行動連鎖を用いた情報検索支援とWebからの行動連鎖の抽出

旭　直人† 山本 岳洋†† 中村 聡史†† 田中 克己††

† 京都大学工学部情報学科 〒606–8501 京都府京都市左京区吉田本町
†† 京都大学大学院情報学研究科社会情報学専攻 〒606–8501 京都府京都市左京区吉田本町
E-mail: †{n.asahi,tyamamot,nakamura,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 本稿では，ある話題や対象に対してユーザがとる行動の連鎖を情報検索に活用する手法を提案する．ユーザが商品購入や，旅行など何らかの行動を起こそうと考え検索を行う場合，そのプロセスに何があるのか，何を注意しなければならないかなどを把握していないと調べることが困難である．我々は，こうした検索において，行動連鎖を発見し，ユーザに提示することで，行動に関するWeb検索やブログ検索の支援を効率化する．提案手法では，ブログに記述された行動情報を抽出し，シーケンシャルパターンマイニングの手法を用いることで，ブログエントリ間での行動連鎖の発見を試みた．また，ユーザのとる行動に役立つ情報を含むページが上位に来るように検索結果をランキングする手法の提案と評価を行った．最後にユーザ実験を行い，提案システムによってどのようにユーザの閲覧スタイルが変わるかを調査した．

キーワード 情報検索，行動指向ランキング，行動連鎖，ブログ

1. はじめに

近年，インターネットで提供される情報の多様化や各種サービスの充実に伴い，ユーザが実生活で直面する問題をインターネット上で解決できるようになってきた．ユーザはインターネット上で電車やバスの時刻検索，オンラインショッピング，レシピ検索，観光情報検索など実生活の問題解決に役立つサービスを気軽にいつでも利用できる．また，携帯電話の普及に伴い，調べる場所をも問わなくなってきた．将来的にはサービスの多角化が進み，実生活で困る問題のほとんどをインターネット上で解決できるようになると期待される．

しかし，依然としてディレクトリ型検索やキーワードベースの検索では特定のページを得る使い方しかできない．例えば，“初めて結婚式に招待されたがどうしたらいいかわからない”というような問題にユーザが直面し，検索エンジンを用いてこの問題を解決していこうとする状況を考える．結婚式へ招待された場合，ユーザは “招待状の返信をする” → “服の準備をする” → “挙式に参加する” → “受付をする” → “ご祝儀を渡す” → “披露宴に参加する” → “二次会に参加する” といったような流れをとると考えられる．ユーザは “結婚式に参加する” という目的を達成するために，先述の流れの中にある行動のそれぞれについてウェブ検索を行い，横断的にページを閲覧していくことになる．例えば，“招待状を返信する” という行動では，どのようにして招待状を返信すればいいのか，ということが書かれているページをユーザは探し，“ご祝儀を渡す” という行動では，ご祝儀はいくらが相場なのかといったことが書かれているページをユーザは探す．このように，結婚式に参加するために必要な情報が記述されているページを横断していくことで，ユーザは結婚式に参加するという目的のための情報を得ることができる．

しかし，初めて結婚式に参加するユーザだとこのような流れを予測することは困難である．そのため，招待状の返信の仕方を調べることはできても，ご祝儀のことを知らなければ，そもそも調べることができない．

このように実生活で直面する行動に関する問題をインターネット上で解決するには特定のページを見つけることよりも，その行動の流れを把握し，流れの中に含まれるそれぞれの行動を達成するページを横断的に発見することが重要となる．

そこで本稿では，ある話題やある対象について行われる一連の行動の連鎖を用いることにより，ユーザが実生活で直面する自分のとろうとする行動に関する情報検索を支援する方法を提案する．

提案手法では，ユーザがクエリをシステムに投げると，クエリにまつわる行動連鎖のグラフをユーザに提示する．そして，その中から行動をユーザが選択することにより，その行動を達成するのに必要な情報を含むページを得られるように検索結果を取得し，目的とする行動に応じて結果をランキングしてユーザに提示する．このようにすることにより，ユーザは対象とするクエリについて検索結果を行動する順番に整理して閲覧していくことができる．また，行動が整理されているため，行動を達成するための情報を効率よく得ることが可能となる．

本稿ではWebから行動連鎖を抽出するために，ブログから行動の連鎖の抽出を行った．ブログを対象としたのは，ブログでは書き手が特定できるうえ，ブログの書き手が実際に体験したことが記述されている可能性が高く，そのエントリの投稿日時が明記されているからである．ここでは，ブログエントリからユーザの行動群を抽出し，ブログエントリ内での行動の連鎖およびブログエントリ間での行動の連鎖を抽出したうえで，PrefixSpanを適用することで，行動連鎖の抽出を試みた．

最後に，行動連鎖に基づく検索システムが，自分のとろうと

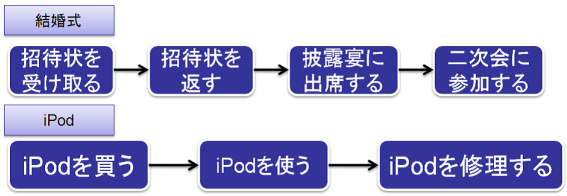

図 1　“結婚式” と “iPod” における行動連鎖の例

する行動を調べることにおいて有効であるか，従来のキーワードによる検索手法と比べてどのような違いが生じるかを調べるためにユーザ実験を行った．その結果，行動の把握，順序に沿った網羅的検索，将来必要になるかもしれない情報を入手することができる可能性を持つという点で，提案システムが有効に働くことが確かめられた．

