

操作イベントの時間分析による

WEB 閲覧履歴検索・提示手法

森田 哲之

システム情報工学研究科  
筑波大学

2010 年 3 月

# 目次

1. はじめに	1
2. WEB 閲覧履歴活用のためのステップ	4
2.1 履歴収集, 履歴解析, および情報提示ステップ	4
2.2 履歴収集ステップにおける履歴収集場所での分類	5
2.3 履歴解析ステップおよび情報提示ステップにおける人間の介在の有無による分類	7
2.4 本研究の位置づけ	10
3. 自分自身による WEB 閲覧履歴検索・提示手法および試作システムの実装	12
3.1 体験獲得情報を想起させる WEB 閲覧履歴検索	12
3.2 体験獲得情報の定義	13
3.3 PC 上での詳細な WEB 閲覧履歴の収集	14
3.4 期間に着目した WEB 閲覧履歴の解析	15
3.4.1 アクティブ期間の抽出, および注目度と重要度の算出	15
3.4.2 集中期間の抽出と重要度の算出	18
3.4.3 WEB ページの重要度の抽出	21
3.5 集中期間検索インタフェース	22
3.6 行動シーケンス表示インタフェース	24
3.7 WEB ページ検索インタフェース	27
3.8 自分自身による WEB 閲覧履歴検索・提示手法の試作システムの実装	28
3.9 自分自身による WEB 閲覧履歴検索・提示手法に関する関連研究	30
4. 自分自身による WEB 閲覧履歴検索・提示手法の評価	32
4.1 評価対象とするインタフェース	32
4.2 比較対象と実験条件	32
4.3 評価実験の手順	34
4.4 評価実験の結果と考察	35
4.5 公開実験と WEB アンケートによる評価	39
4.6 今後の展望	44
5. 第三者による WEB 閲覧履歴提示手法と試作システムの実装	46
5.1 対話型オンラインショップ	46
5.1.1 対話型オンラインショップの機能	46
5.1.2 対話型オンラインショップにおける顧客の嗜好情報の必要性	49
5.1.3 抽出対象とする顧客の嗜好情報	50
5.2 WEB サーバのアクセスログによる閲覧履歴の収集	50
5.3 対話型オンラインショップの店員に提示するための閲覧履歴の解析	51
5.4 顧客の嗜好情報抽出のための WEB 閲覧履歴提示インタフェース	55
5.4.1 Bubble-graph 提示インタフェース	55
5.4.2 Category 提示インタフェース	57
5.4.3 Highlight 提示インタフェース	59
5.5 第三者による WEB 閲覧履歴提示手法の試作システムの実装	61
5.6 第三者による嗜好抽出手法に関する関連研究	63
6. 第三者による WEB 閲覧履歴提示手法の評価	65
6.1 評価実験の手順	65
6.2 評価実験のパラメータ	66
6.3 評価実験の結果と考察	69
6.3.1 Bubble-graph 提示インタフェースの評価結果と考察	69

6.3.2	Category 提示インタフェースの評価結果と考察	70
6.3.3	Highlight 提示インタフェースの評価結果と考察	72
7.	結論	74
	謝辞	76
	参考文献	77
	著者論文リスト	82

## 図一覧

図 1-1	インターネットの利用頻度	3
図 1-2	パソコンからのインターネット利用機能・サービス	3
図 2-1	WEB 閲覧履歴活用のためのステップ	4
図 2-2	履歴分析，情報提示ステップでの分類	9
図 3-1	行動履歴自動収集モジュール	14
図 3-2	アクティブ期間の開始	16
図 3-3	直前のアクティブ期間の終了と新しいアクティブ期間の開始	16
図 3-4	アクティブ期間の抽出と注目度の概念図	18
図 3-5	再検索キーワードに対する適合度の概念図	18
図 3-6	集中期間の抽出	20
図 3-7	集中期間検索インタフェース	23
図 3-8	行動シーケンス表示インタフェース	25
図 3-9	注目箇所表示	26
図 3-10	WEB ページ検索インタフェース	27
図 3-11	試作した WEB 閲覧履歴検索・提示システムの流れ	29
図 4-1	従来手法の画面イメージ	33
図 4-2	実験手順（体験ステップおよび想起ステップ）	35
図 4-3	全被験者の平均獲得スコア	37
図 4-4	全被験者の単位時間当たりの平均獲得スコア	37
図 4-5	アンケート結果（役立つ機能）	38
図 4-6	アンケート回答者の年齢（N=647）	40
図 4-7	アンケート回答『履歴を保存することに対する現在の意識を教えてください』（N=647）	41
図 4-8	アンケート回答『期間単位と WEB ページ単位がありますが，どちらのほう が使いやすかったですか』（N=369）	42
図 4-9	アンケート回答『期間単位がよいと感じた理由はなぜですか？』（N=148）	42
図 4-10	アンケート回答『行動詳細表示のランク表示（赤，オレンジ，青）は重要だ と思ったページを示していますが，適切でしたか』（N=148）	43
図 4-11	ユーザの負荷なしに集中期間を気付かせるインタフェースの例	44
図 4-12	気付き機能と，3 章および 4 章で述べたインタフェースとの関係図	45

図 5-1	対話型オンラインショップの構成	47
図 5-2	対話型オンラインショップの顧客向け画面の例	48
図 5-3	実店舗陳列棚画像表示	48
図 5-4	商品一覧ページ表示	49
図 5-5	短期的嗜好，長期的嗜好，および商品購買時の嗜好の関係	50
図 5-6	Bubble-graph 提示インタフェース 1	56
図 5-7	Bubble-graph 提示インタフェース 2	56
図 5-8	Category 提示インタフェース 1	58
図 5-9	Category 提示インタフェース 2	58
図 5-10	Highlight 提示インタフェース 1	59
図 5-11	Highlight 提示インタフェース 2	60
図 5-12	試作システムの構成図	62
図 5-13	店員向けアプリケーションの画面例	62
図 6-1	リスト提示インタフェースで表示する項目	66
図 6-2	リスト提示インタフェースの画面例	66
図 6-3	Bubble-graph 提示インタフェースの正解率	70
図 6-4	Category 提示インタフェースの正解率	71
図 6-5	Highlight 提示インタフェースの正解率	73

## 表一覧

表 2-1	履歴収集ステップでの分類 .....	6
表 2-2	人間を介すか否かによる特徴 .....	9
表 2-3	本研究における WEB 閲覧履歴活用のための各ステップの概要 .....	11
表 3-1	HTML タグによる重み .....	21
表 4-1	被験者の属性 .....	33
表 4-2	注目度算出パラメータ .....	34
表 4-3	タスク例および被験者の記述例 .....	35
表 6-1	実験参加者に指示した顧客の状況リストの例 .....	68

## 1. はじめに

多くのユーザがインターネットを利用した体験を日々積み重ねている。

2008年および2009年に総務省が行った調査[1]では、パソコン経由でインターネットを利用する者のうち、毎日利用している者は約4割、少なくとも週に一回は利用するという者を合わせると7割以上といった報告がなされている(図1-1)。そして、WEBページの閲覧は代表的なコンピュータの利用方法であることが報告されている(図1-2)。

それらWEB閲覧の履歴を活用し、過去のWEB閲覧で獲得した情報や、興味を持った内容を効率的に探し出すことは有用だと思われる。

例えば、家族旅行に行く計画をたて、温泉、旅館、グルメや観光スポットをインターネットを使って調べるとしよう。一度に決めることができればよいが、空き時間を見つけて徐々に調べる場合も多い。そのとき、『確か良さそうな雪見温泉があった』といった、明確には思い出せないが『過去に見た』という体験を思い出し、その温泉のホームページや温泉への行き方など、過去に得た知識を再度見つけ出そうとすることは多いのではないだろうか。このようなとき、過去に得た知識を短時間で思い出すことができれば、その温泉についての口コミをさらに調べ、温泉宿の空室情報のチェックや予約を行うなど、新たな行動を効率的に行うことができる。

WEB閲覧履歴を活用し、自分が過去に抱いていたが忘れていたり、あるいは自分自身でも気付かずに潜在的に抱えている嗜好情報を探し出すことも興味深い。

例えば、近年広く行われているブログなどで、自分が興味を持った事柄を書きたいときがある。しかし、いざ思い出そうとしても、漫然としていて、自分が何に興味を持って過ごしていたのかを思い出せず、書くのをあきらめてしまうことがある。あるいは、実は、ビーチバレーについてのインターネット上の記事を多く読んでおり、ビーチバレーに興味を持ち始めているといった、潜在的な自分への気付きにより、興味が広がっていくことも考えられる。

上で述べた例は、自分自身で自分のWEB閲覧履歴を活用して知識や嗜好情報を抽出する例であるが、第三者がWEB履歴を活用して知識や嗜好情報を抽出することも広く行われている。

例えば、ネットワーク上のサービス提供者は、利用者のWEB閲覧履歴から嗜好情報を抽出することで、個々の利用者に適したお勧め情報を提示している。このようなきめ細かいサービスを提供することで、売り上げを増加させ、顧客満足を高めることができる。また、WEB閲覧履歴から訪問者ごとの嗜好情報を抽出し、ホームページにあらかじめ用意さ

れた広告を表示する領域に、訪問者ごとに最適化された広告を表示することで、広告収入を高めることが行われている。また、大人数の WEB 閲覧履歴を収集して世の中の流行をいち早く探し出すといった、マーケティング情報を獲得する手段としても活用されている。このように、WEB 閲覧の履歴を活用し、過去に得た知識や、興味を持った内容を効率的に探し出すことは有用だと思われる。

本論文では、WEB 閲覧履歴を自動的に収集し、時間に着目して解析し、ビジュアルにユーザに提示することにより、過去に得た知識や、嗜好情報を獲得させる手法を述べる。

2 章では、WEB 閲覧履歴を活用するために必要な、履歴収集、履歴解析、および情報提示という 3 つのステップについて述べ、それらステップを整理する。

3 章では、自分自身の WEB 閲覧履歴と印刷などの PC 上の詳細な履歴を利用する WEB 閲覧履歴検索・提示手法について述べる。詳細な WEB 閲覧履歴を収集する手法を示すとともに、再検索のために入力したキーワードに関連した行動を集中して起こした期間および WEB ページをアクティブに表示している期間に着目し、それらの重要度を算出する手法、および WEB 閲覧履歴を効果的に提示するためのインタフェースについて述べる。

4 章では、3 章で述べた自分自身による WEB 閲覧履歴検索・提示手法を評価し、従来手法に比べ、過去に得た知識を、短時間でより多く抽出できることを示す。また、試作システムを一般公開し、数百人規模で行った WEB アンケートの結果を示す。

5 章では、顧客の商品ページ閲覧履歴を利用して、サービス提供者が顧客の嗜好を抽出する手法、および試作システムについて述べる。実店舗の商品棚をリアルタイムに提示して時間とともに変わる商品を販売する EC ショップ、また、顧客と店員が TV 電話で対話する EC ショップを想定し、店員が顧客と対話する直前に、顧客の WEB 閲覧履歴を見て顧客の嗜好情報を短時間で抽出させる閲覧履歴解析手法、および履歴提示インタフェースを示す。

6 章では、5 章で提案した商品閲覧履歴の提示手法により、第三者がどれだけ顧客の嗜好情報を抽出できたかを評価する。従来手法に比べて、提案手法が短時間でより正確にユーザの嗜好を抽出できることを示す。

最後、7 章でまとめを述べる。

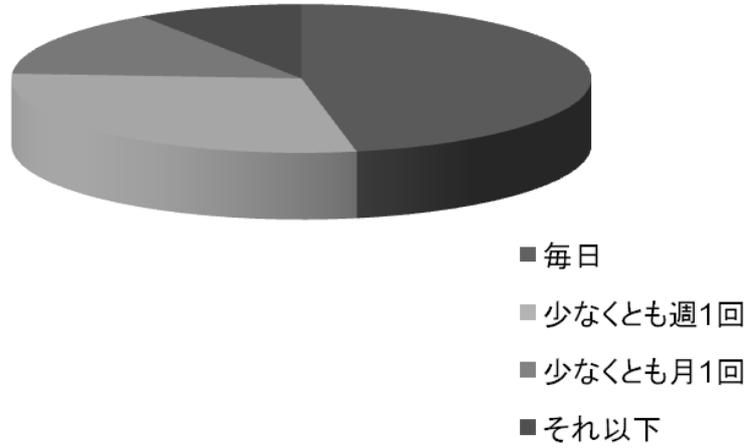


図 1-1 インターネットの利用頻度

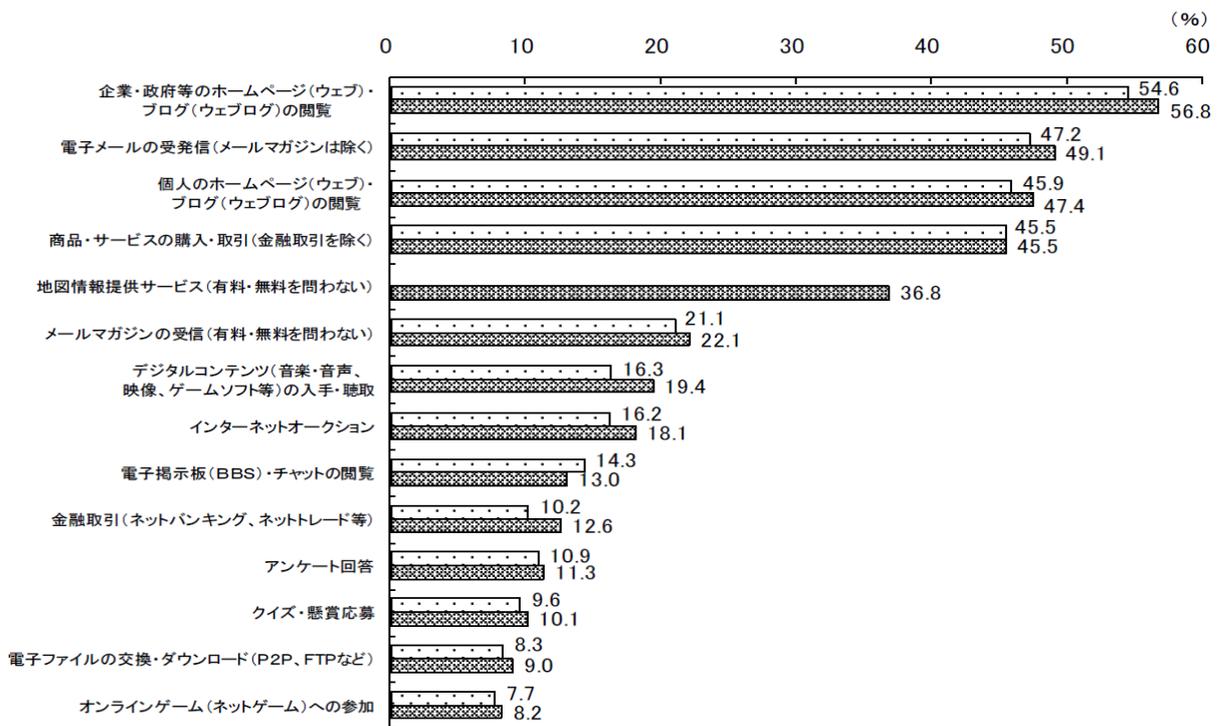


図 1-2 パソコンからのインターネット利用機能・サービス

## 2. WEB 閲覧履歴活用のためのステップ

### 2.1 履歴収集，履歴解析，および情報提示ステップ

WEB 閲覧履歴を活用するためには，大きく分けて 3 つのステップが必要だと考えられる（図 2-1）。

最初に必要となるのが，WEB 閲覧履歴をどのように収集するかという，履歴収集ステップである．できるだけ広い範囲で詳細な履歴を収集することが望まれるが，現実のシステム構築にあたっては，ユーザ負担やプライバシーに考慮する必要がある．次に必要となるのが，収集した WEB 閲覧履歴を解析し，閲覧履歴の特徴や最終的に目的とする情報を抽出する履歴解析ステップである．広く詳細な履歴を収集した場合，WEB 閲覧履歴は大量に収集されるため，それら履歴の中から意味のある情報を抽出する必要がある．最後に，解析した履歴を利用して，利用者に情報を提示する情報提示ステップがある．提示する情報には，その人に合ったニュース記事，WEB 広告，あるいは閲覧履歴自体などが考えられる．

