジェクトである CRES [4] によって行われた実験の一部で以下のような課題に取り組んだ 14 名分の探索的検索履歴データを用い，提案法の各処理について評価を行った．

課題

次のような状況を想定して，課題に取り組んでください．あなたは，全国紙の新聞社でニュース雑誌のインターンとして編集の仕事をしています．あなたは編集長から，次号から始まる様々な**環境問題を紹介する連載記事の情報**を集めるよう指示を受けました．編集長は，2 時間後に行われる他の部門の編集長との編集会議で連載記事について話し合いをする予定です．編集長は，会議で連載 1 回分の記事について具体的な記事の内容を示したいと考えています．そのため **複数の問題に関する情報よりも，特定の問題についての記事を書くための詳細な情報**を必要としています．もちろん，読者の興味を引くものが良いと考えています．あなたに与えられた **時間は 45 分**です．サーチエンジンを使って情報を探し，編集長が求める情報を収集してください．

6.2 探索空間マップの作成法に関する評価

探索空間マップの作成に関して，作成されるマップにおける，話題クラスタの位置関係について実際の話題の意味的近さを反映したものになっているのかどうか評価した．今回用いた，データの探索課題は「環境問題」に関するものであることから，探索空間マップの作成には平成２２年版 環境・循環型社会・生物多様性白書 [5] の HTML 版を用いた．これは，環境に関する話題は環境・循環型社会・生物多様性白書にならば多く書かれ網羅されているだろうと考えられるためである．また，一つのまとまった文書であるため各章などで話題の遷移が明確であり，その文書から作成する探索空間マップでは，環境に関する話題の遷移構造を適切に表現出来ると考えたためである．探索空間マップの大きさについては 20 × 20 ユニットとした．

特定の文書内で，続いている章や節では，話題の流れが読み手が理解のしやすいように関連するものへ徐々に移り変わって行くと考えられる．そこで，話題クラスタの位置関係の評価について，正解とする文書内の連続する章や節の話題の流れを定め，提案する探索空間マップ上においてその遷移の様子を観測し遷移の総距離について評価する．正解文書から作られるキーワードベクトルが，探索マップ上のどのユニットに配置されたのか，その順序でユークリッド距離の合計を計算した．ここで，表 1 に正解とする話題遷移について示す．話題遷移 A と B は

表 1　正解データとする節の話題遷移

話題遷移 A	3 章 1 節 加速する生物多様性の損失
話題遷移 B	3 章 4 節 地球のいのちの行方を決める生物多様性条約
話題遷移 C	1 章 1 節 地球温暖化問題の現状

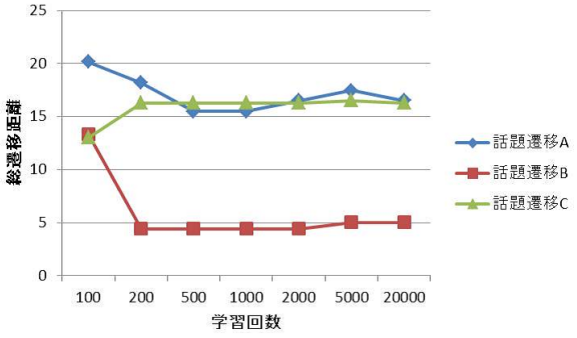

図 5　学習回数の増加による話題遷移の総距離の変化

同章内での違う節，話題遷移 C は別章の初めの節であり，話題遷移一つの節を一つのまとまりとして抽出した．また，話題遷移 D は，同章内で 2 つの節を横断するまとまりとした．

6.3　SOM の学習回数に関する実験

自己組織化マップ計算の際に，重要であるパラメータの一つに入力とする自らのベクトルの学習回数がある．自己組織化マップは，入力とする自らのベクトルの特徴を学習しながら少しずつ，マップ内のクラスタのベクトルを更新していく．この学習回数が，少なすぎれば特徴を反映したマップを作ることができない．本手法においては，キーワード間の意味的な関連を距離として正しく反映できないことになってしまう．また，学習回数は無限に増加させることができるがある程度で，マップ内部のベクトルの更新も収束してくることが考えられる．そこで，本手法において探索空間マップ作成時の SOM の学習回数を変化させたマップを複数作成し，その上に前項で示した，正解とする話題遷移をマッピングしていきマップ上での話題遷移の過程を，ユークリッド距離によって計算した．この場合も SOM 内のベクトルが安定していけば，マッピング結果の文書間のユークリッド距離も安定してくると考えられる．