2. 行 動 連 鎖

2.1 行動連鎖とは

例えば，結婚式に呼ばれた人は “招待状を受け取る” → “招待状を返す” → “披露宴に出席する” → “二次会に参加する” といったような行動をとる．また，“iPod を買う” → “iPod を使う” → “iPod が壊れて修理する” のようにある対象（ここでは iPod）に対して何らかの行動をとった後で対象に対する別の行動が発生することがある（図 1）．このように，ある話題の中で行われる一連の行動の流れを本稿では行動連鎖と呼ぶ．ここで，招待状を受け取る場合であれば “招待状” を “受け取る” という関係，披露宴に出席する場合であれば “披露宴” に “出席する” という関係，iPod を買う場合であれば “iPod” を “買う” という関係のように，動作にはその対象があることが多い，そこで本稿では，動作とその対象を組としてシーケンシャルに並べることで行動連鎖を実現する．

2.2 行動連鎖の持つ性質

行動連鎖には以下のような性質をもつものが考えられる．

プロセス性　行動連鎖の中には，ある目的を達成するために行わなければならない一連の行動があらかじめ定められているような手順と呼べるものがある．例えば，結婚式に呼ばれ参加を決めたユーザは “招待状を受け取る” → “招待状を返す” → “ご祝儀を用意する” → “披露宴に出席する” といったような一連の行動を行う必要があろう．また，バジルを育てる場合は “種を植える” → “間引きする” → “定植する” といったような一連の行動を取る必要があるだろう．このように行動連鎖にはプロセスを表す性質をもつものがある．

原因結果性　先述したプロセス性とは対照的に行動連鎖には，必ずしもある行動の後にはこれをしなければならないということはない，というものがある．例えば，2.1 節で述べた iPod に対する行動連鎖の場合では，“iPod を買う” → “iPod を使う” というものが考えられるが，ユーザは必ずしも iPod を買った後に使わなければならない，というわけではない．しかし，それにもかかわらず多くのユーザが類似した一連の行動をとっているということがある．

時期依存性　年賀状に対する行動連鎖では，例えば，“年賀状を買う” → “年賀状を書く” → “年賀状を送る” → “年賀状を受け取る” というようなものが考えられる．この場合，それぞれの行動は時期に依存している．“年賀状を買う” という行動は 11 月下旬以降であり，年賀状を書き，年賀状を送るのは 12 月，年賀状を受け取るのは 1 月になってからである．このように行動連鎖の中には時期に依存するものがある．

順序可換性　引越しをする場合，事前に “電力会社に連絡”，“水道会社に連絡”，“ガス会社に連絡” といった行動をとらなければならないと考えられる．これらの行動に実際とりかかるときにはその順序が存在するが，先に “電力会社に連絡” しようが，“水道会社に連絡” しようが，その順序に重要性はない．このように，行動連鎖の中には順序を入れ替えても問題がないものが存在する．

並列性　料理をする場合，“解凍しながら野菜を切る” → “煮ながら洗い物をする” → “テレビを見ながら食べる” というように，2 つ以上の動作を同時に実行することが多い．特に料理については技術の高い人であればあるほどこの並列性は高いといえる．このように行動連鎖の中には並列して 2 つ以上の動作が実行されるものがある．

繰り返し性　チューリップを栽培する場合，“チューリップを植える” → “チューリップの世話をする” → “チューリップの球根を回収する” → “チューリップを植える” →…，というように行動がループすることが考えられる．このように行動連鎖の中には同じ行動を繰り返す部分をもつものもある．

分岐　結婚式の例において，“招待状を受け取る” の後の行動には “出席する” と “欠席する” という 2 つの選択肢が考えられる．このように行動連鎖には次の行動が分岐して，以後の行動連鎖が大きく変わってくるものが存在する．

時間間隔　それぞれの行動と行動の間には時間間隔が存在する．その間隔は iPod の例であれば，“家電量販店に行く” → “機種を比較する” といったように数分～数時間単位のものから，“iPod を買う” → “iPod を修理する” というようにその間隔が数ヶ月～数年といったものまで粒度は様々である．以上のように，行動連鎖には行動と行動間に時間的間隔が存在し，その長さによっていろいろな粒度でみることができる可能性がある．

3. 行動連鎖を用いた Web 検索支援

3.1 基本的アイデア

本提案システムでは，図 2 のように，ユーザがクエリを入力すると，入力されたクエリにまつわる行動連鎖をユーザに提示する．その中からユーザが行動を選ぶと，システムはその行動に関して検索結果を取得し，ランキングした上でユーザに提示する．

システムがユーザに対してクエリに関する流れを提示し，その情報に基づく検索及びランキングを可能とすることで，ユーザはまず一連の行動の流れを一目で把握することができ，そして自分のとるべき行動に沿ってページを閲覧していくことが可能になる．