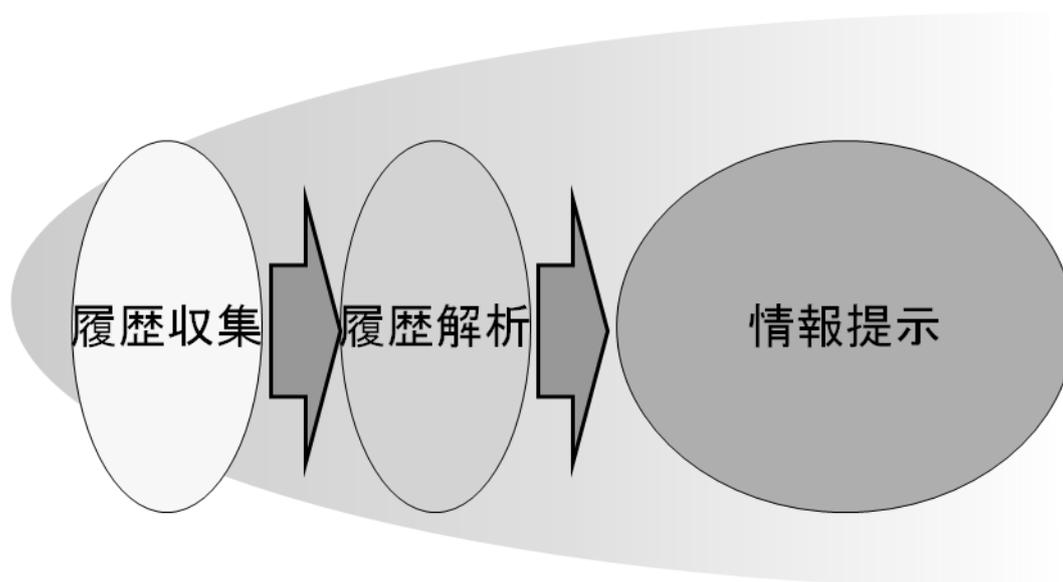


図 2-1 WEB 閲覧履歴活用のためのステップ

## 2.2 履歴収集ステップにおける履歴収集場所での分類

PC 上での WEB 閲覧履歴を収集する履歴収集ステップは、履歴を収集する場所によって大きく 3 つに分類される。一つは、WEB ページを提供する WEB サーバで履歴を収集するもの、もうひとつは、WEB ページを閲覧するローカルの PC 上で履歴を収集するもの、もうひとつはプロキシサーバにより WEB サーバとローカル PC の間で履歴を収集するものである。それぞれの特徴を表 2-1 に示す。

WEB サーバ側で履歴を収集する場合、ユーザの環境にはなにも手を加える必要がない。ユーザには非常に負担が少ない方法といえる。また、WEB サーバは WEB ページを作成するためのデータベースと連携しているため、例えば、WEB ページに掲載している商品はこの商品であるなど、WEB ページについてのメタ情報を正確に獲得できることが多い。しかし、基本的には WEB サーバ単位でユーザの履歴情報が蓄積するため、WEB サイトを横断してユーザの履歴を収集することが難しい。つまり、ショッピングモール A での WEB 閲覧履歴は取得できるが、そのユーザが次に移ったショッピングモール B での WEB 閲覧履歴は取得できない。また、WEB サーバにアクセスしたという履歴は取得できるが、アクセスした後の履歴については取得できない。つまり、WEB ページにはアクセスしたがすぐ別のブラウザタブに移ったので、実際はその WEB ページは見えていないといった履歴は取得するのが難しい[38]。また、WEB サーバ間で提携し、提携した WEB サーバが作成する WEB ページに JAVASCRIPT を記述して、WEB 閲覧履歴を同一の場所に送信するようにして集約することにより、複数の WEB サーバでの閲覧履歴を収集することが行われている[31]。これにより複数の WEB サーバでの閲覧履歴を、ユーザに負担をかけることなく詳細な履歴を収集できる。しかし、複数とはいえ提携した WEB サーバに限られる、ユーザの許可なく閲覧履歴を集約するため、プライバシーの問題がある、などの課題がある。

また、ローカル PC で履歴を取得する手法では、WEB サイトを横断して幅広い WEB 閲覧履歴を収集できる[40]。たとえば、ショッピングサイト A である商品を閲覧し、次にショッピングサイト B で価格を比較し、その後、ユーザ投稿サイトで商品の口コミ情報を調査したといった WEB 閲覧履歴を収集できる。また、WEB サーバにアクセスした履歴だけでなく、WEB サーフィンを中断してメールを読むなど他のアプリケーションに移るまでの時間の長さや、マウスの動きはどうだったか[35]といった詳細な履歴を収集できる。さらに、視線を取得することで、WEB ページの中でどの部分を閲覧していたのかをある程度正確に取得することも可能である[36]。しかし、ローカルの PC に履歴を収集するツールをインストールするなど、利用者に何らかの負担を要求するものが多い。基本的には、詳細な履歴を獲得しようとするほど、ユーザに与える負担は多くなる。例えば、視線認

識を行う場合、ユーザはカメラをつけたヘッドセットを用意する、あるいはユーザの視線を捕らえるためのビデオカメラをモニタのそばに準備し、利用時にはキャリブレーションを行うなど、比較的大きな負担が必要となる。

また、スニッファなどのネットワーク機器により HTTP トラフィックデータを収集する手法[39]や、プロキシサーバを準備して WEB 閲覧履歴を収集する手法[32][33][34][37]は、アクセスするすべての WEB サーバへのアクセス履歴を収集することができる。しかし、プロキシサーバを経由するようにブラウザの設定をユーザに行わせる必要がある、他のアプリケーションを利用した時やマウスの動きなど通信が発生しない履歴は収集することができない、といった課題がある。

一方、WEB 閲覧履歴以外の様々な履歴を収集する手法も広く研究されている。携帯を利用した履歴を収集することで、PC を利用しない者の履歴を収集することが行われている[41]。また、場所に応じた適切なサービスを提供するため、GPS を用いる方法や[42]や無線のアクセスポイントを用いる方法[43][47]、家の中に多数のセンサを配置する方法[45]で、実世界の場所履歴を収集することが行われている。また、バイタルデータを収集し、医療や健康管理に活用しようという研究が行われている[44]。日々の食事のログを、食卓の写真を撮影して画像処理することによりユーザへの負担を抑えて収集する手法が提案されている[46]。ビジネス上で誰と話をしたかという履歴を収集し、職場の円滑なコミュニケーションを支援する研究がおこなわれている[45]。このような PC 上の WEB 閲覧以外の履歴を収集し、必要に応じて WEB 閲覧履歴と連携してサービスを行うことは、ライフログというキーワードで広く研究されており、興味深い。

表 2-1 履歴収集ステップでの分類

		履歴の範囲	履歴の詳細さ	ユーザ負担	サーバDBの利用	プライバシー
ローカル PC で取得	ソフトのみ	○すべてのサイト	○	△	×困難	△個人が管理
	ソフト+ハード	同上	◎	×	同上	同上
WEBサーバで取得		△限定サイト	×通信の発生する履歴のみ	○なし	○容易	×サーバ管理者が管理
NW上で取得(プロキシ等)		○すべてのサイト	×通信の発生する履歴のみ	△プロキシの設定、あるいは機器の設置	×困難	×NW管理者が管理

## 2.3 履歴解析ステップおよび情報提示ステップにおける人間の介在の有無による分類

WEB 閲覧の履歴を解析するステップおよび情報を提示するステップは、人間の介在の有無によって、大きく二つに分類できる（図 2-2）（表 2-2）。

ひとつは、目的とする情報を抽出するために、人間の介在を前提とする手法である。収集した WEB 閲覧履歴の解析において、最終的に目的とする情報を直接抽出するのではなく、閲覧履歴の特徴を抽出し、利用者にそれらを提示する。利用者は提示された閲覧履歴の特徴をもとに、目的とする情報を推測あるいは想起して獲得する。

例えば、WEB 閲覧履歴の元データから、いつその WEB ページを見ていたか、どのような順番で WEB ページを見ていたか、WEB ページを見ていた前後の時間に利用していたファイルは何か、WEB ページのタイトルは何か、あるいは WEB ページを提供していたサーバの URL は何か、などの閲覧履歴の特徴をコンピュータで抽出する。そして、それらをビジュアルにわかりやすく利用者に提示する。利用者は、提示された履歴情報を判断材料にして、WEB 閲覧体験で得た知識などの目的とする情報を、利用者自身で抽出する。

本手法の利点は、閲覧履歴情報の特徴を、人間が効率的に解釈できる程度までコンピュータで解析し、最後は人間の能力によって目的とする情報を抽出するため、人間の能力、特に暗黙知を活用できることである。例えば、履歴データが少量である場合、目的とする情報をコンピュータのみで抽出すると精度が低くなることが予想される。また、『この温泉は、12月であっても、ほぼ毎日雪景色が見られるので良い』といった粒度の細かい知識は、コンピュータのみで抽出することは非常に難しく、現在の研究水準では抽出したとしても精度が低くなると思われる。このような場合、人間の能力を利用することで精度の向上が期待できるであろう。また、閲覧履歴が過去の自分自身の体験の履歴である場合、人間は様々なことを頭の片隅で憶えている。ほぼ忘れてしまっていた体験の履歴であっても、提示されたなにかの情報をきっかけに、目的とする情報を想起することを期待できる。

しかし、単純に履歴ログデータの数字や文字を羅列するだけでは人間は理解できない。閲覧履歴をまず人間に理解させるため、人間が理解できる形式にして提示する必要がある。また、履歴情報を理解できたとしても、その情報が大量過ぎる場合、人間が理解するために多くの時間がかかり、現実的でなくなることが予想される。よって提示する情報量は、コンピュータではなく人間が理解できる量に抑える必要がある。また、人間の能力を利用するため、一人の人間が一人分の嗜好情報を抽出するといったことは難しく、効率的に抽出できる嗜好情報の量には限界がある。

もうひとつは、目的とする情報を抽出するために、人間の介入を必要としない手法である。収集した WEB 履歴の解析において、目的とする情報を直接抽出し、利用者に提示する。目的とする情報そのものを提示するため、利用者は提示された情報から新たに何かを推測あるいは想起する必要はない。

例えば、WEB 閲覧履歴の元データから、利用者の嗜好情報を抽出し、その嗜好情報に関連する新しいニュースや WEB 広告などの目的とする情報を提示する。

本手法の利点は、コンピュータですべての処理を行うため、人間に負荷をかけないことである。例えば、全世界中の人間ひとりひとりの毎日の嗜好情報を抽出しようとした場合、人間が介入して行うのは非現実的であるが、コンピュータのみで行う場合、アルゴリズムやシステムによっては実現可能であると思われる。

しかし、人間の能力、特に暗黙知を利用しにくいいため、人間であれば直感的にわかることであっても、コンピュータで論理的に計算させなくてはならない。履歴が非常に少ない場合、目的とする情報の抽出精度が大きく低下する、といった問題が考えられる。また、前述したように、『この温泉は、12月であっても、ほぼ雪景色が見られるので良い』といった粒度の細かい知識は、コンピュータのみで抽出することは非常に難しく、現在の研究水準では抽出したとしても精度が低くなると思われる。また、閲覧履歴が過去の自分自身の履歴である場合、人間の記憶を活用して、なにかきっかけとなる情報を利用して記憶を想起させるだけで、目的とする情報を抽出できる場合でも、コンピュータでは人間の記憶は考慮せずゼロから計算して抽出しなければならない。

また、閲覧履歴を解析するステップにおいて、一人の履歴のみを解析する手法、複数人の履歴を解析する手法がある。複数人の履歴を利用する場合、特定の人間の履歴と、その他多くの人間の履歴を比較し、特定の人間の特徴を抽出することも可能である。しかし、複数人の詳細な履歴を収集することはプライバシーの問題がある。そこで、本研究では、一人の履歴のみ、あるいは複数人の履歴であっても自分が管理するサイトへの閲覧履歴のみを解析する手法を検討する。