図 5 は，探索空間マップ作成の際の SOM の計算において学習回数のパラメータを 100，200，500，1000，2000，5000，20000 回に変更した際の，各正解話題遷移をマッピングし総ユークリッド距離を比較したものである．話題遷移 A は，学習回数 100 回の探索空間マップの合計距離が最大であり学習回数が増えるにつれて総距離は短くなっていき，500 回以降は総距離の変化は少なくなっている．また，話題遷移 B と話題遷移 C においては学習回数 200 回あたりで学習回数を増加させても合計距離がほとんど変化しなくなっている．似た内容の話題遷移では，マップ上での遷移の移動も短くなることと，学習回数によってマッピング結果が大きく変動しないことが望ましいため，提案法において探索空間マップの作成の際の学習回数は 500 回程度が適切であると言える．

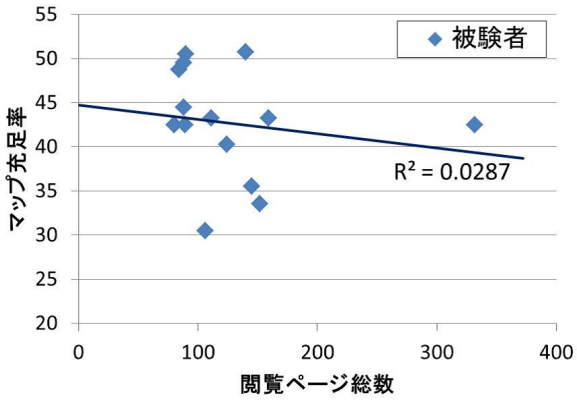

図 6　各ユーザの探索終了後のマップ充足率と閲覧ページ数

提案する，探索空間マップ上に特定の話題遷移を，マッピングすることで探索空間マップの評価を行った．話題遷移 D について，平成２２年版 環境・循環型社会・生物多様性白書から作成した 20 × 20 ユニット探索空間マップ上で，どのように遷移していくのかをグラフにて示したものが図**??**である．前項の実験より，学習回数は 500 回としその他のパラメータ等の条件もすべて，同じ値にし SOM の計算を行った．

6.4　一人の履歴マッピング結果の分析

図 6 は各ユーザの閲覧総ページ数と，探索後のマップ充足率を示したものである．各ユーザの閲覧ページ総数とそれぞれのマップ充足率の相関を計算したところ約-0.17 であり相関は無いことが示された．このことから，閲覧したページ数が多ければ多いほど，マップの充足率が上がるとは言えない．

ユーザは Web 探索を通して，様々な情報を入手しある程度の情報要求が満たされるとその探索を終える．その探索中の情報収集の過程の可視化という点について，本手法ではどの程度行えているのかを評価する．提案法では，探索空間マップにユーザが閲覧したページとキーワードをマッピングしてゆくが，そのマッピングされたキーワードの全対数に対する割合を，マップの充足率と定義した．探索空間キーワードは閲覧度の値を持っているが，閲覧履歴ページのマッピングによって，探索空間キーワードの閲覧度値が 0 以上となったものすなわち一度でも出現したものをカウントし，それらのキーワードの探索空間マップキーワードにおける割合が充足率となる．図 6 は各ユーザが探索を終了した際の，探索空間マップの充足率を示している．ユーザ間では，マップの充足率についてある程度の差が見られた．

次に，図 7 はそれぞれのユーザの探索空間マップの充足していく過程を示したものである．全ユーザで探索を行っていた時間は同じため，時間についてはより詳細なデータの解析が必要であるが，各ユーザ共に閲覧済みページ数の増加によって徐々にマップの充足率は増加していっていることがわかる．ここで，グラフの傾きはどれだけ少ないページで，マップの充足率を上げることができたかを示している．傾きについてはユーザ間では，ある程度の差がみられる．マップ充足率の高かった，u007 のユーザはここでも少ないページ数で，高いマップの充足率を