ユーザが入力したクエリに応じて関連するトピックを一覧で

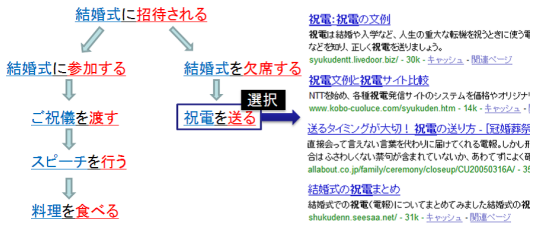

図 2　“結婚式” における行動連鎖を用いた検索の例

提示するシステムに Clusty(注1) などのクラスタリングサーチエンジンがある．一般にクラスタリングサーチエンジンでは名詞ベースでラベルを作り，検索結果をクラスタリングしている．単純に名詞だけではそれに対して何をするか分からない．例えば，“スピーチ” と提示されただけではスピーチを依頼するのか，スピーチをするのかが分からない．それに対し，本提案システムでは動詞と対象でラベルを表しているので，ユーザの実際にその名詞に対して何をするのかという観点から検索結果を眺めることができる．

また，実際の行動の順序にあわせて提示することで行動のつながりが明確になると期待できる．そして，ユーザは行動の分岐を見ることで自分にとって関係のある情報，関係のない情報の判断を容易に行うことができると期待できる．そのため，本提案システムで順序だてて行動連鎖をユーザに提示することで，ユーザは自分にとって必要な情報を効率よく調べることが可能となる．

そして，結婚式の場合，“招待された人” と “新郎新婦” のように立場が違えば，求める情報は変わってくるであろう．ここで，招待された人の行動連鎖と新郎新婦の行動連鎖を分けて提示すると，それぞれのユーザに合った検索結果を得ることができるようになるであろう．

3.2　システムの流れ

行動連鎖に基づく情報検索システムでは，予めブログを集め行動連鎖を求める部分とユーザにサーチ結果を見せる部分がある．ユーザとシステムの流れは図 3 の通りである．

（1）　ユーザはシステムにクエリを入力
（2）　システムは予め Web から抽出した行動連鎖を保存してあるデータベースからクエリに適合する行動連鎖を取得
（3）　システムは取得した行動連鎖に基づきグラフを描画し，ユーザに提示
（4）　ユーザは提示された行動の中から調べる行動を選択
（5）　システムは選択された行動に対する検索結果を検索エンジンを用いて取得し，ランキングを行ってからユーザに提示
（6）　ユーザは提示された検索結果の中からページを閲覧
（7）　(4) に戻る

4.　ブログからの行動連鎖の抽出

4.1　ブログの収集

データソースとしてブログを対象とする理由は，ブログには

(注1)：http://clusty.jp/

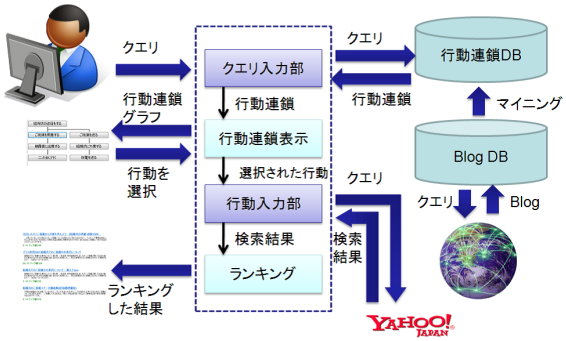

図 3　行動連鎖を用いた情報検索システムのイメージ図

実際に書き手が体験したことが記述されている可能性が高いこと，書き手が同一かどうか判定することが容易であること，投稿日時が明記されていることである．行動連鎖を集めるには，実際に同一人物がある対象に対してどのような行動を起こしていったのかという流れが重要となるため，実際の体験が記述されているブログは行動連鎖の抽出に最適であると考えられる．また，多数のブログから情報を収集することにより，一般的な流れだけでなく，意外な流れや最終的に失敗したという流れも得ることができるようになると期待できる．さらに，それぞれの行動と行動がその日のブログエントリとして投稿されていれば，その発生間隔を投稿日時の差から求めることができるであろう．

頻出する行動連鎖を抽出するためには，大量のブログからシーケンシャルパターンマイニングを行う必要がある．そのために，我々はブログを大規模に収集するコレクタを実装した．コレクタはブログサービスのドメインをサイト指定して，検索エンジンに問い合わせ，得られた URL からさらにブログ単位でサイト指定検索をするという簡易なものである．

4.2　ブログエントリ内での行動連鎖抽出

ある対象とそれに対する動作の流れを求めるために，エントリ本文をまず文単位に分解し，それぞれの文に対して形態素解析を行い，係り受け解析を行う．そして，動詞及びサ変接続名詞とそれに係る文節の組を抽出する．ただし，今回は，どの程度手法が実現できるかを調べるため，手動で十数語程度の辞書を作り，その語を含む文節のみを対象とした．そして，文のブログエントリ冒頭からの出現順に得られた組を並べる．これをブログエントリ内での行動連鎖として扱う．

4.3　ブログエントリ間での行動連鎖抽出

同一のブログのエントリでの解析結果を投稿日時順に並べ，連結する．そして得られた結果を 1 つのブログエントリ間でのシーケンスとし，これを大量のブログで行い，得た行動の系列に PrefixSpan を適用し，頻出する行動連鎖を抽出する．