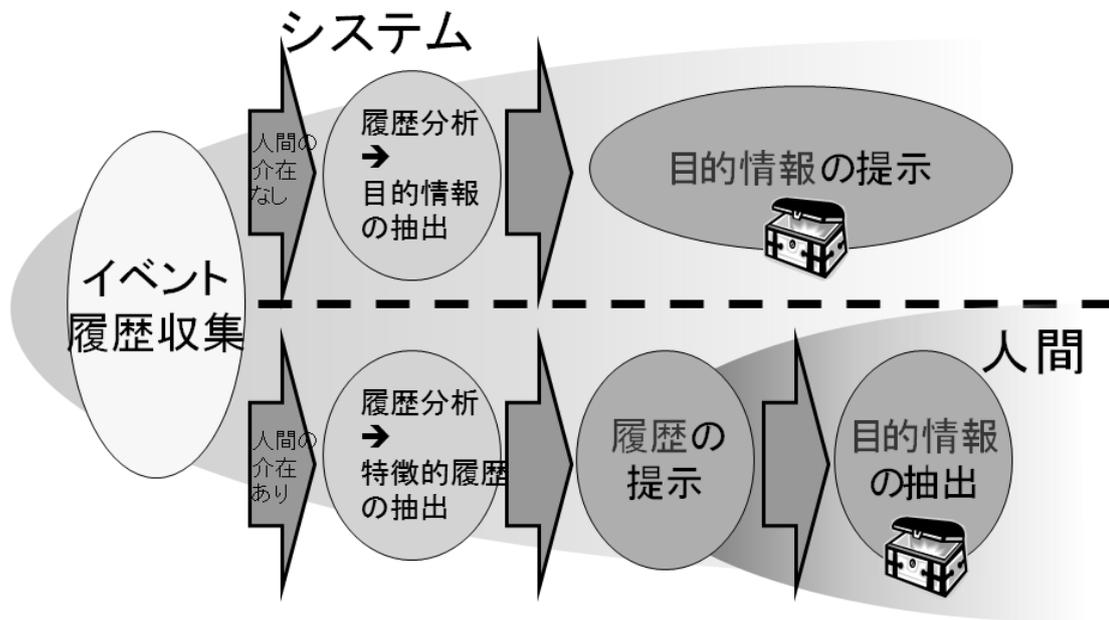


図 2-2 履歴分析，情報提示ステップでの分類

表 2-2 人間を介するか否かによる特徴

	人間への 負荷	抽出でき る情報量	人間の知 識の利用	記憶の活 用	粒度の細 かい情報 の抽出
人間を 介さない 手法	○ 負荷なし	○ 自動的に 大量抽出	△ 無し	△ 記憶はあ まり利用し ない	△ 困難
人間を 介す手法	△ 負荷大	△ 人間の処 理能力が 限界	○ 暗黙知の 活用	○ 過去の記 憶を想起	○ 可能

## 2.4 本研究の位置づけ

本章で整理した内容と、本論文の3章および5章で具体的に提案する二つの手法との対応関係を整理しておく(表 2-3)。ただし、提案手法の詳細については各章にゆだねることにし、対応関係が分かる程度にとどめておく。

本論文は、人間を介在させて、目的とする情報を抽出する手法を検討し評価した内容である。

3章で提案する手法では、ローカルなPC上で、ブラウザのウィンドウ状態などの詳細な履歴を自動収集する。履歴の解析・情報の提示においては、利用者に提示すべき閲覧履歴の特徴を時間に着目して解析し、グラフィカルなインタフェースで詳細なWEB閲覧履歴の特徴を提示する。具体的には、アクティブ期間、集中期間という新たな概念を提唱し、それらの期間の着目度や、キーワードに対する重要度を算出するアルゴリズム、およびそれらを検索、提示するインタフェースを提案する。利用者は、インタフェースに提示された閲覧履歴を見て、目的とする情報を抽出する。ここで、本手法は、粒度の細かい知識を含む自分自身の体験獲得情報(3章にて定義)を抽出することを目的としている。

5章で提案する手法では、自動的に収集されるWEBサーバ上のWEB閲覧履歴を利用する。WEBサーバと連携した商品DBを利用し、利用者が閲覧した商品を精度高く特定する。多くの利用者を対象とするため、WEB閲覧履歴はWEBサーバで収集され、かつ短期間の履歴を対象とする。それらの履歴から時間に着目した特徴や、WEBページに表示した画像内で注目した個所を抽出し、ビジュアルなインタフェースで提示する。利用者は提示された履歴をもとに、第三者の嗜好情報を抽出する。

表 2-3 本研究における WEB 閲覧履歴活用のための各ステップの概要

	履歴収集	履歴解析	情報提示	利用者に出させる情報(目的とする情報)
3章：自分自身による WEB 閲覧履歴検索・提示手法	<ul style="list-style-type: none"> <li>・ PC 上の履歴収集モジュールにより自動収集された行動を含む WEB 閲覧履歴</li> </ul>	利用者に履歴を提示するため主に下記を抽出． <ul style="list-style-type: none"> <li>・ アクティブ期間</li> <li>・ 集中期間</li> </ul>	<ul style="list-style-type: none"> <li>・ 集中期間検索インタフェース</li> <li>・ 詳細行動表示インタフェース等</li> </ul>	自分自身の体験獲得情報
5章：第三者による WEB 閲覧履歴検索・提示手法	<ul style="list-style-type: none"> <li>・ WEB サーバ上で自動収集されたアクセス履歴</li> <li>・ WEB サーバの商品データベース</li> </ul>	利用者に履歴を提示するため，下記を抽出． <ul style="list-style-type: none"> <li>・ アクティブ期間(推測)</li> <li>・ 閲覧した画像内の場所</li> </ul>	<ul style="list-style-type: none"> <li>・ Bubble-graph インタフェース</li> <li>・ Category インタフェース</li> <li>・ Highlight インタフェース</li> </ul>	第三者の嗜好情報

### 3. 自分自身による WEB 閲覧履歴検索・提示手法および試作システムの実装

#### 3.1 体験獲得情報を想起させる WEB 閲覧履歴検索

過去の WEB ページの閲覧体験で得られた情報を想起するために、過去に閲覧した WEB ページの内容を再検索することは多い。そのために、デスクトップ検索技術に代表される、ローカルな PC 上で取得した WEB 閲覧履歴を検索するツールが公開されている [4][5]<sup>1</sup>。デスクトップ検索ツールでは、過去の WEB ページの閲覧体験で得られた情報を思い出すために入力したキーワード(本論文では『再検索キーワード』と呼ぶ)と WEB ページの内容のマッチングを行い、WEB ページのリストを提示する。リスト中の WEB ページをひとつ選択することで、選択した WEB ページの内容を閲覧することができる。

しかし、過去の WEB ページの閲覧体験で得られた情報は、ひとつの WEB ページの内容だけではない。同一体験中にその WEB ページ以外に見た他の WEB ページの内容や、それら WEB ページ間の関係、さらには WEB ページの内容を見て得られた知識等の様々な情報があると考えられる。ひとつの WEB ページの内容だけを発見したとしても、それらの情報を思い出すことは困難である。一方、これらの情報は、過去に一度自分で得た情報であるため、自分の記憶の中になんらかの情報が眠っている可能性が高い。

そこで、過去の WEB ページの閲覧体験で得られた情報を想起させるために、詳細な履歴を自動的に収集し、WEB ページの URL だけでなく、WEB 閲覧にかかわる『行動』(ここで、本研究では、ユーザがコンピュータ上で行った操作を単に『行動』と呼ぶ)を含めた WEB 閲覧履歴を検索・提示する手法を検討する。本手法によってユーザに過去の行動を理解させ、最終的に体験獲得情報を想起させる。

---

<sup>1</sup> 多くのデスクトップ検索では、Web ページだけでなく、ドキュメントファイルやメール文書など多数の形式の文書の検索が可能である。

## 3.2 体験獲得情報の定義

WEB ページの閲覧体験で得られた情報のひとつは、過去に閲覧した WEB ページの記事など、体験に含まれる WEB ページの内容である。また、例えばユーザが何かを調べるとき、一つの WEB ページだけでなく複数の WEB ページの内容を閲覧することは多い。信頼できる特集ページに推薦されていたので商品の詳細ページを閲覧したというように、ユーザが辿った WEB ページ間の関係や、その WEB ページをどのように解釈したのかといった WEB ページから得た知識も、過去の WEB ページの閲覧体験で得られた情報である。体験によって WEB ページの内容とはまったく関係無いことをひらめくといったケースもある。

このように、過去の WEB ページの閲覧体験で得られた情報は、ひとつの WEB ページの内容といった単純なものではなく、様々な情報が含まれている。われわれはコンピュータ上の WEB 閲覧体験によって得られた情報のなかで、本手法の対象とする情報を次のように整理し、体験獲得情報と呼ぶことにする。

### 体験獲得情報：

WEB ページの内容

WEB ページ間の関係

WEB ページから得た知識

ユーザが体験獲得情報を想起しようとする時、ある一つの WEB ページの内容や一つの WEB ページから得た知識をピンポイントで思い出したい場合がある。例えばある商品の重さが知りたい時は、商品のスペックが記載された WEB ページの内容だけを思い出せばよい。逆に、過去に得られた多くの体験獲得情報を思い出したい場合もある。例えば複数メーカーの商品を比較調査した体験で得られた情報を幅広く想起したいときなどである。

本論文では、多くの体験獲得情報の想起が要求される後者の状況を対象とする。しかし、過去の WEB ページの閲覧体験で得られた体験獲得情報を、すべて完璧に思い出すのは、困難であると思われる。また、短期間で想起できた重要な体験獲得情報を元にインターネットの WEB ページを探し直すことで、想起できなかった体験獲得情報を再発見することも考えられる。そこで、効率的に体験獲得情報を想起させることを提案手法の目的とし、次のように定めた。

### 提案手法の目的：

多くの体験獲得情報を想起させること

短期間で体験獲得情報を想起させること

重要な体験獲得情報を想起させること

### 3.3 PC 上での詳細な WEB 閲覧履歴の収集

WEB 閲覧履歴を解析・提示し、体験獲得情報の想起を促すためには、元データとなる詳細な履歴が必要である。そのために、コンピュータ上の詳細な行動を含む WEB 閲覧履歴を自動的に収集し保存する行動履歴自動収集モジュールを開発した(図 3-1) [12][40]。本モジュールでは、オペレーションシステム(OS)のイベントメッセージを監視することにより、アプリケーションに依存せずに行動を含む WEB 閲覧履歴の取得が可能である。具体的には、マウス、キーボード、ウィンドウ状態、コピー、印刷などの履歴を収集する。また、アプリケーション毎に特化した詳細な履歴、たとえば WEB ブラウザが表示した URL やソースファイルなどは、OS のイベントメッセージを監視するだけでは取得できない。そこで、WEB ブラウザおよびメール向けに、行動履歴自動収集モジュールのプラグインを作成し、プラグインによりアプリケーション毎のイベントメッセージを取得できるようにした。具体的には、WEB ブラウザの表示した WEB ページの URL、ソースファイル、サムネイル、HTTP ヘッダ情報、マウスで選択した文字列、およびメールが表示したメール文書等を取得する。行動を含む WEB 閲覧履歴は自動的に取得/保存され、ユーザは行動を含む WEB 閲覧履歴を残すために特別な追加操作を行う必要はない。よって、閲覧した殆んどすべての WEB ページを保存し、必要に応じて閲覧することができる。また、プライバシーの問題に考慮し、WEB 閲覧履歴データを暗号化し、かつローカルのコンピュータ内だけに保存することとする。

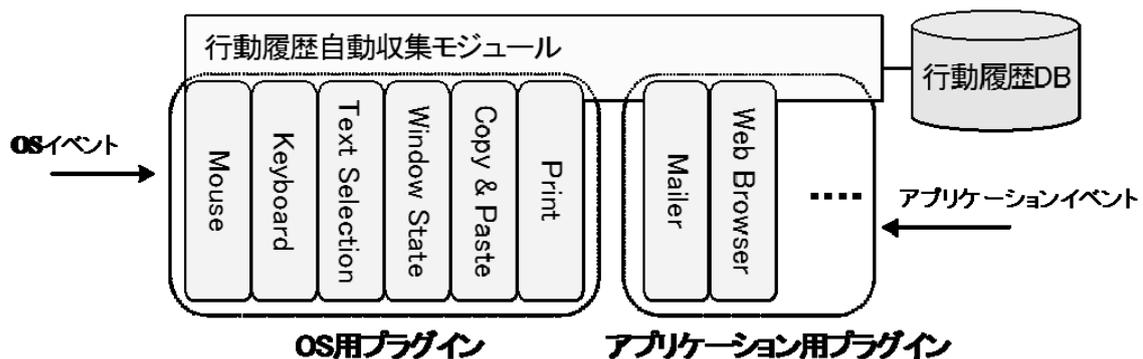


図 3-1 行動履歴自動収集モジュール

### 3.4 期間に着目した WEB 閲覧履歴の解析

体験獲得情報を効率よく想起させるために、時間的に連続した WEB 閲覧履歴をひとつのグループとして抽出し、理解することはよい方法だと思われる。たとえば、デジタルカメラについて調査した体験があったとする。デジタルカメラに関して詳しく調査した期間を一つ発見すれば、その調査期間内の多くのデジタルカメラに関する WEB ページの内容を一度に想起可能である。調査した期間内の WEB ページの閲覧行動を、時系列に詳細に追っていくことにより、WEB ページ間の関係を想起できる。また、複数の WEB ページの内容や WEB ページ間の関係を把握すれば、WEB ページの内容を見たときに、WEB ページから得た知識を想起しやすいと思われる。そこで、WEB ページという物理的なファイルではなく、ユーザが過去に行動を起こした期間に着目する。具体的な期間として、一つのウィンドウまたはブラウザタブがひとつの WEB ページをアクティブに表示している期間(『アクティブ期間』と定義する)、および再検索キーワードに関連する複数の WEB ページを集中して閲覧した期間(『集中期間』と定義する)を抽出し、再検索キーワードに対するそれらの期間の重要度を算出する手法を検討する。WEB ページ単位で閲覧履歴を扱うのではなく、集中期間単位で閲覧履歴を提示することにより、提示する履歴を効率的に絞り込むことができると思われる。

#### 3.4.1 アクティブ期間の抽出、および注目度と重要度の算出

現在普及しているコンピュータの多くは、ウィンドウシステム[6]を採用したオペレーションシステム(OS)を利用している。ユーザは WEB ページをウィンドウ内に表示して閲覧や作業を行っている。アクティブなウィンドウ(あるいはタブ)に表示している WEB ページを変更する、あるいは別のウィンドウ(あるいはタブ)をアクティブすることで、閲覧する WEB ページを切り替えている。たとえば、興味のある WEB ページをリンククリックにより探し、いくつか注目すべき WEB ページが見つかり、WEB ブラウザ(あるいはタブ)を複数起動して両者を切り替えながら WEB ページを比較することなどを行っている。つまり、WEB サーバにアクセスしたからといって、その WEB ページをしっかりと見ているとは限らない。