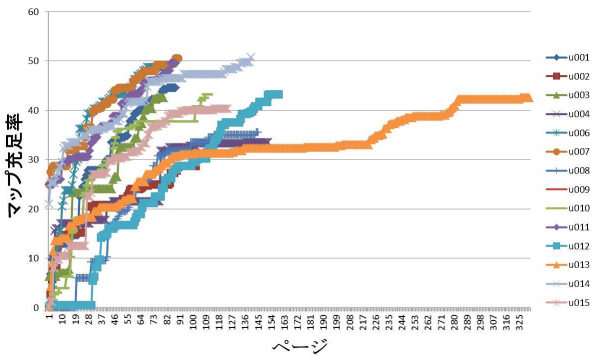

図 7　各ユーザの探索中のマップ充足の過程

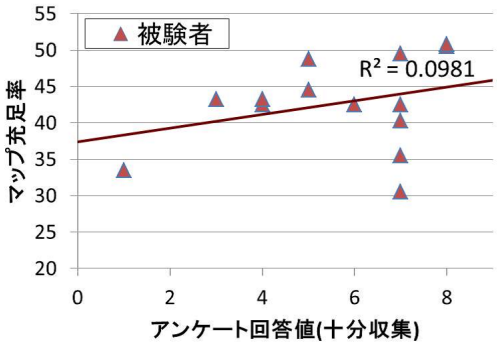

図 8　アンケート項目 (十分収集) の回答値とマップ充足率

得ている．このことは，他ユーザより多くのステップを経て，情報を見つけて行っているといえる．また，一番閲覧したページの多かった u013 ユーザのについても，緩やかであるがこのグラフから後半にも新たな情報を入手していることでマップの充足率の増加が起こっていることがわかる．

さらに実験後ユーザには，アンケートより時間は十分であったか，十分収集できたか等を 1 が不十分，9 が十分として 9 段階で評価し回答してもらった．図 8 は,「十分に収集できた度合い」に関する回答の値と，マップ充足率を示したものである．「十分の収集できた」と実際の充足率間の相関を計算した結果，決定係数 $R^2 = 0.0981$ であるものの，相関係数 $R = 0.32$ となりある程度の相関が示された．つまり，提案法によって得られたマップは，ユーザの主観的な収集度にも近い結果が得られているのではないかと言える．

6.5　異なる探索空間マップを用いた分析

探索空間マップの作成法の評価として，元となるデータの違いが完成されたマップ及び，その可視化結果や充足率にどのように影響しているのかを実験した．ここでは，環境・循環型社会・生物多様性白書をベースとした「環境白書ベースマップ」と，すべてのユーザが探索したページを収集しそれをもとにキーワード間の関連を計算しマップを作る「ユーザ閲覧ページベースマップ」について比較する．図 9 は，探索マップの違いによる，各ユーザ探索終了時のマップ充足率の様子である．ユーザ閲覧ページベースマップについて，ほとんどのユーザで探索後マップの充足率が上がっていることがみられた．閲覧ページ数が他ユーザと比べて多かった，u013 ユーザについてみてみると，環境白書マベースマップよりもその充足率は増加していることがわかる．これは，ユーザ閲覧全文書における u013 の閲覧した文書の割合が多いことが原因と考えられるが，その充足率の増加は他のユーザについても観測できる．また，その充足率の増え幅についても，他ユーザと大きな差は見られなかった．