4.4　行動連鎖抽出実験

“バジル栽培”，“出産”，“インフルエンザ” に関するブログの解析を行った．バジル栽培では，1462 ブログの 7078 件のブログエントリに対して，出産では，1383 ブログの 7732 件のブログエントリに対して，インフルエンザでは，1761 ブログの 8892 件のブログエントリに対してブログエントリ内，及びブ

表 1　ブログエントリ内の行動連鎖抽出例

成功例	バジル
	種をまく⇒水をかける
	苗をいただく⇒土を購入
	出産
	陣痛が来る⇒分娩室へ移動
	初期検査を受ける⇒妊娠が分かる
	インフルエンザ
	インフルエンザにかかる⇒タミフルを飲む
	タミフルを飲む⇒熱が下がる
失敗例	バジル
	種をまく⇒芽がくる
	水気をしぼる⇒水気をきる
	出産
	赤ちゃんに行く⇒赤ちゃんに送る
	病院にいく⇒病院に行く
	インフルエンザ
	咳が始める⇒熱が始める
	動物病院で受ける⇒治療を受ける

表 2　ブログエントリ間での行動連鎖抽出例

成功例	バジル
	苗を買う⇒苗を植える
	種をまく⇒芽が出る
	出産
	検診にかかる⇒陣痛が来る⇒出産費用を踏み倒す
	初期検査をうける⇒妊娠が分かる⇒赤ちゃんを産む
	インフルエンザ
	熱が出る⇒インフルエンザにかかる⇒タミフルを飲む
	病院に行く⇒予防接種を受ける⇒高熱が出る
失敗例	バジル
	芽を出す⇒種をまく
	水をあげる⇒水で作る
	出産
	出産費用に関する出産⇒出産費用を聞く⇒病院に聞く
	出産を終える⇒超音波で決める⇒赤ちゃんを産む
	インフルエンザ
	治療を受ける⇒インフルエンザにしまう⇒高熱が出る
	感染力を持つ⇒治療を受ける⇒タミフルが効く

表 3　行動連鎖抽出結果

	バジル	出産	インフルエンザ
エントリ内サポート値	3	4	3
エントリ内最大 GAP 値	6	5	3
エントリ内抽出個数	131	207	267
エントリ内正解率	39.7%	49.7%	38.2%
エントリ間サポート値	3	3	3
エントリ間最大 GAP 値	6	5	指定なし
エントリ間抽出個数	185	106	197
エントリ間正解率	27.0%	31.1%	49.7%

ログエントリ間の行動連鎖抽出を行った．ノイズを減らすため，バジルの場合では，“苗” や “種” といったバジルに重要な語を手動で十数語選び，その語を含む文のみを解析対象とした．出産，インフルエンザの場合でも同様に行った．そのようにして，ブログ本文を解析した結果を PrefixSpan に適用した．次に，ブログ本文の解析結果を時系列順に結合して，PrefixSpan に適用した． 抽出例は，表 1 と表 2 の通りである．抽出結果は，表 3 の通りである．抽出に失敗したものの原因として考えられるのは，係り受け解析の失敗や目的格を省略した主語を含む節だけを得てしまっているということである．その結果，日本語として正しくないものが得られてしまっている．また，ブログエントリで記述が現実の時間関係と前後して記述されている場合に意味的に間違った順番のものが得られてしまっている．エントリ内で過去の行動を振り返ったり，願望を述べたり，他人の行動に関して記述されていると，正しい順番が取得できない．さらに，同じものが何度も出たり，表記ゆれのものが含まれてしまっている．このような問題や精度の面を今後解決していく必要がある．また，重要度の低い系列や順番がおかしい系列が多いために正しい系列の頻度が少なくなり，精度が低くなっている．今後は正しい系列に重みをつけられるようなアルゴリズムを考えなければならない．

5. 行動に基づく検索結果のランキング

システムが行動連鎖を提示し，ユーザがそのうちの行動を選択したとき，その行動を実際に実行する際に必要となる情報を記述しているページを提示することが望ましい．しかし，単純にその行動をクエリにして検索するだけでは，その動詞を直接的に含むページがその行動を解決するのに必要な情報を提供しているページであるとは限らず，目的にあったページを得られることが難しいという問題がある．そこで本稿では，行動を達成するのに必要なページを得るための検索結果の ランキング手法を提案する．

5.1 df 差分法

ある対象に対する 1 つの行動を特徴付ける語集合は，その対象に対するそれ以外の行動と共起する語集合を除くと得られるのではないか，という仮定に基づき，2 つの検索結果間の語の出現頻度の差をランキングに利用する．

以下，主題とは，ユーザがシステムに投げたクエリのことであるとし，対象と動詞はシステムが提示した行動連鎖の中からユーザが選択した行動の動詞とその対象であるとする．df 差分法の流れは以下の通りである．