そこで、一つのウィンドウ(あるいはタブ)がひとつの WEB ページをアクティブに表示している期間を、『アクティブ期間』と定義し、アクティブ期間をユーザの行動をまとめる特徴的な単位であると考えた。よって例えば、WEB ブラウザ(あるいはタブ)に表示している WEB ページを変更するたびに、あるいはブラウザのアクティブ/非アクティブを切り替えるたびに、新しいアクティブ期間が発生する(図 3-1)(図 3-2)。

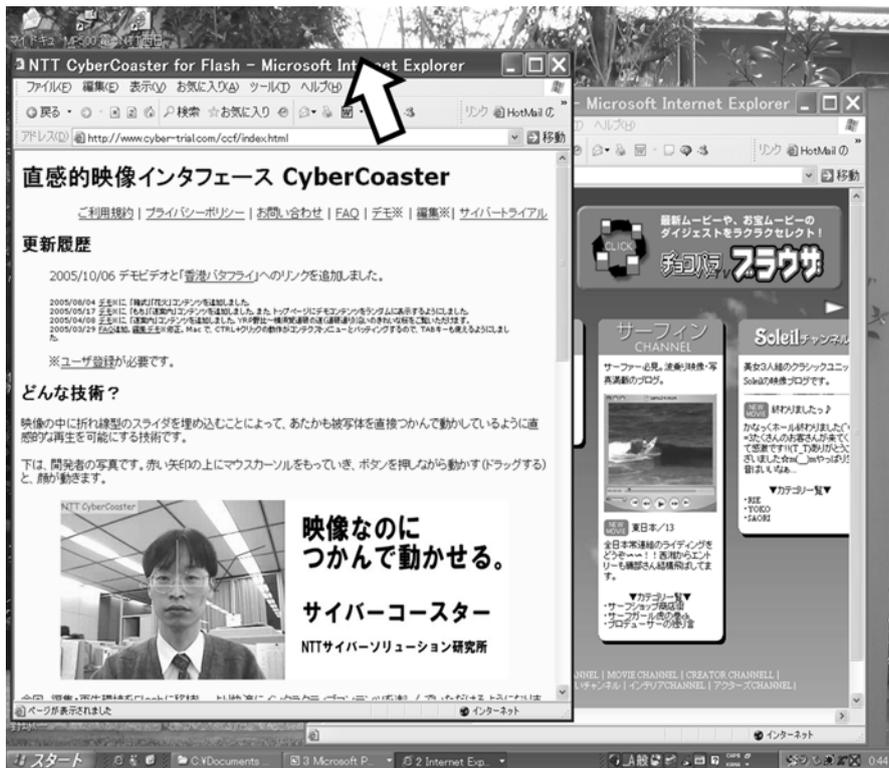


図 3-2 アクティブ期間の開始

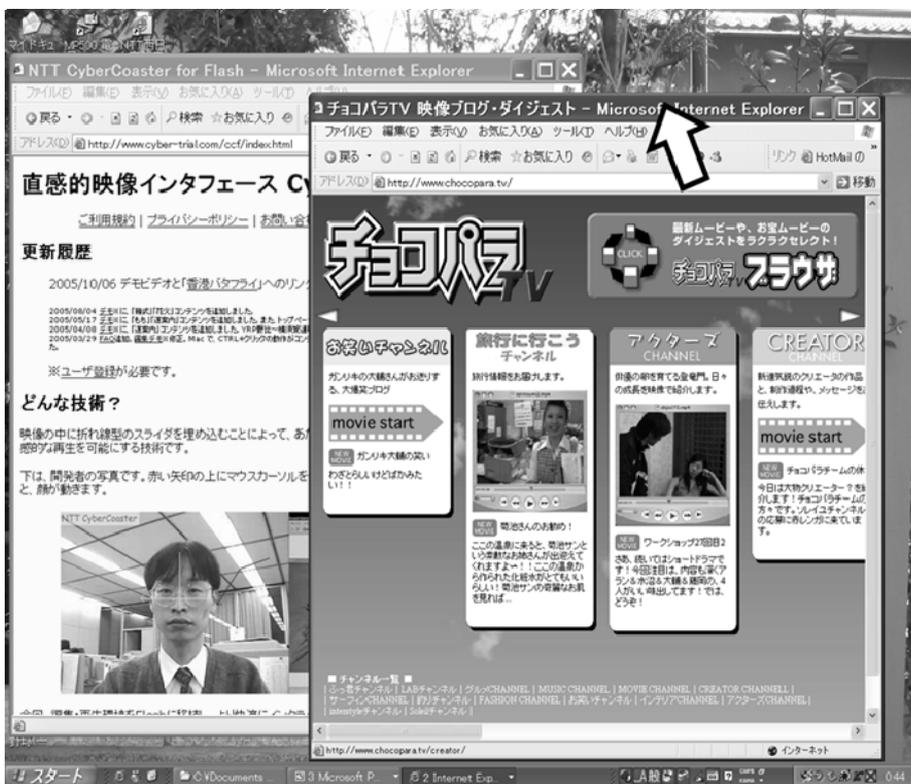


図 3-3 直前のアクティブ期間の終了と新しいアクティブ期間の開始

ユーザはそれぞれのアクティブ期間内に、WEB ページを閲覧し、注目した文章をコピーし、特に有用であると判断すれば印刷するなど、さまざまな行動を起こす。そこで、ユーザが多く行動を起こしていたアクティブ期間はユーザが注目した可能性が高いと仮定し、アクティブ期間に対するユーザの注目の強さをあらわす指標としてアクティブ期間の注目度を定義する。

アクティブ期間の注目度は、アクティブ期間内にユーザが行ったそれぞれの行動に対し、印刷などユーザが意識的に起こした行動ほど重み付されたポイントを与え、アクティブ期間内のすべての行動のポイントを積算することで求める((1)式) (図 3-4)。更に、任意の再検索キーワードに対するそれぞれのアクティブ期間の重要度を、アクティブ期間の注目度と、アクティブ期間中に表示されている WEB ページの再検索キーワードに対する適合度を乗ずることで、式(2)のように求める(図 3-5)。

$$Att(ap) = \sum_i (E_i \times Fr_i) \quad (1)$$

$$IAP(k, ap) = Att(ap) \times R(k, ap) \quad (2)$$

$$R(k, ap) = tf(k, ap) \times \log(n/n_k) \quad (3)$$

$Att(ap)$  : アクティブ期間  $ap$  の注目度

$IAP(k, ap)$  : 再検索キーワード  $k$  に対するアクティブ期間  $ap$  の重要度

$R(k, ap)$  : アクティブ期間  $ap$  に表示されている WEB ページの再検索キーワード  $k$  に対する適合度

$E_i$  : 行動種類  $i$  の重み付け係数

$Fr_i$  : アクティブ期間  $ap$  内における行動種類  $i$  の発生回数

$ap$  : アクティブ期間

$k$  : 再検索キーワード

$i$  : 行動種類

$tf(k, ap)$  : アクティブ期間  $ap$  に表示されている WEB ページ中での再検索キーワード  $k$  の出現頻度

$n$  : WEB ページ集合中の WEB ページの総数

$n_k$  : WEB ページ集合中の再検索キーワード  $k$  を含む WEB ページの総数

ここで、アクティブ期間  $ap$  に表示されている WEB ページの再検索キーワード  $k$  に対する適合度  $R(k, ap)$  は、行動履歴自動収集モジュールが保管したすべての WEB ページを WEB ページ集合とした TF-IDF 値[7]を用いる((3)式)。 $tf$  値は、HTML タグによる重み付

けを行う（表 3-1）。また，システムの実装においては，TF 値および IDF 値を求める計算に時間がかかるため，上記 WEB ページ集合は一日毎など定期的，あるいはユーザ指定時に最新状態に更新する．行動種類としては，WEB ページ閲覧時間，コピー，印刷，マウス操作，キーボード入力，文字列選択を考慮する．

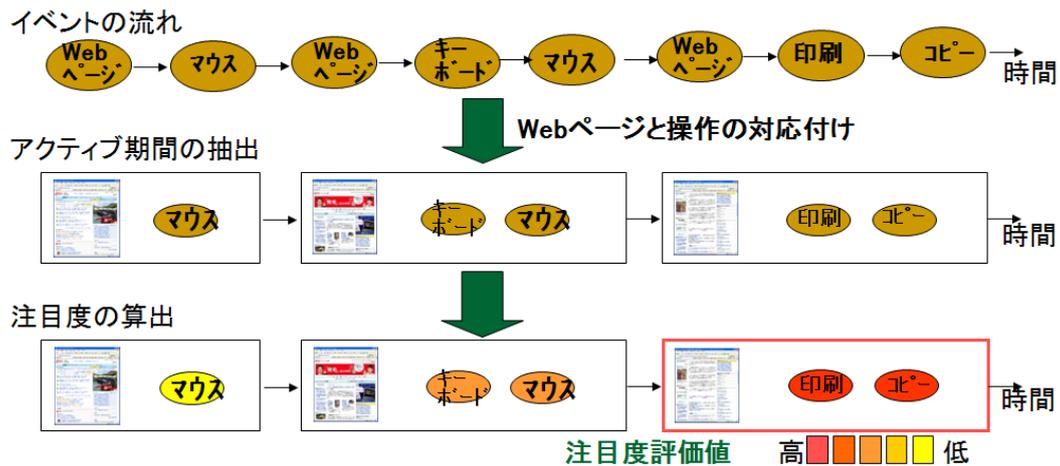


図 3-4 アクティブ期間の抽出と注目度の概念図

### <注目度から再検索キーワードに対する重要度を算出>

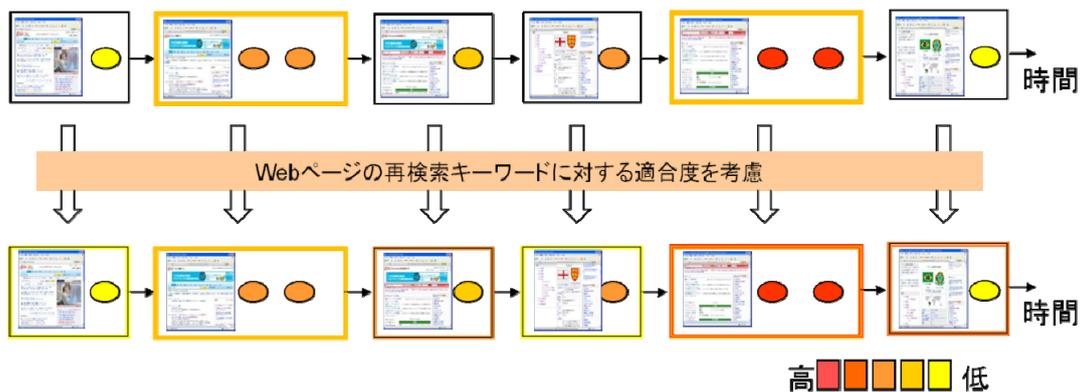


図 3-5 再検索キーワードに対する適合度の概念図

### 3.4.2 集中期間の抽出と重要度の算出

ある再検索キーワードに関連する複数の WEB ページを集中して閲覧した期間には，多

くの重要な体験獲得情報が含まれると思われる。よって、そのような期間（集中期間）を抽出して、集中期間内の詳細な WEB 閲覧履歴を提示することで、効率よく体験獲得情報を想起させることができる。集中期間に閲覧した多くの WEB ページには再検索キーワードそのものが含まれていると考えられる。しかし、再検索キーワードを含まないが再検索キーワードに関連している WEB ページも存在する。また短期的に再検索キーワードに関連せず再検索キーワードを含まない WEB ページを訪れ、その後、再検索キーワードに関連する WEB ページの閲覧に戻る行動も考えられる。再検索キーワードを含む WEB ページを見たが、その WEB ページには注目しなかったといった行動も考えられる。

そこで、次のように集中期間を抽出する。まず、再検索キーワード  $k$  に対する任意の時間  $t$  の重要度を、再検索キーワード  $k$  に対するアクティブ期間  $ap$  の重要度をアクティブ期間  $ap$  の時間幅で割った値とする ((4)式)。次に再検索キーワード  $k$  に対する時刻  $t$  の平均重要度を、再検索キーワード  $k$  に対する  $[t-a, t+a]$  における重要度の平均であるとする。そして、平均重要度がある閾値  $b$  以下であれば、再検索キーワード  $k$  に関連する行動が中断されたとみなす(図 3-6)。言い換えると、継続判定関数  $B(k, t)$  が 1 以上である連続した期間を、ひとつの集中期間として抽出する((5)式)。一定時間幅  $a$ 、閾値  $b$  は、どのように集中期間を抽出するかを調整するパラメータである。一定時間幅  $a$  により、再検索キーワード  $k$  を含まない WEB ページを短期的に閲覧していたとしても、前後の時間帯の行動から一つの集中期間であると判断することができる。閾値  $b$  により、再検索キーワード  $k$  に対する重要度がある程度高い時刻でも、再検索キーワードに関連する行動が中断されたとみなすことができる。例えば  $(a, b) = (300, 0)$  とすれば、10 分以上再検索キーワードを含むページを閲覧しなかった場合に、再検索キーワード  $k$  に関連する行動が中断されたとみなすことになる。ユーザが脇道に逸れた WEB ページをある程度閲覧していても、一つの集中期間として抽出したい状況において有効である。 $(a, b) = (0, 5)$  の場合、ユーザが特に集中して再検索キーワード  $k$  に関連する行動を起こしていた期間を一つの集中期間として抽出したい状況などにおいて有効である。これらのパラメータは、集中期間の抽出を行うアプリケーションが、集中期間抽出時のユーザの要求に応じて変更できる。

$$I(k, t) = IAP(k, ap) / (apet - apst) \quad (4)$$

$$AI(k, t) = \int_{t-a}^{t+a} I(k, t) / 2a dt$$

$$B(k, t) = \begin{cases} 0, & \text{if } AI(k, t) \leq b \\ 1, & \text{otherwise,} \end{cases} \quad (5)$$