表 2　2 つの探索空間マップの元となる文書群のデータ

	抽出キーワード数	抽出ブロック数
環境白書	15632	261
ユーザ閲覧ページ	14897	1788

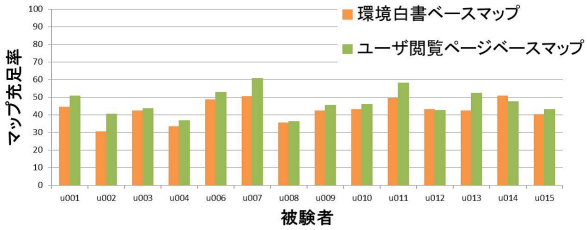

図 9　探索空間マップの違いによる充足度の違い

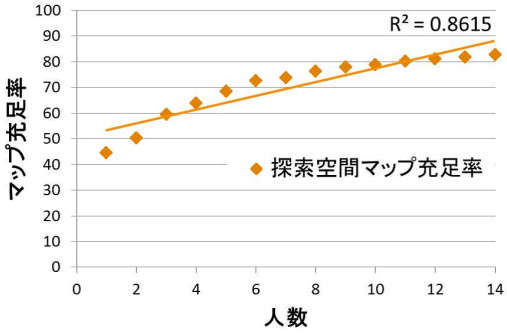

図 10　複数人でのマップ充足度の変化

6.6　複数人の履歴マッピング結果の分析

次に図 10 は複数人の履歴データを統合し，探索空間マップ上のでの充足率の増加を示したものである．

ここでは，人数の増加に伴って，マップの充足率が増加していることがわかる．履歴データを統合する人数と，マップの充足率についてそれぞれの項目間の相関を計算したところ 0.93 となり，高い相関があることが示された．それぞれのユーザ個人では，探索空間マップの充足度に極端な差は見られなかったことと対し，複数のユーザの履歴を統合した結果では，充足率は 1 人の場合の約 40%から 14 人で約 80%まで増加した．

図 11 は，ユーザ全員の履歴データを統合し，すべての探索を終えた際のシステムが示す，探索空間マップの充足の様子である．マップの左上の部分にマッピングされているページと左下にマッピングされているページが多いことがわかる．

7.　考　　察

7.1　探索空間マップの作成に関する考察

各評価実験の結果についてより詳細に考察を行う．自己組織化マップの学習回数に関する実験結果より，マップ内へのページマッピングの結果が 500 回程度で安定していることが示され

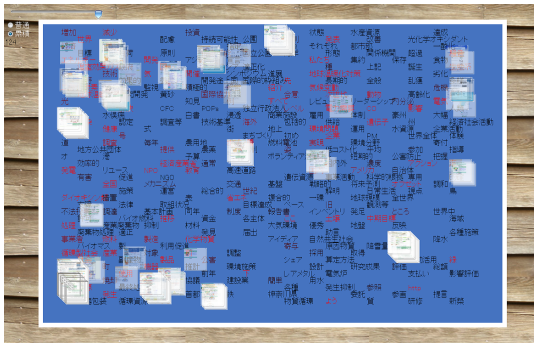

図 11　ユーザ全員の履歴を統合した探索空間マップの充足の様子

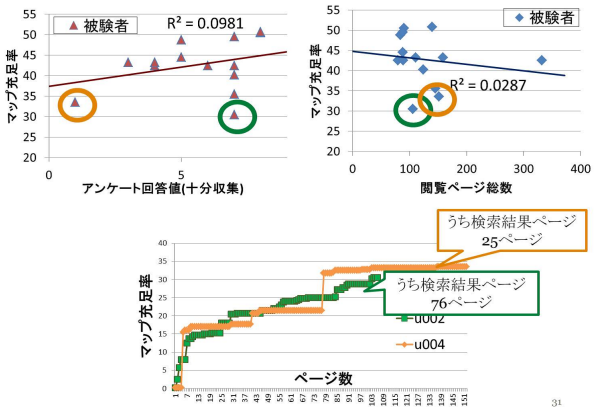

図 12　特定のユーザ 2 人に関して各値の比較考察

た．一般的に学習回数は完成するマップの数の 500 倍必要だといわれている．しかし，その値はあくまで一般化した値であり，本手法においては 500 回程度で十分にその効果が得られ，マップが安定していると考えられた．また，計算の対象とした環境・循環型社会・生物多様性白書内には，多くのトピックが存在しているがきちんと統制した文書と統制した語で書かれているため，その特徴がベクトルとして表現しやすく学習回数も少なくて済んだのではないかと考えることもできたが，ユーザが閲覧したページのみから作成した異なる探索空間マップとの比較によって，ユーザが探索を終えた際のマップの充足率については大きな差が見られなかったことがあり，手法はページの文体の統制や語の統制の影響はあまり受けていないと考えることができた．また複数人の履歴を統合することでマップの充足率を上昇させることができた結果より，ユーザごと探索した範囲に違いが探索空間マップによって表現できており，その違いが統合した際の探索空間マップの充足率増加に繋がっているのではないかと考えられた．