（1） クエリ $q =$ “主題 AND 対象 AND 動詞” で検索を行い，検索結果集合 R を得る．

（2） クエリ $q' =$ “主題 AND 対象 NOT 動詞” で検索を行い，検索結果集合 R' を得る．

（3） R 中に出現する全ての語 $t \in T(R)$ について，出現頻度の差 $sub(t)$ を次式に従って求める．（ただし，$T(R)$ は R 中に出現する語の集合）

$$sub(t) = df_t(R) - df_t(R')$$

ただし，$df_t(R)$ は R 中で t を含む検索結果の数．

（4） 次に得られた $sub(t)$ を用いて，R 中の全ての要素（タイトルとスニペット）$r \in R$ に対して，次式に従ってランク値 $Rank(r)$ を計算する．

$$Rank(r) = \sum_{t \in T(r)} sub'(t)$$

$$\text{ただし，} \quad sub'(t) = \begin{cases} sub(t) & (sub(t) > 0) \\ 0 & otherwise \end{cases}$$

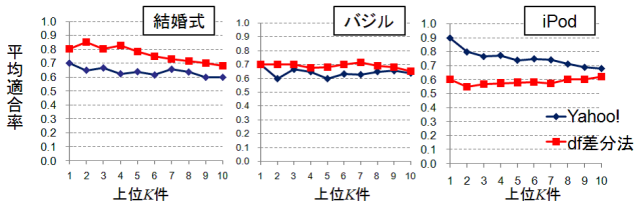

図 4　上位 K 件での平均適合率

（5）　r のタイトルが対象を含まなかった場合，その検索結果はユーザの求める話題とは関係のないページであることが多いため，(4) で得られた $Rank(r)$ を大きく減点し，下位へ下げる．

（6）　最後に，$Rank(r)$ の降順で検索結果を並びかえる．$Rank(r)$ が同じ場合は，並びかえる前の検索結果で順位が高かったものを上位に並べる．

5.2　評 価 実 験

df 差分法を用いてランキングした結果と再ランキングする前の検索結果との上位 K 件でのそれぞれ 10 クエリに対して行った時の平均適合率の比較を行った．ここでいう適合とは，ページがその行動を実行する際に役立つ情報を含んでいるかどうかであり，著者が判断を行った．

結果は図 4 の通りである．図中の Yahoo!とは，df 差分法を適用する前の Yahoo!からクエリ q =“主題 AND 対象 AND 動詞” で返ってきた検索結果で評価した場合の結果である．横軸は上位 K 件の K に相当し，縦軸は上位 K 件における 10 クエリでの平均適合率を表している．結婚式，バジル栽培においては df 差分法が有効に働いている．これは，df 差分法を用いることで，リンクしかはっていないようなページや多くの話題を扱っているページのトップページが省け，特定の行動を説明しているページが上位に来るため適合率が上がっていると考えられる．

しかし，iPod に対して行った場合適合率は df 差分法を適用する前と比べ，結果が悪くなってしまった．

これは，サプライヤ側のページがあまり行動に関する語を含まないために下位へ下がってしまうことや，df 差分法では語ベースで行っているために，古くてもはや有用でない情報が上位に上げられてしまう，といった問題が存在するためであると考えられる．

今後，これらの手法のそれぞれがもつ有用な点を見極め，組み合わせていく必要があると考えられる．

6.　実　　装

提案システムのプロトタイプを C# を用いて実装した．システムは図 5 のように 4 つのモジュールで構成されており，各モジュールが連携を行って動作する．

6.1　モジュール

6.1.1　クエリ入力部

クエリ入力部では，ユーザからのクエリが与えられると，行動連鎖を格納したデータベースに与えられたクエリを送信する．そしてデータベースから行動連鎖を XML 形式で取得する．

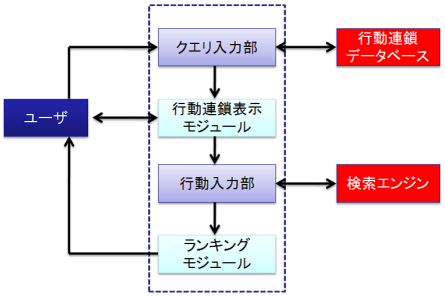

図 5　システムの設計

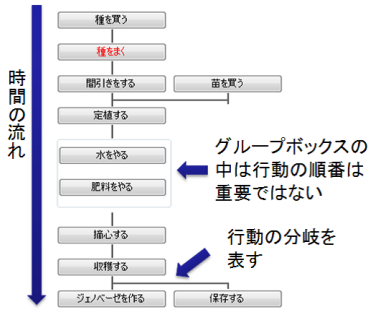

図 6　“バジル” に対して提示される行動連鎖グラフ

XML データには，それぞれの行動の目的語や動詞，次にどの行動につながるのかということが記されている．そして，受け取った XML データを行動連鎖提示モジュールに渡す．

6.1.2　行動連鎖提示モジュール

行動連鎖提示モジュールでは，クエリ入力部から受け取った行動連鎖の XML データをパースし，得られた情報から 1 つ 1 つの行動を表すノードとなるボタンを生成する．順序が重要でないものについては，グループボックスを作成し，グループボックス内にボタンを並べて生成する．また，どのノードがどのノードにつながるかという情報からノード間をつなぐ線を生成する．グラフは上から下に向かって時系列が進んでいくようになっている．その結果できた図 6 のようなグラフをユーザに提示する．ユーザが行動を選択した時（その行動をあらわすボタンをクリックした時），行動連鎖提示モジュールはその行動の情報を行動入力部へ渡す．