$apst$  : アクティブ期間  $ap$  の開始時刻

$apet$  : アクティブ期間  $ap$  の終了時刻

$I(k,t)$  : 再検索キーワード  $k$  に対する時刻  $t$  の重要度

$AI(k,t)$  : 再検索キーワード  $k$  に対する時刻  $t$  の平均重要度

次に、抽出したそれぞれの集中期間  $ip$  について、再検索キーワード  $k$  に対する重要度を (6) 式のように算出する。

$$IIP(k,ip) = \int_{ipst}^{ipet} I(k,t) dt \quad (6)$$

$IIP(k,ip)$  : 再検索キーワード  $k$  に対する集中期間  $ip$  の重要度

$ipst$  : 集中期間  $ip$  の開始時刻

$ipet$  : 集中期間  $ip$  の終了時刻

$ip$  : 集中期間

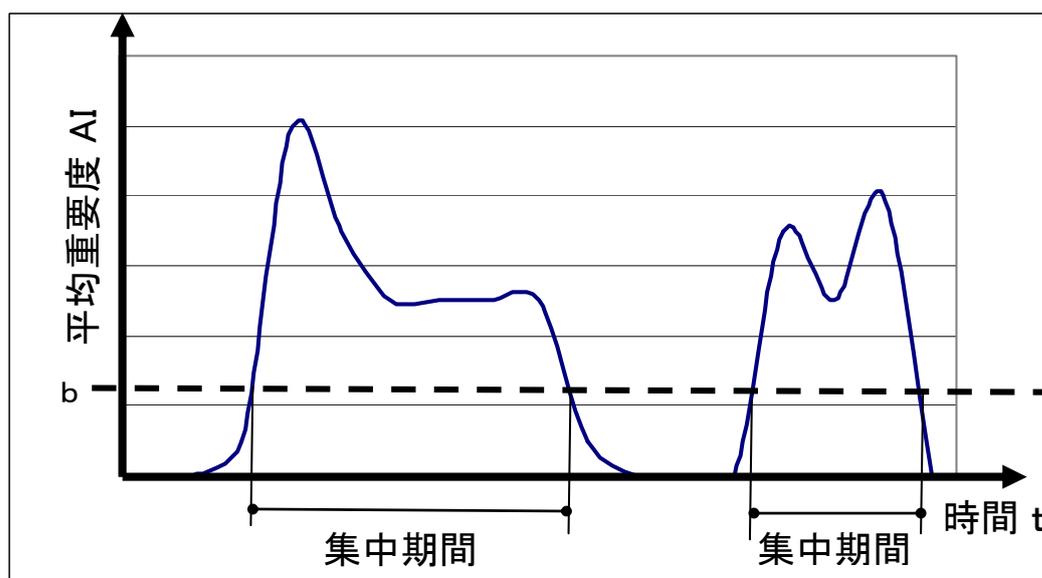


図 3-6 集中期間の抽出

表 3-1 HTML タグによる重み

HTMLタグ	重み
<TITLE>	16
<H1>	8
<H2>	7
<H3>	6
<H4>	5
<H5>	4
<H6>	3
<A>	4
<STRONG>, <EM>, <CODE>, <KBD>, <SAMP>, <CITE>, <VAR>, <ABBR>, <ACRONYM>, <DFN>	2

### 3.4.3 WEB ページの重要度の抽出

指定した期間における再検索キーワード  $k$  に対する WEB ページ  $url$  の重要度は ,再検索キーワード  $k$  に対する指定された期間内に WEB ページ  $url$  を表示していたアクティブ期間  $ap_j$  の重要度の和であるとする ((7)式) .

よって再検索キーワードに高く適合し , かつ , 注目度の高いアクティブ期間に多く表示されていた WEB ページが , 高い重要度を持つ . この指標は後述する WEB ページ検索インタフェースで利用される .

$$IW(k, url, st, et) = \sum_j IAP(k, ap_j) \quad (7)$$

$IW(k, url, st, et)$  : 期間  $[st, et]$  における再検索キーワード  $k$  に対する WEB ページ  $url$  の重要度

$ap_j$  : 期間  $[st, et]$  において , WEB ページ  $url$  を表示していたアクティブ期間

$url$  : WEB ページ

### 3.5 集中期間検索インターフェース

WEB 閲覧履歴を解析した結果は、複数のインターフェースによってユーザに提示される。図 3-7 に集中期間を検索するインターフェースを示す。ユーザはまずこのインターフェースを利用することを想定している。

ユーザが画面左上の入力窓に再検索キーワードを入力すると、再検索キーワードに対する集中期間の重要度の高い順、あるいは開始時間順に、集中期間を表示する。それぞれの集中期間に対して、集中期間内に行われた行動の概要を提示し、どのような行動をした集中期間かを理解できるようにする。具体的には、集中期間の開始時刻と終了時刻(図 3-7(a))を表示する。集中期間の開始時刻と終了時刻に張られているリンクをクリックすると、クリックした集中期間内の行動を行動シーケンス表示インターフェース(図 3-8, 後述)で表示する。また、再検索キーワードに対する集中期間の重要度(図 3-7(b))、集中期間内にインターネット上の検索エンジンに入力されたキーワード(検索キーワードと呼ぶ)(図 3-7(c))、および集中期間内に閲覧した WEB ページに出現する頻出名詞と未知語(図 3-7(d))を表示する。また、集中期間における再検索キーワードに対する WEB ページの重要度の高い上位 5 つのサムネイルとタイトルを表示する(図 3-7(e))。

本インターフェースを利用することで、ユーザは再検索キーワードに対する重要度の高い集中期間を容易に発見できる。上位にランキングされた再検索キーワードに対する重要度の高い集中期間には、再検索キーワードに関連する多くのアクティブ期間が含まれている。よって、次節で説明する行動シーケンス表示インターフェースを用いて上位の集中期間内の詳細な行動を提示 / 理解することで、効率的に体験獲得情報を想起できると思われる。

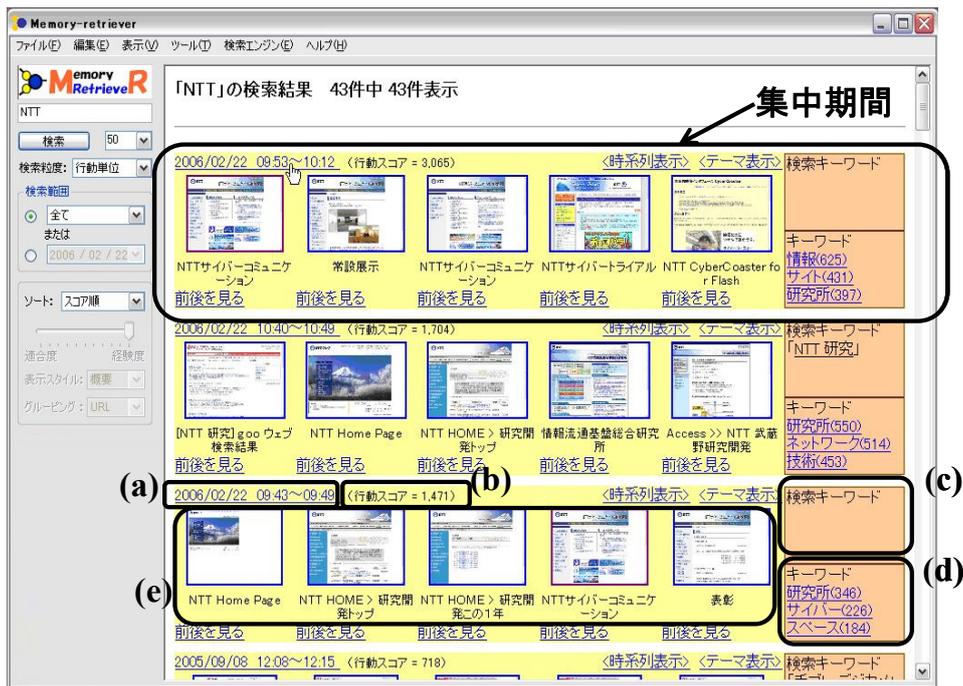


図 3-7 集中期間検索インターフェース

### 3.6 行動シーケンス表示インタフェース

集中期間検索インタフェースで表示された集中期間の一覧の中から、より詳細な WEB 閲覧履歴を見たい期間を選択すると、集中期間内の詳細な行動が、行動シーケンス表示インタフェースを用いて提示される。図 3-8 に行動シーケンス表示インタフェースの例を示す。

行動シーケンス表示インタフェースでは、アクティブ期間の遷移を WEB ページのサムネイルを用いて時系列に表示する。サムネイル上にマウスポインタを重ねると、拡大されたサムネイルを表示する(図 3-8(a))。サムネイルを右クリックすると更に画面いっぱいに大きくサムネイルを表示する。これにより、ユーザは複数の WEB ページの内容を発見できる。

WEB ページを閲覧する場合、インターネット上の検索エンジンに検索キーワードを入力し、該当する WEB ページを見つけ出すことは多い。また、検索キーワード、WEB ページに含まれるタイトル、およびアンカーテキストは情報検索において有益な手がかりであることが判明している[11]。本インタフェースでは、個々のサムネイルの下に WEB ページのタイトルを表示する(図 3-8(b))。サムネイルの上には、検索エンジンに入力した検索キーワード(図 3-8(c))が、表示されている WEB ページに遷移するために直前の WEB ページ上でクリックしたアンカーテキスト(図 3-8(d))を表示する。これにより、WEB ページの内容の理解を高めるだけでなく、ユーザにとっての WEB ページ間の関係を想起できると思われる。たとえば、検索エンジンに『デジタルカメラ 薄型』と入力して薄型のデジタルカメラを探し始め、ニュースサイトでお勧めのカメラを調べた後、メーカーの商品ページで詳細にスペックを調べたといった過去の行動を理解し、この WEB ページはこのレビューページで推薦されているといった WEB ページ間の関係を想起できると思われる。

また、表示されているアクティブ期間のうち、再検索キーワードに対するアクティブ期間の重要度の高い上位 5 つを赤色、次の 5 つを橙色の枠によって強調表示する。逆に、再検索キーワードに対する重要度の低いアクティブ期間をフィルタリング操作バーを調整することで表示しないようにするフィルタリング機能を備える(図 3-8(e))。注目度の高い、あるいは低いアクティブ期間を素早く理解させることにより、効率的に体験獲得情報を想起させることができる。

サムネイルの右下端には、URL 毎の番号を提示し、同じ URL の WEB ページであるかどうか把握できる。重複削除チェックボックスをチェックすると、同じ URL を表示するアクティブ期間が複数あった場合、初回以外のアクティブ期間を表示しない。また、ウィンドウの新規起動や切替えをサムネイル間の矢印を用いて提示する。具体的には、同じ

ウィンドウで遷移したアクティブ期間は矢印線でサムネイルを結び、新しいウィンドウを立ち上げた場合は<NewWindow>と書かれた矢印線でサムネイルを結び(図 3-8(f)), ウィンドウを切り換えた場合は、サムネイル間に矢印線を表示しない。

また、それぞれの機能は、チェックボックスで表示/非表示を選択できるようにした。

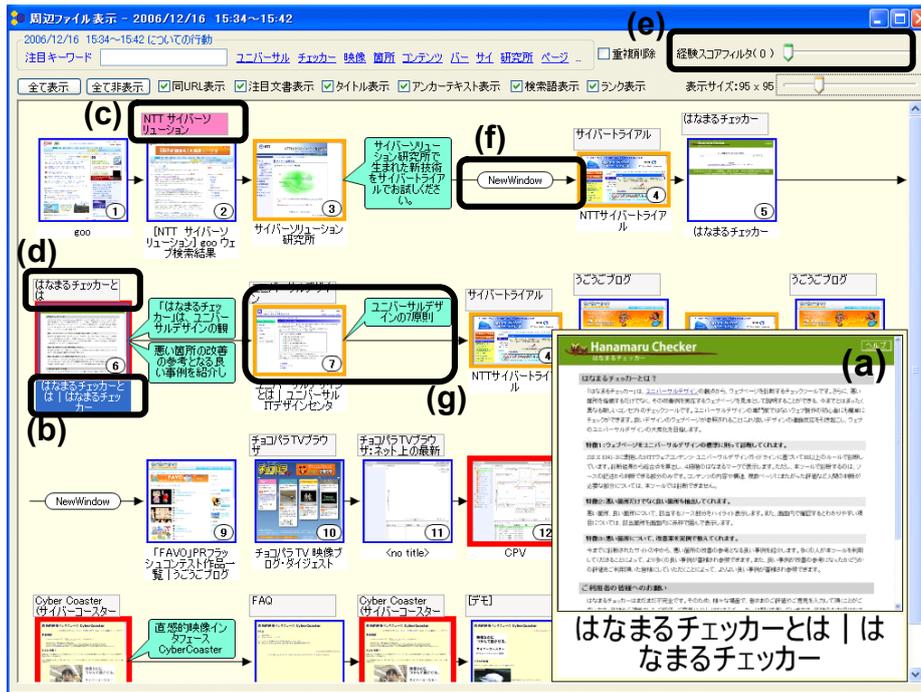


図 3-8 行動シーケンス表示インターフェース

ここで、本を読んでいるときに重要な箇所を鉛筆などで線を引く人は多い。同じような操作を WEB ページ上で実現する技術を適用した[8]。サムネイルの横に表示された吹き出し(図 3-8(g))の形で、選択したテキスト情報を表示する。また、吹き出しのあるサムネイルをクリックすると、選択した部分の背景を黄色で表示する(図 3-9)。この注目箇所表示機能を用いて WEB ページ内の重要と考えた文章を理解することにより、効率的に WEB ページから得た知識を想起できると思われる。



図 3-9 注目箇所表示

### 3.7 WEB ページ検索インタフェース

体験獲得情報を想起するために、集中期間やアクティブ期間ではなく WEB ページを直接検索したい場合がある。その場合に利用するインタフェースが、WEB ページ検索インタフェースである。図 3-10 に WEB ページ検索インタフェースを示す。

ユーザが再検索キーワードを入力すると、行動を含む WEB 閲覧履歴を蓄積してきたすべての期間における再検索キーワードに対する WEB ページの重要度の高い順、あるいは最も新しく閲覧した時間順に WEB ページを表示する。よって再検索キーワードに高く適合し、かつ、注目度の高いアクティブ期間に多く表示されていた WEB ページが上位に示される。それぞれの WEB ページには、サムネイル、タイトル、スニペット、および URL を表示する。更に、閲覧回数、閲覧時間、コピーや印刷したかどうかなどの行動を含む WEB 閲覧履歴を表示する(図 3-10(a))。また、検索された WEB ページを閲覧していた集中期間を表示するリンクが検索結果に付加される。そのリンク(図 3-10(b))をクリックすると、WEB ページを閲覧した時刻を含む集中期間を、行動シーケンス表示インタフェースで表示する。