7.2　複数人での Web 協調探索実現に関する考察

複数人の履歴をマッピングした評価実験と分析から，ページ数の増加によるマップの充足率の増加の間での相関は計算の結果からは見られなかった．このことから，それぞれの人ひとりで探索できる話題の範囲というものはある程度決まっており今回行った 45 分程度では，その関連話題の多さにもよるがある探索課題に対して収集できる情報の量には限界があるのではないかということが考えられた．

しかし，ユーザのデータを統合し人数の増加によるマップの充足率の増加に関する相関は計算によってある程度確認することができた．このことから，人数を増やすことによって一つの探索課題に複数人で取り組む，事によりより探索空間マップを充足させることが出来ると考えられた．つまり，本提案における探索空間マップの充足は，探索課題に対する関連情報の網羅度に類似したものであるといえるため，複数人での協調探索は有効ではないかと考えることができる．さらに，本手法によって，探索空間マップ上で閲覧済み空間を把握したり，協調相手の閲覧話題の状況を把握できるため，協調探索へ適用することでのその実現を可能にできるのではないかと考えられた．

また，図 12 は，同程度のマップ充足率であるが，アンケートの十分収集の回答値に大きな違いのあった u002 と u004 の 2 名について各値を比較したものである．探索履歴内のページについて，改めて分析を行ったところ，この 2 人には検索結果ページの閲覧数に大きな違いがあったことが明らかになった．検索結果ページの閲覧数の多かった u002 については，検索結果をより詳細に深い順位まで吟味したか，または多くの検索を行ったと考えられる．この検索に関する何らかの行動の違いが，ユーザの十分収集できたかという主観的な満足度につながっている可能性が示唆されたため今後は，ユーザの探索中のページ滞在時間に注目して詳細な分析を行う必要もあると考えられた．

8.　ま　と　め

ある程度大きな探索的検索課題において，そのテーマを示すような文書をもとに探索空間マップを作成し，マップ上へ閲覧履歴ページを配置することで探索的検索の過程を可視化する手法を提案した．提案法によって，その探索空間マップの充足率から，ユーザの話題の収集の広さの度合いを示す指標を提案し，実験よりユーザの主観的な収集度合を反映した値であることを確認した．また，複数のユーザの探索履歴を統合することで，探索空間マップの充足率を増加させることができ，一人のユーザが多くのページを見る探索よりも，複数のユーザで同じテーマの探索を行うことの有効性が示唆された．

謝　　辞

本研究の一部は科研費 (21500091) の助成を受けたものである．ここに記して謝意を示す．

文　　献

[1] Timo Honkela, Jorma Laaksonen, Hannele Torro, and Juhani Tenhunen. Media map: a multilingual document map with a design interface. In *Proceedings of the 8th international conference on Advances in self-organizing maps*, WSOM'11, pp. 247–256, Berlin, Heidelberg, 2011. Springer-Verlag.

[2] 村上晴美, 平田高志. Www からの情報獲得・整理支援：思考・興味空間ブラウザ. 情報処理学会研究報告. 情報学基礎研究会報告, Vol. 2001, No. 20, pp. 167–174, 2001-03-05.

[3] 種市淳子, 逸村裕. エンドユーザーの web 探索行動：短期大学生の実験調査にもとづく情報評価モデルの構築. *Library and information science*, No. 55, pp. 1–23, 2006.

[4] Cres（cognitive research for exploratory search）. *http://cres.jpn.org*.

[5] 環境省. http://www.env.go.jp/.