6.1.3　行動入力部

行動入力部では，行動連鎖提示モジュールから受け取った行動の情報から 5.1 節で述べた 2 つのクエリを作成し，検索エンジンにクエリを送信する．そして，検索エンジンから取得した検索結果をランキングモジュールに渡す．

6.1.4　ランキングモジュール

ランキングモジュールでは，行動入力部から受け取った 2 つの検索結果から 5.1 節で述べた計算を行い，検索結果を並びかえる．そして，並びかえた検索結果をユーザに提示する．

6.2　実　行　例

システムの実行例を図 7 に示す．この例の場合，ユーザはバジルの栽培に関する検索を行っている．“バジル” の行動連鎖の XML データを取得し，行動連鎖提示部に “バジル” に関連する行動連鎖をグラフとして展開している．ここで，ユーザが

図 7　プロトタイプシステム

“種をまく” というボタンをクリックすると，システムは右側のブラウジング領域に “バジル/種まき” や “種まきの豆知識” といったより種まきに役に立つページが上位に来るように検索結果をランキングして提示する．ユーザは随時この行動連鎖のビューにある “摘心する” や “苗を買う” といったような行動をクリックしていくことで，様々な情報を得ることができるようになる．

7.　行動連鎖に基づく情報検索の観察実験

7.1　実験概要

行動連鎖を用いた情報検索がどれだけ有効に働くか，またユーザの振る舞いが本システムによりどのように変化するかを調べるために，実装したプロトタイプを用いてユーザ実験を行った．

実験では，結婚式主催，iPod 購入，就職活動，バジルの栽培という 4 つの行動に関する詳細な前提条件を設定し，検索を行ってもらった．これらの 4 つのタスクについて，既存の Web 検索エンジンのみを用いる方法と，提案システムと Web 検索エンジンを併用する方法を比較した．ここでは順序効果を考慮し，ユーザを 2 つのグループに分け，片方のグループでは奇数番のタスクを Web 検索エンジンのみを用いて，偶数番目のタスクを提案システムと Web 検索エンジンを用いてタスクをこなしてもらった．もう一方のグループは，奇数番のタスクを提案システムと Web 検索エンジンを用いて，偶数番のタスクをを Web 検索エンジンのみを用いてこなしてもらった．なお，タスクを実行する時に提示する行動連鎖グラフの元になる行動連鎖 XML データについては理想の行動連鎖が与えられたという仮定のもと，手作業で作成した．実験ではユーザがタスクが完了したと判断したら次のタスクに進んでもらうようお願いした．全てのタスクが終わった後にユーザにアンケートを答えてもらった．

実験の被験者は 1 グループ 5 人，計 10 人の大学生であり，日頃よりコンピュータや検索などに慣れ親しんでいる．提案システムでどのようにユーザのブラウジングが変わるのかを分析するために，ユーザのブラウジングをユーザの後ろで監視するとともに，以下に挙げるログをとり，分析を行う．

表 4　ログの解析結果

	検索エンジンのみ	提案システム
平均閲覧時間（分）	9.1	12.8
平均閲覧ページ数	26.35	41.45
1 分当たりの閲覧ページ数	1.62	1.31
平均クエリ数	4.35	12.0
検索結果からアクセスしたページからさらに別のリンクをクリックする割合	75.4 %	33.2 %

- 訪れたページのタイトルと URL
- 使用したクエリ
- ブラウジングにかかった時間
- ページ滞在時間
- 押したボタンの種類とその順序

上記のログを解析し，考察を行った．

7.2　実験結果

ブラウジングログを解析した結果は，表 4 の通りである．提案システムを使うことで，閲覧時間は延び，閲覧ページ数も増えている．これは，検索エンジンのみを使う場合，ユーザはまず何をすればいいのかを調べるところから始まるため，話題に関してまとめてあるページを探さなければならない．そのために数個のクエリを試した結果，ユーザは話題に関してまとめてあるページを発見し，そのページ上で情報の収集を始めていた．そのページ上である程度の情報を得ることができるため，それでタスクは完了したと感じ，短くなる傾向にあると考えられる．

それに対し，提案システムを用いた場合では，すでにどういうことを調べればいいのかということがシステムによって示されているため，ユーザはシステムによって提示されたボタンをクリックし，初めからそれぞれの行動を深く調べるようになった．1 つの行動について詳しく解説しているページにたどり着きやすくなるため，ユーザのページ滞在時間は増えている．

また，検索結果でクリックしたページから他のページへ遷移しない割合は提案システムを使った場合，66 %も占める．これは，提案システムを使わない場合では，まとめページ内でトップページからそれぞれのコンテンツへ遷移しなければならないが，提案システムを使った場合だと直接それぞれの行動に関して説明しているページへアクセスできるためだと考えられる．

平均クエリ数が増えているのは，検索エンジンのみの場合では，まとめページを探すための数個のクエリしか作らないが，提案システムではボタンを押すことで，気軽に検索でき，多くのことを調べられるからだと考えられる．また，行動連鎖グラフを提示することにより，検索エンジンのみの場合とは異なったクエリを生成していた．