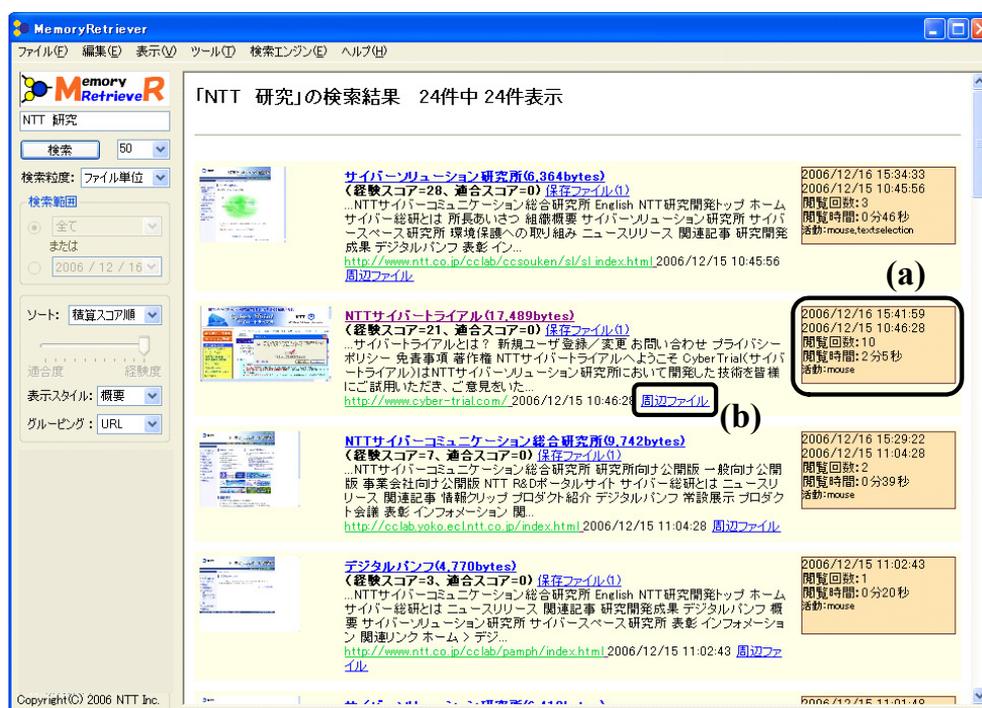


図 3-10 WEB ページ検索インタフェース

### 3.8 自分自身による WEB 閲覧履歴検索・提示手法の試作システムの実装

3.3 章で述べた PC 上での WEB 閲覧履歴の自動収集モジュールを実装し、3.4 章で述べた WEB 閲覧履歴解析を行い、3.5, 3.6 および 3.7 章で述べたインターフェースを用いて、WEB 閲覧履歴を検索・提示する試作システムを実装した。試作システムは、履歴自動収集ツールと、履歴検索ツールの二つのプログラムで構成される。

履歴自動収集ツールは、マイクロソフト WINDOWS XP および VISTA 上で動作する単独アプリケーションとして実装した。タスクトレイに常駐することで、アプリケーションに依存しない履歴を収集する。具体的には、キーボードの押下、マウスボタンのクリック、マウスホイールの回転操作、ウィンドウのアクティブ操作、非アクティブ操作、コピー操作、および印刷操作を収集して行動履歴 DB に記録する。また、WEB 閲覧履歴の詳細なデータを取得するため、インターネットエクスプローラ Ver.6 および Ver.7 用のプラグインを作成し、ブラウザに組み込んだ。このプラグインにより、インターネットエクスプローラのタブの切換え操作、ブラウザのサムネイル、閲覧した URL、ソースファイル、HTTP ヘッダ情報、マウスで選択した文字列、およびマウスで文字選択したときのサムネイルを収集し、行動履歴 DB あるいはローカル PC のハードディスクに保存する。また、プライバシーの観点から、WEB 閲覧履歴を保存したくない時間や WEB ページがあると思われる。よって、WEB 閲覧履歴すべての記録を休止する機能、および、個々の WEB 閲覧履歴の種別(マウス、キーボード等)ごとに記録を休止する機能を持たせる。ホワイトリスト、ブラックリストの URL フィルタ機能を実装し、URL フィルタに従って閲覧履歴を記録できるようにする。また、履歴自動収集ツールは、一定間隔で自動的に、アクティブ期間を抽出してアクティブ期間の着目度を計算する機能を持たせた。さらに、他のユーザが WEB 閲覧履歴を見ることができないようにするため、WINDOWS のログインパスワード以外に、履歴自動収集ツール独自のパスワードを設定できるようにした。

履歴検索ツールは、独立したひとつのアプリケーションとして実装した。履歴検索ツールを起動すると、集中期間検索インターフェース(図 3-7)が表示される。集中期間検索インターフェースの再検索キーワード入力ボックスに再検索キーワードを入力すると、集中期間を検索できる。集中期間は、再検索キーワードに対する集中期間の重要度順、あるいは時間順にソートできるようにする。また、検索ツールに選択ボックスを設け、履歴検索を集中期間単位で行って集中期間検索インターフェースで提示するか、WEB ページ単位で行って WEB ページ検索インターフェース(図 3-10)で提示するかを選択できるようにする。そして、集中期間検索インターフェースで表示された集中期間を選択するか、WEB ページ検索インターフェースで表示される『周辺ファイル』ボタンをクリックすることで、行動詳細表示インターフェース(図 3-8)を用いて集中期間内のユーザの行動を表示するようにする。

また、履歴検索ツールは、再検索キーワードと検索結果の入出力だけでなく、定期的にアクティブ期間のシーケンスを解析することでクリックしたアンカテキストの抽出を行う。また、goo[22]やYahoo![23]などのメジャーな検索エンジンの検索結果のURLから検索語を抽出する。さらに、WEBページのソースファイルを形態素解析し、頻出語句を抽出する。ここで、形態素解析には茶筌[24]、全文キーワード検索にはLucene[25]、WEBページのサムネイル作成にはImageMagick[26]、暗号ライブラリにはCamellia[27]、IEプラグイン作成では、BandObject[28]を利用した。これらの優れたライブラリを提供していただいた開発者の方に感謝したい。

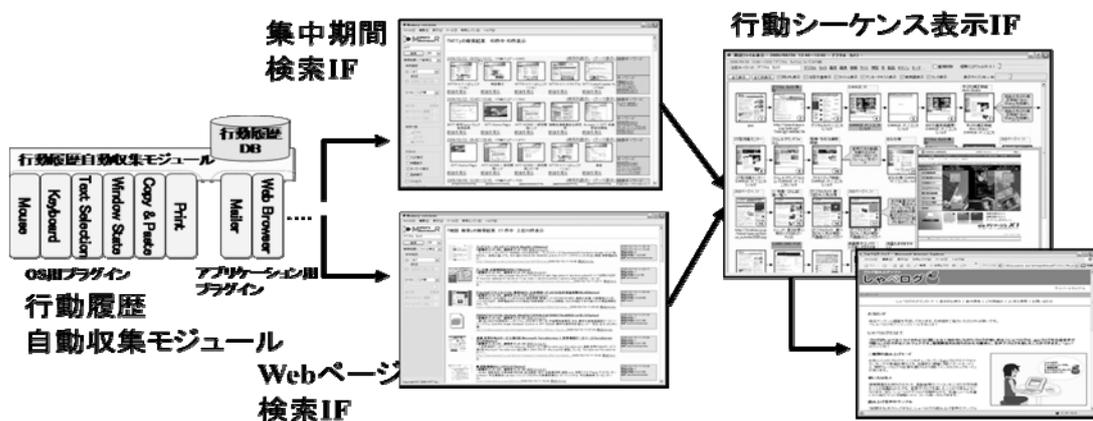


図 3-11 試作した WEB 閲覧履歴検索・提示システムの流れ

### 3.9 自分自身による WEB 閲覧履歴検索・提示手法に関する関連研究

コンピュータ上でのユーザの行動の履歴を記録し、活用する研究は広く行われている。

Chirita ら[2]は閲覧した WEB ページや e-mail に、クリック先の WEB ページの URL やメールの送信先などのメタデータを自動付加することで、コンテキスト情報を用いたデスクトップ検索を提案している。MyLifeBits[3]では、人間の行動にかかわるすべてのものを記録するという方針のもと、WEB ページだけでなく、メール、紙、書籍、写真、音声、CD/DVDなどを記録する。メールは一日 100 通、WEB ページ一日 100 ページ、紙は一日 5 枚、書籍は 10 日で一冊、写真は一日 10 枚、音声は一日 8 時間、CD は 10 日で一枚と見積もって、5 年で 80G バイト程度だと述べている。これらを検索するため、メタデータをファイルに付加する簡単な手法等を多数提供している。これらの技術や、前章で述べたデスクトップ検索ツールは、主に再検索キーワードに関連するユーザが閲覧した WEB ページの内容を効率よく探し出すことに主眼を置いており、ユーザが過去の WEB ページの閲覧体験で得た様々な情報を想起させることに主眼を置いている本提案とは異なる。

ユーザの行動の時間に着目している手法では、単純に WEB ページのダウンロード時間順に表示するツールが WEB ブラウザの履歴機能やデスクトップ検索ツールのタイムライン表示機能として提供されている[4]。SmartBack[13]では、WEB ページ閲覧の軌跡、検索/ブックマークなどの特徴的操作等に基づいてキーとなる WEB ページを抽出し、WEB ブラウジングの戻る操作を補完する機能を提案している。Retrospector[14]は、コンピュータ画面上のほぼすべてのテキストを自動的に記録し、また、キーボード、マウス、クリップボード、かな漢字変換、および手動で加えたメモ書き等を収集する。日付の指定や、キーワードの入力によって表示する画面キャプチャを絞り込み、時系列に表示/再生する。また、3 次元視覚化機能と連携し、画面キャプチャを放物線上に連続配置してオンマウスで浮き上がらせて紙幣の束をめくるような操作で閲覧できるようにすることで、視覚性を向上している。TimeMachineComputing[15]は、個々のファイルを時間順に配列するのではなく、デスクトップ環境やコンピュータ環境の状態そのものを時間順に管理する方式を提案している。単に過去の情報を検索するだけではなく、過去の作業環境をできるだけ完全に復活することを目的としている。作業環境を保存するため、利用者は、目的のファイルを見つけるだけでなく、そのとき同時に画面に登場する関連する情報を抽出することができる。基本的には、移動すべき日時(時刻)を指定するが、その指定には様々な方法を提案している。日付を直接指定する方法のほかに、ファイルを作成した時といったユーザ操作を指定することができる。履歴の提示方法としては、半透明表示、タイムラインやカレンダービューを提案している。俺デスク[16]では、ユーザの行動を含む WEB 閲覧履歴から WEB

ページやファイル毎の着目度を算出し，着目度を縦軸，時間を横軸にして過去に閲覧した WEB ページ等のサムネイルを表示する手法を提案している．表示するサムネイルを指定する方法として，日付を入力するインタフェースを提案している．また，あるファイルと同時に利用していたファイルを検索するインタフェースを提案している．増井ら[17]は，さらに進んで，ファイルの更新日時，ディレクトリ構造，内容等から指定したファイルの近傍だと判断したファイルを自動提示する手法を提案している．Memory-organizer[18]は，ユーザがクリックしたアンカテキスト内の単語の共起関係を視覚化し，指定したある時刻でフィルタリングして表示する．山田ら[19]は WEB 閲覧履歴をクラスタリング手法により要約し，時系列に 3D 表示することで大域的なユーザの興味の遷移を表示する手法を提案している．これらは時間に着目し，ユーザの閲覧履歴を活用する機能を備えている．しかし，本研究では再検索キーワードに関連した行動を起こしていた期間を抽出／提示するという点で，従来の手法と異なる．

また，WEB 閲覧履歴を用いてニュースサイトの記事をカスタマイズして提示するシステム[20][21]が提案されている．これらのように，WEB 閲覧履歴を用いて未知の WEB ページの検索を支援する研究も広く行われているが，本手法は WEB 閲覧履歴自体を検索するという点でこれらと異なる．

## 4. 自分自身による WEB 閲覧履歴検索・提示手法の評価

3章で提案した WEB 閲覧履歴検索・提示手法が，ユーザに体験獲得情報を想起させることが可能であるかを評価する．評価にあたっては，重要な体験獲得情報が短時間で多く想起できているか分析する．

### 4.1 評価対象とするインタフェース

評価実験では，行動シーケンス表示インタフェースが，体験獲得情報を想起させているかを定量的に評価する．行動シーケンス表示インタフェースは，体験獲得情報を想起させるインタフェースであると共に，アクティブ期間の妥当性も同時に評価できると考えられるからである．

想起させる体験獲得情報としては，WEB ページから得た知識を選択する．これは，WEB ページの閲覧によってユーザが最終的に得る体験獲得情報であり，重要度が高いと考えられるからである．たとえば，石垣島について調査した体験があった場合，

『www.ishigaki.tmp.co.jp/ishigaki.html』という URL ではなく，『石垣島でサーフィンは困難である』という知識を思い出すことに相当する．

また，行動シーケンス表示インタフェースの有効性について，定性的に分析する．

### 4.2 比較対象と実験条件

#### (1)比較対象

本実験の比較対象としては，WEB ページの閲覧履歴の閲覧時間の特徴ををテキストで時系列に表示する手法(従来手法)があげられる[4]．具体的には，閲覧した WEB ページをダウンロードした時間順に表示する．それぞれの WEB ページの情報として，ダウンロードした時刻，アイコン，タイトル，URL，および PC 内に保存された WEB ページへのリンクを表示する(図 4-1)．

ウェブ イメージ ニュース マップ グループ デスクトップ [more](#) »

検索  デスクトップ設定  
検索オプション

タイムライン [すべて](#) - [メール](#) - [ファイル](#) - [ウェブの履歴](#) - [チャット](#) イベントを削除

[前のアイテム](#)

2006年9月19日

- 19:57:32 Privacy Policy - www.cyber-trial.com/privacy.html - 1個のキャッシュが見つかりました
- 19:57:29 よくある質問 - www.cyber-trial.com/webcomp/faq.html - 1個のキャッシュが見つかりました
- 19:57:26 WebCompツールバー機能詳細 - www.cyber-trial.com/webcomp/function.html - 1個のキャッシュが見つかりました
- 19:57:10 WebComp Toolbarトップ - www.cyber-trial.com/webcomp/index.html - 1個のキャッシュが見つかりました
- 19:57:08 NTTサイバートライアル - www.cyber-trial.com/index.html - 2個のキャッシュが見つかりました
- 19:57:03 シェバログ:お問い合わせ - www.cyber-trial.com/shabelog/contact.html - 1個のキャッシュが見つかりました
- 19:57:01 シェバログ:よくある質問 - www.cyber-trial.com/shabelog/FAQ.html - 2個のキャッシュが見つかりました
- 19:56:37 シェバログ:ご利用規約 - www.cyber-trial.com/shabelog/caution.html - 2個のキャッシュが見つかりました
- 19:56:35 シェバログ:動作環境 - www.cyber-trial.com/shabelog/config.html - 2個のキャッシュが見つかりました
- 19:56:33 シェバログ:基本的な操作 - www.cyber-trial.com/shabelog/how2use.html - 2個のキャッシュが見つかりました
- 19:56:31 シェバログ:アンインストール - www.cyber-trial.com/shabelog/uninstall.html - 1個のキャッシュが見つかりました
- 19:56:28 シェバログ:トップ - www.cyber-trial.com/shabelog/index.html - 2個のキャッシュが見つかりました
- 19:55:40 Liability - www.cyber-trial.com/liability.html - 1個のキャッシュが見つかりました
- 19:55:34 お問い合わせ - 電車内での手話による緊急情報提供実験 - www.super-space.jp/contact.html - 1個のキャッシュが見つかりました

タイムラインを表示  
2004年8月 から 2006年9月

8月 2006年9月 10月 >

日	月	火	水	木	金	土
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
今日: 9月19日						

次の期間のアイテムに移動  
2006年 1月 実行

図 4-1 従来手法の画面イメージ

## (2)被験者

インターネットでの WEB 閲覧が問題なく行えるよう、コンピュータ初心者ではないユーザを対象とする。被験者の属性を表 4-1 に示す。

表 4-1 被験者の属性

被験者数	20人
年齢	20代～40代
男女比	1:1
コンピュータ利用頻度	週1回～ほぼ毎日
ユーザインタフェースの習熟度	各手法に対し、30分の練習時間あり

## (3)アクティブ期間のパラメータ

アクティブ期間の注目度を算出するための行動種類毎の重み付け係数  $E_i$  は、たとえば印刷行動はユーザに与えた影響が高いと推測されるので大きくするというように、ヒューリスティックに決定した(表 4-2)。

表 4-2 注目度算出パラメータ

行動の種類(単位)	重み付け係数
閲覧(秒)	0.17
コピー(回)	10
文字列選択(回)	50
印刷(回)	50
マウス操作(回)	0.1
キーボード入力(回)	0.1

### 4.3 評価実験の手順

WEB ページを閲覧する体験ステップと、提案手法あるいは従来手法を利用して体験獲得情報を想起する想起ステップの 2 段階で行う(図 4-2)。得られた実験データを分析し、提案手法と従来方法の有効性を比較する。また、実験終了後に、被験者にアンケートを行った。各手順について説明する。

#### (1) 体験ステップ

被験者は、与えられた課題に対して 30 分間自由に WEB ページを閲覧する。そして閲覧した WEB ページから得た知識を記述する。特に重要と考えた知識 10 個に必ず印をつける。ユーザの行動を含む WEB 閲覧履歴は履歴自動収集ツールによって記録される。また課題はサーベイ検索タスク[10]を独自に作成した。表 4-3 にタスク例、被験者の記述例、および一つの課題に対して一人の被験者が獲得した知識数の平均を示す。

#### (2) 想起ステップ

体験ステップの一週間後に、行動シーケンス表示インタフェース、あるいは従来手法を用いて、体験獲得情報の想起を行う。被験者は、想起しようとする体験ステップの期間の WEB 閲覧履歴を、行動シーケンス表示インタフェース、あるいは従来のインタフェースに表示し、想起した知識を解答用紙に随時記述する。回答の記述時間を含め、実施時間は 15 分とした。また、開始から 4、7、10、15 分でどこまで想起できたかを区別できるように記述した。

それぞれの手法で、被験者および課題の偏りがないようにした。つまり、それぞれの被験者は提案手法を用いて 2 課題、従来手法を用いて 2 課題回答した。各課題は、提案手法あるいは従来手法に 2 回ずつ均等に割り当てられた。最終的に、それぞれの手法につき 40 件のデータを収集した。

### (3) アンケート

実験後に被験者全員にアンケートを行った。『思い出すのに役に立った機能に丸をつけてください(複数選択可)』という質問項目に対して、被験者全員から回答を得た。



図 4-2 実験手順 (体験ステップおよび想起ステップ)

表 4-3 タスク例および被験者の記述例

検索課題の例
3連休を利用して離島に行くことにしました。3つの候補地(種子島、石垣島、佐渡島)について自由に調べてください。
被験者の記述例
・石垣島でサーフィンが困難
・種子島までは鹿児島から9410円
・古座間味ビーチは魅力的
被験者が記述した知識数の平均
19.9 (個/人・課題)

#### 4.4 評価実験の結果と考察

どれだけ体験獲得情報を想起できたかを評価するため、式(8)のように獲得スコアを算出した。重要な体験獲得情報を優先的に思い出すことが望ましいと考えられるため、重要であるとユーザが印をつけた体験獲得情報は印の無い体験獲得情報の2つ分の重みをつけた。

$$\text{獲得スコア} : \text{Score} = (2a + b) / (2c + d) \quad (8)$$

$a$  : 想起した重要な知識数

$b$  : 想起した通常の知識数

$c$  : 閲覧時に獲得した重要な知識数

$d$  : 閲覧時に獲得した通常の知識数

行動シーケンス表示インタフェースと従来手法のそれぞれについて、全被験者の獲得スコアの平均を図4-3に示す。提案手法を利用した場合、従来手法に比べ、ユーザが体験獲得情報を想起していることがわかる。対応のない場合のt検定[9](有意水準を5%)によると、本評価結果の4分、7分、10分、15分のt値は3.88、4.97、5.73、4.04であり、tの臨界値は1.99であるため、提案手法は従来手法と比べ、すべての期間において有意差があると判断される。

図4-4に、想起ステップの開始からの経過時間に沿った、全被験者の単位時間当たりの獲得スコアの平均を示す。この図4-4において提案手法は従来手法に比べ、想起開始から7分までにおいて得に高い値を示している。

図4-5にアンケートの集計結果を示す。サムネイル表示と注目箇所表示は、多くの被験者が有効であったと回答した。アクティブ期間の重要度を利用するサムネイル枠の強調表示/フィルタリング機能は、半数弱の被験者が有効であると回答した。タイトル表示、アンカテキスト表示についても、半数弱の被験者が有効であると回答した。また、矢印表示機能は、ほとんどの被験者が有効だと回答しなかった。

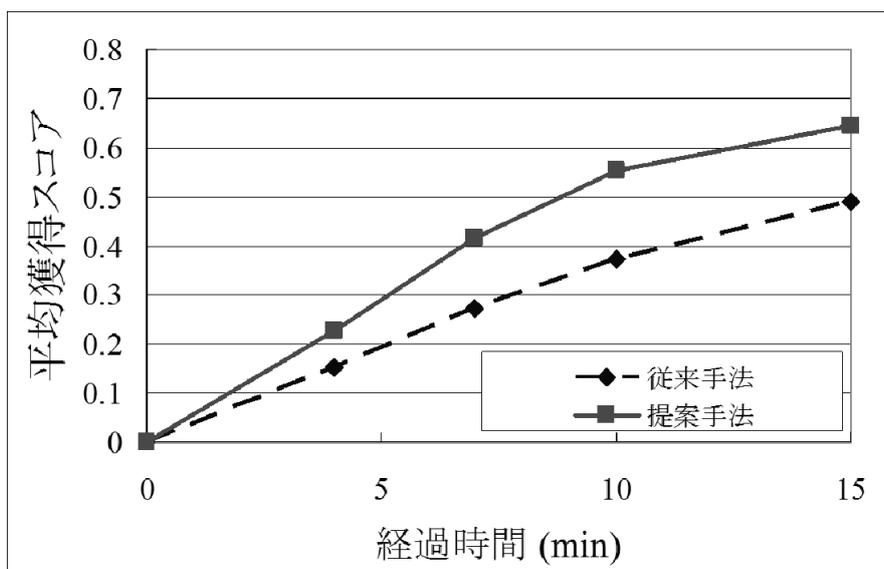


図 4-3 全被験者の平均獲得スコア

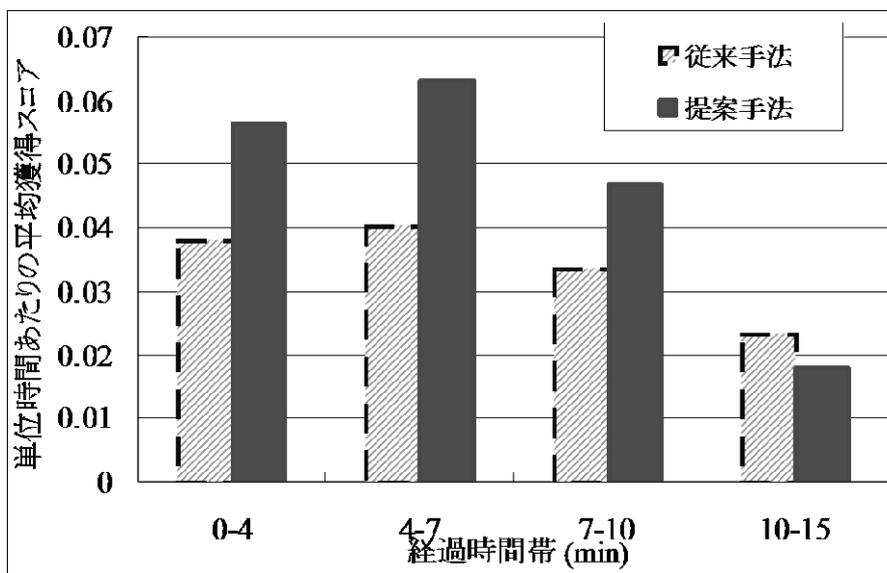


図 4-4 全被験者の単位時間当たりの平均獲得スコア

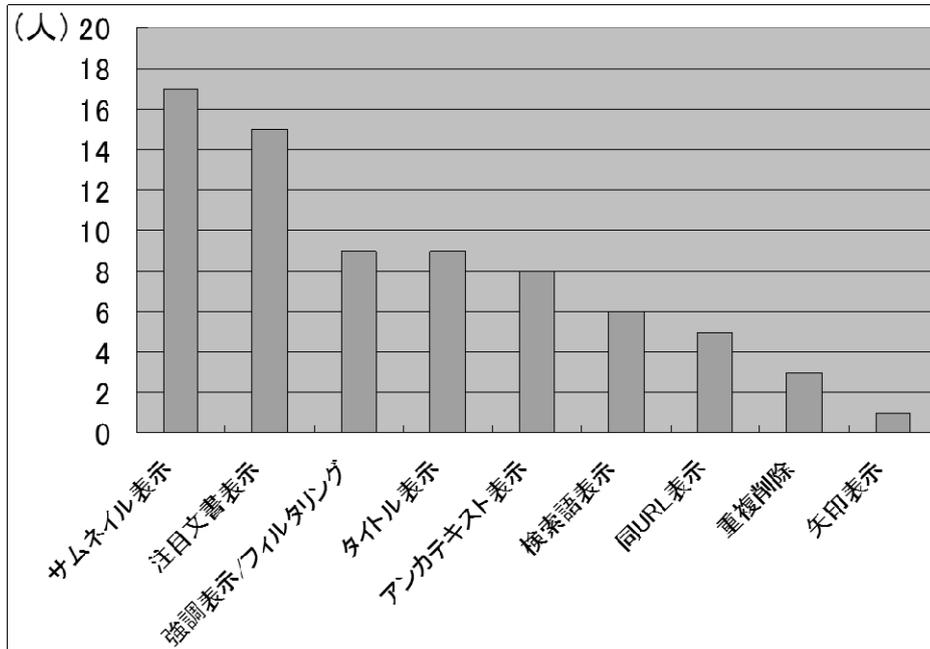


図 4-5 アンケート結果 (役立つ機能)

全被験者の獲得スコアの平均(図 4-3)により、今回のサーベイ検索タスクにおいて、行動シーケンス表示インタフェースが、従来の手法に比べ WEB ページから得た知識を効率的に想起させていることがわかった。

全被験者の単位時間当たりの平均獲得スコア(図 4-4)により、行動シーケンス表示インタフェースは、特に想起開始からの初期の期間で、従来手法に比べ効率的に体験獲得情報を想起させていることがわかった。今回の実験では、30 分間の WEB 閲覧体験を想起する状況を設定したが、ユーザはそのために 30 分の期間を費やすとは考えられない。より短い間に効率よく想起することを望んでいると考えられる。この点で、提案手法の有効性が確認できる。

次に行動シーケンス表示インタフェースのどの機能が WEB ページから得た知識を想起させるのに有効であったかを考察する。被験者の行動シーケンス表示インタフェースの利用方法を観察したところ、

- 時系列に沿って、サムネイルを右クリックして順次拡大する
- 吹き出しのあるサムネイルを、想起開始からの初期の期間にクリックして注目箇所を表示する
- フィルタリングの操作バーを調整し、重要度の低いアクティブ期間を非表示にする
- 赤枠やオレンジ枠で囲まれたサムネイルを、想起開始からの初期の期間に拡大表示する

という操作が多く見られた。アンケート結果において、サムネイル表示と注目箇所表示は、多くの被験者が有効であったと回答しており、サムネイル枠の強調表示/フィルタリング機能、タイトル表示、アンカテキスト表示については、半数弱の被験者が有効であると回答している。よって、下記のように定性的に評価できる。