ページに滞在している時間の分布を求めると，図 8 のようになった．提案システムのほうが滞在時間が長くなっている．検索エンジンのみの場合で，5 秒以下の割合が大きくなっているのは，ページが不適合ですぐにページから出て行く場合とトップページですぐに他のリンクをたどる場合があるからだと考えられ，提案システムでページ滞在時間が長くなっているのは，先に述べたように直接詳細な情報を含むページにアクセスできるからだと考えられる．

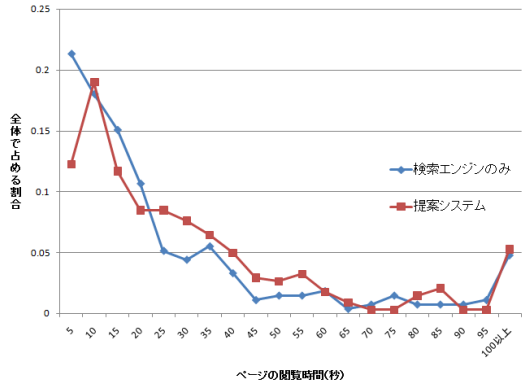

図 8　ページ閲覧時間の分布

表 5　アンケートの結果

質問内容	評価平均
本システムで自分の調べようとする話題を網羅できたか.	4.20
本システムを使うことで次に何をすべきかわかりましたか.	4.20
ボタンを押すことで出てくる検索結果はそのボタンに書かれた行動に関係のあるページでしたか.（適合性）	3.50
ボタンを押すことで出てくる検索結果のページをみて行動をするのに役立つ情報を得られましたか.（内容を見て）	3.90

アンケートでの結果は表 5 の通りである．ここでは被験者に 1 から 5 までのは 5 段階で評価してもらった．なお，数値が大きいほど肯定的な意見となっている．表 5 を見ると，網羅性と次に自分がとるべき行動を調べるという点で，提案システムはユーザの支持を得ていることが分かる．しかし，表 5 のランキング結果の適合性に関する項目の結果をみると，その結果は十分であるとはいえない．

提案システムに関して，ユーザから得た肯定的な意見は以下の通りである．

- 何を調べればよいかがある程度分かった
- 行動の流れをすぐ把握できる
- 行動プロセスが明示されることによって検索効率が非常によくなった
- 次に何をするべきか，将来的にこういうことが起こりそうだから先にこうしておこうというのに有用

上記の意見から，システムの良い点は，行動のプロセス提示により，ユーザは行動の流れを把握でき，行動に関して調べるべき点が分かる点であるといえる．また，ユーザが思いつかないようなページ，将来的に必要になるかもしれない情報を得ることに有用であることがわかった．

次にユーザから得た否定的な意見は以下の通りである．

- 検索結果の適合性に不満がある
- よく知らない語が提示されて困惑する
- やるべきことが本当にそれだけなのかという不安が残る
- 行動の粒度が適切でないことがある，プロセスの粒度を操作できるとよい

上記のように適合性に関しての不満点が見受けられた．表 5 での結果からもそのことがわかる．今後，ランキングのアルゴリズムの改善が課題である．また，提案システムによって知らないことが出てきてうれしい反面，バジル栽培の例において，“定植” や “摘心” といった一般になじみの薄い語が提示されているなど，その語を見てユーザは困惑してしまったのは問題である．実際ログを観察すると，まず wikipedia や百科事典，語の定義を説明してあるようなページを web 検索エンジンで検索して調べてから，その行動について調べていた．ユーザにとってなじみのない語をどのように見せていくか，ということも今後の課題であるといえる．

上記の被験者のコメントからも明らかなように，行動の粒度に関しては多くの人が注目していた．行動が多すぎると，ユーザは困惑してしまうし，逆に提示された結果が少なすぎても不満を抱くだろう．この問題を解決するには，初めは大きな粒度で見せておき，ユーザが行動を選択すると，細かい粒度で提示していく，という仕組みが必要となるだろう．また，その行動連鎖が本当にすべてを網羅しているのかといったことや，分岐がある場合に，どちらのほうが一般的で多くの人に行われているのかということを一律的に提示してしまっているために，ユーザは提示されている情報を把握しにくいという問題がある．こうした問題を解決するには，ノードの大きさを出現確率によって変えたり，得られた結果がどの程度尤もらしいのかを数値化して提示する必要がある．

以上が，ユーザ実験によって得られた知見である．行動に関する情報検索において，提案システムを用いることで，するべき行動を把握したり，順序に沿って網羅的に検索したり，将来必要になるかもしれないことまで調べたり，自分の知らないことまで調べることができるといった点で効率が高まるということが分かった．ユーザ実験によって挙げられた問題点は今後の課題である．

8. 関 連 研 究

8.1 PrefixSpan

大量のデータの中から頻出のアイテムの組合せを相関ルールとして発見するデータマイニングの技術 [1] が考案されているが，アイテムの出現順を維持したまま頻出するパターンを発見する手法 [2]〜[4] もいくつか考案されている．我々はその中で深さ優先探索で頻度の多い系列パターンを発見する PrefixSpan [5] をブログからの頻出する行動連鎖の発見に利用した．

8.2 ブログからの体験抽出

倉島ら [6] はブログ上から街の話題語を集め，その場所で人々が体験することを集約し，地図上にマッピングし，ユーザに提示している．この手法では対象を地名・ランドマークと助詞の共起に着目して抽出している．また，倉島ら [7] はブログから状況（時間，空間），行動（動作，対象），主観（感情，感情）をマイニングし，相関ルール抽出技術を用いて，状況，行動，主観との間の関係をルール形式で抽出を行っている．乾ら [8] や阿部ら [9] は，トピック，経験主，事態タイプ（ポジティブ/ネガティブな出来事・状態・性質，入手・利用等の行為），事実性という 4 つの軸に基づいてブログから経験をマイニングする．これらはブログから体験情報をマイニングするという点で本研究と同じであるが，行動と行動の関係は考慮されていない．

8.3 ブログからの代表的な行動経路とコンテキスト抽出

郡ら [10] は，ブログに記述された地名と場所や移動に関係する助詞に注目し，ブログの書き手が実際に辿った経路を抽出し，それを大量に集め，PrefixSpan を適用することで代表的な行動経路を求めている．また，経路を通った人が共通に含むテーマをコンテキストとして捉え，各文書に現れる名詞から特徴ベクトルを作り，経路上のコンテキストを求めている．行動の流れを追うという点で本研究と同じであるが，場所と場所の移動に限定している点が本研究と異なる．

8.4 実世界セマンティクス

長沼ら [11] は，インターネット上のサービスに実世界の人間行動との関連性をメタデータとして付与することにより，問題解決に必要とされるサービスを容易に発見できるシステムを提案し，またオントロジーの構築方法についても議論している．携帯電話向けに提供されている Web コンテンツからランダムに抽出し，実際にそれらのサービスが実現するものを分類している．そして実生活で直面する問題を階層化し，概念木を作成している．そのサービスが現実世界のどの問題を解決するのかということを結び付けている点が本研究と似ているが，ページに対してメタデータを付与している点が異なる．本研究では，行動連鎖に着目してページの分類を試みたものである．

9. ま と め

本稿では，行動連鎖を用いて情報検索を支援するための手法の提案を行った．ここでは，ブログのブログエントリ内，ブログエントリ間という２つの観点からシーケンシャルパターンマイニングを用いて行動連鎖を抽出する手法を示し，提案システムのユーザ評価を行うことにより，ユーザがとろうとする行動に関して情報検索する際に，ユーザのブラウジングスタイルが大きく変わることが明らかになった．また，システムの問題点，改善点も洗い出すことができた．今後は再ランキングアルゴリズムの改良を行う予定である．また，行動連鎖をブログに限らず Web からも抽出することで，行動連鎖の抽出の精度を高めていく予定である．そして，行動連鎖に基づいてページを推薦したり，立場の違う人の行動連鎖を分離してうまく提示する方法，検索クエリログと行動連鎖の関連性についても今後考えていく予定である．

謝辞 本研究の一部は，グローバル COE 拠点形成プログラム “知識循環社会のための情報学教育研究拠点”，計画研究 “情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究”(研究代表者：田中克己，A01-00-02, 課題番号 18049041)，によるものです．ここに記して謝意を表すものとします．

文　献

[1] Agrawal, R., Imielinski, T. and Swami, A.: Mining association rules between sets of items in large databases, *Proc. of the 1993 ACM SIGMOD international conference on Management of data*, pp. 207–216 (1993).

[2] Srikant, R. and Agrawal, R.: Mining Sequential Patterns: Generalization and Performance Improvements (1996).

[3] Zaki, M.: SPADE: An Efficient Algorithm for Mining Frequent Sequences, *Machine Learning*, Vol. 42, No. 1, pp. 31–60 (2001).

[4] Ayres, J., Flannick, J., Gehrke, J. and Yiu, T.: Sequential PAttern mining using a bitmap representation, *Proc. of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 429–435 (2002).

[5] Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U. and Hsu, M.: PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern, *IEEE Int. Conference on Data Engineering* (2001).

[6] Kurashima, T., Tezuka, T. and Tanaka, K.: Mining and Visualization of Visitor Experiences from Urban Blogs, *Proc. of the 17th International Conference on Database and Expert Systems Applications (DEXA 2006)* (2006).

[7] 倉島健, 藤村考, 奥田英範: 大規模テキストからの経験マイニング, 電子情報通信学会 第 19 回データ工学ワークショップ (2008).

[8] 乾健太郎, 原一夫: 経験マイニング：Web テキストからの個人の経験の抽出と分類, 言語処理学会第 14 回年次大会論文集, pp. 1077–1080 (2008).

[9] 阿部修也, 江口萌, 隅田飛鳥, 大崎梓, 乾健太郎: みんなの経験：ブログから抽出したイベントおよびセンチメントのＤＢ化, 言語処理学会第 15 回年次大会 (2009).

[10] 郡宏志, 服部峻, 手塚太郎, 田島敬史, 田中克己: ブログからのビジターの代表的な行動経路とそのコンテキストの抽出, 電子情報通信学会技術研究報告. DE, データ工学, Vol. 106, No. 149, pp. 29–34 (2006).

[11] 長沼武史, 礒田佳徳, 倉掛正治: 人間行動に基づく実世界セマンティクスの構築, 情報処理学会研究報告. 情報学基礎研究会報告, Vol. 2003, No. 98, pp. 55–60 (2003).