- ◆ サムネイルを時系列に表示し、クリック一つで拡大表示できる機能は有効であった。多くのWEBページを短期間で理解させ、また、どのような順でWEBページを閲覧したかを短期間で理解させたと思われる。
- ◆ 注目箇所の概要を行動シーケンス表示インタフェースで表示し、注目箇所表示で確認できる機能は、有効であった。知識を得たWEBページやWEBページ内の文章が短期間で発見できた。WEBページの内容を発見してからWEBページから得た知識を想起するまでの時間を短縮したと思われる。
- ◆ 強調表示/フィルタリング機能は、多くの被験者に利用されており、ある程度有効であった。短期間で知識を得たWEBページを発見させたと思われる。また、被験者からインターネット検索エンジンの検索結果のページは重要ではないが強調表示されるので改善してほしい、などの意見があった。本機能は、強調表示する必要の無いページの識別や、アクティブ期間の重要度の算出精度を向上させることで、より有効になるであろう。
- ◆ タイトル表示、アンカテキスト表示機能は、利用者や利用状況によって有効となる場合がある。
- ◆ ウィンドウの遷移方法を示す矢印表示は有効ではなかった。これは、時系列にサムネイルが表示されており、その情報だけで利用者にとって十分であったからではないか。

#### 4.5 公開実験とWEBアンケートによる評価

インターネット上で試作システムの公開実験を行うとともに、WEB閲覧履歴を用いたサービスへのWEBアンケートを行った。その結果、647名から有効回答を得た。また、647名のうち、369名が実際に試作システムを利用した。この369名からは試作システムに関するWEBアンケート項目の回答も得た。WEBアンケートの回答者は、男性364名、女性384名であり、年代では男女ともに20代から40代が中心であった(図4-5)

N=647

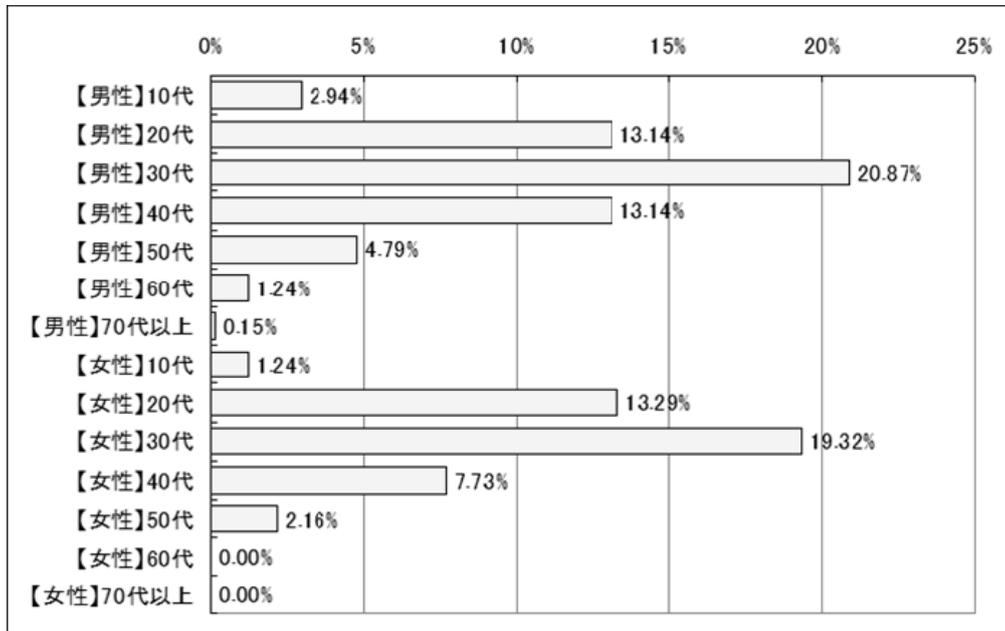


図 4-6 アンケート回答者の年齢 (N=647)

図 4-7 に『履歴を保存することに対する現在の意識を教えてください』という質問に対する回答結果を示す。回答結果を見ると、サーバ、ローカルデスクにかかわらず履歴を保存したくない者は、全体の 22.0%であり、逆に、WEB 閲覧履歴の保存を許容する者は 78%と高い結果であった。この結果は、WEB 閲覧履歴の保存は、プライバシー的な嫌悪感等の理由から受け付けられないという考えは、比較的少数であること示している。また、本試作システムを利用前では WEB 閲覧履歴の収集に否定的であったが、本試作システム利用後では WEB 閲覧履歴を保存してもよいと答えた者が、全体の 23.8%（そのうち、サーバに保存してもよいと思った者は 7.1%、ローカルデスク内であれば保存してもよいと思ったものは 16.7%）であった。この結果は、履歴を活用するサービスに魅力があれば、WEB 閲覧履歴収集の抵抗が下がることを示している。よって、WEB 閲覧履歴をさらに有効に利用できる手法を提供すれば、WEB 閲覧履歴を保存を許容する者はさらに増加すると思われる。

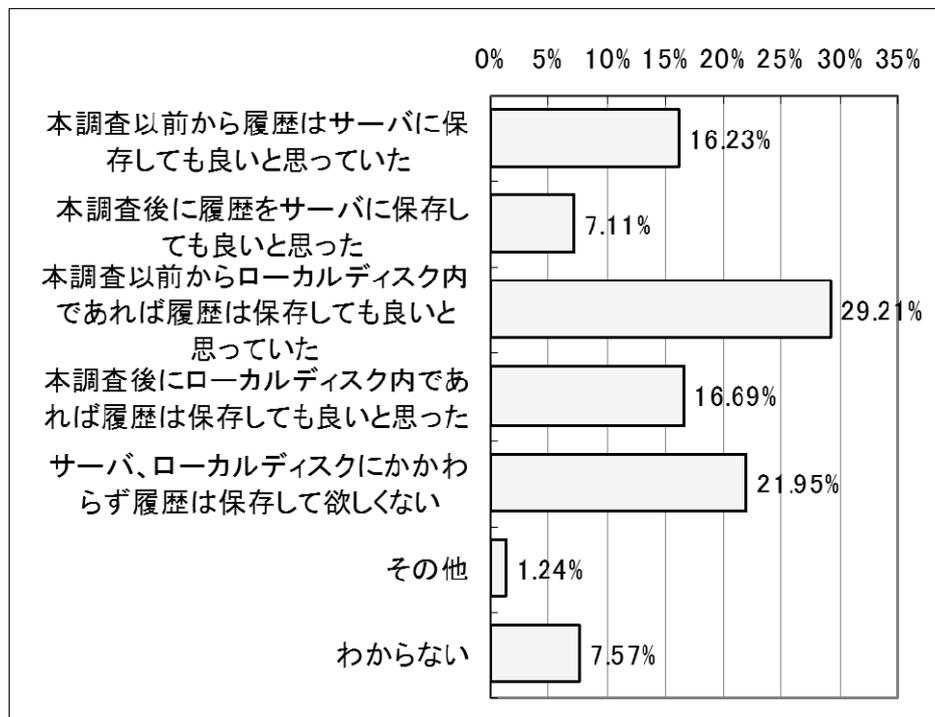


図 4-7 アンケート回答『履歴を保存することに対する現在の意識を教えてください』  
(N=647)

図 4-8 に『期間単位と WEB ページ単位がありますが、どちらのほうが使いやすかったですか』という質問を、試作システムを利用した者に対して行った結果を示す。全体の 40.1% が期間単位のほうが使いやすかったと答えており、集中期間検索インタフェースおよび行動シーケンス表示インタフェースを用いた集中期間の検索と提示は、新しい検索方法として有効であることが分かった。次に、期間単位のほうが使いやすかったと答えた者に対して、『期間単位がよいと感じた理由はなぜですか?』という質問を行った結果を図 4-9 に示す。本研究で想定した通り、『複数の WEB ページをまとめて見つけれられたから』『なぜ WEB ページを見たのか思い出しやすかったから』という回答が上位を占めた。しかし、WEB ページのどの文章に注目したのかがわかったから』という回答は 16.2% と少数を占めた。これは、注目した文章を表示するためには注目した文章を選択するという手間が必要であること、および、注目した文章を選択すればあとで表示されるという履歴検索ツールの使い方を十分ユーザに伝えられなかったことが原因であると思われる。

N=369

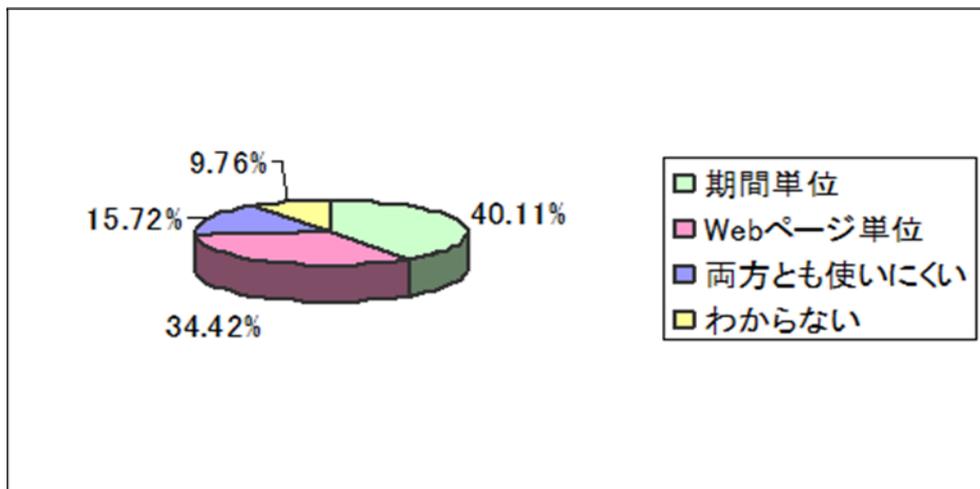


図 4-8 アンケート回答『期間単位とWEBページ単位がありますが、どちらのほうが使いやすかったですか』（N=369）

N=148 ※複数回答

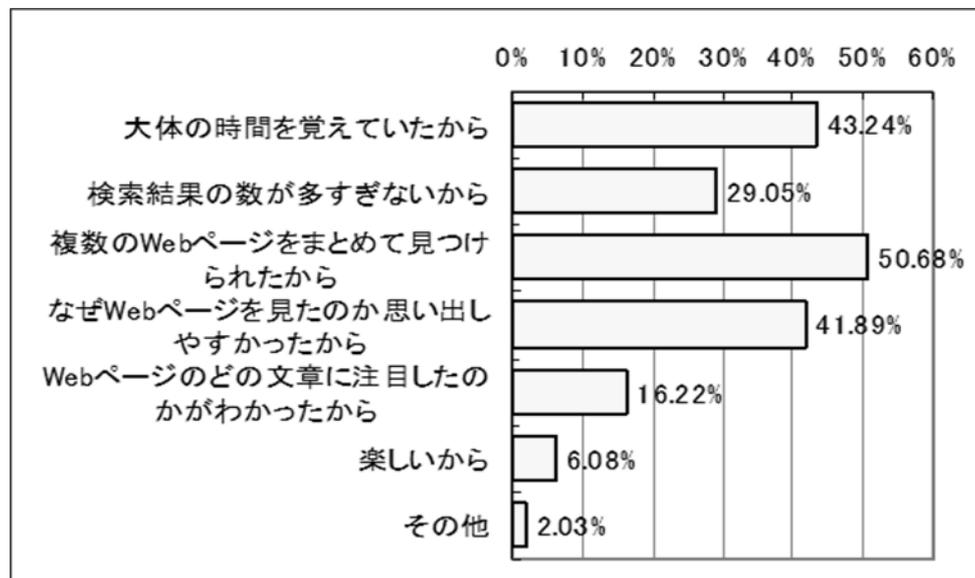


図 4-9 アンケート回答『期間単位がよいと感じた理由はなぜですか？』（N=148）

図 4-10 に『行動詳細表示のランク表示（赤，オレンジ，青）は重要だと思ったページを示していますが、適切でしたか』という質問の結果を示す。とても適切だった、および、

まあまあだったと回答した利用者は 62.1%であった。これは、再検索キーワードに対するアクティブ期間の重要度がある程度の満足度を得られたことを示している。ただし、あまり適切ではなかった、およびまったく適切ではなかったと回答した者も 14.4%おり、再検索キーワードに対するアクティブ期間の重要度の算出方法は、さらにチューニングする余地があることを示している。

N=369

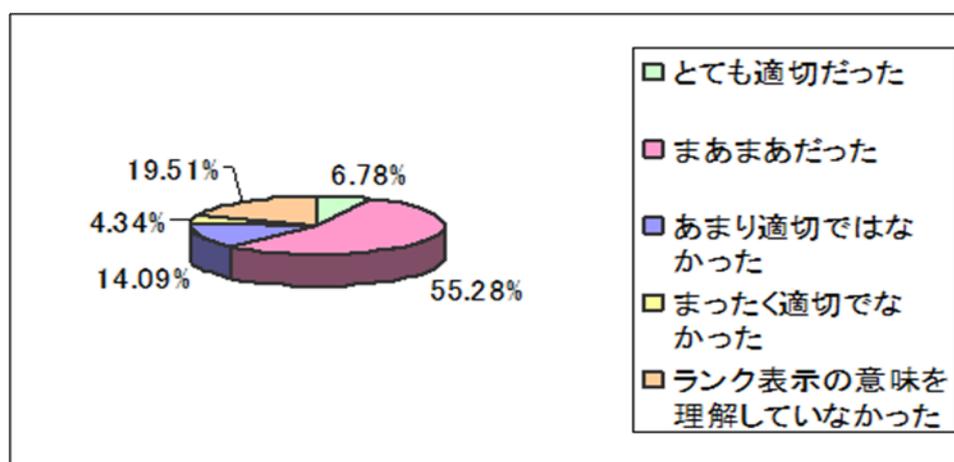


図 4-10 アンケート回答『行動詳細表示のランク表示（赤，オレンジ，青）は重要だと思ったページを示していますが、適切でしたか』（N=148）

#### 4.6 今後の展望

3章および4章では再検索キーワードを入力することによってWEB閲覧履歴を検索・提示する手法について述べたが、本研究を通じて、そもそも気付いていないWEB閲覧履歴は探せない(探そうと思えない)ことが大きな課題であることがわかった。そこで、過去のWEB閲覧履歴の存在をユーザに気付かせ、提示する機能が有効であると考えられる。過去の行動を気付かせるため、閲覧中のWEBページのテキスト情報をリアルタイムに解析してキーワードを抽出し(気付きキーワードと呼ぶ)、過去の行動を自動的に検索する機能は有用ではないだろうか[29]。気付きキーワードは、過去に閲覧したWEBページを母集合とし、HTMLタグによって重み付けされたTF-IDF値の高いものとする事で抽出できると思われる。そして、気付きキーワードを抽出することができれば、集中期間やWEBページをWEBブラウザのパーティカルバーに自動的に表示するなど[30][48]のインタフェースが考えられる(図4-11)。本研究との関係を図4-12のようになる。

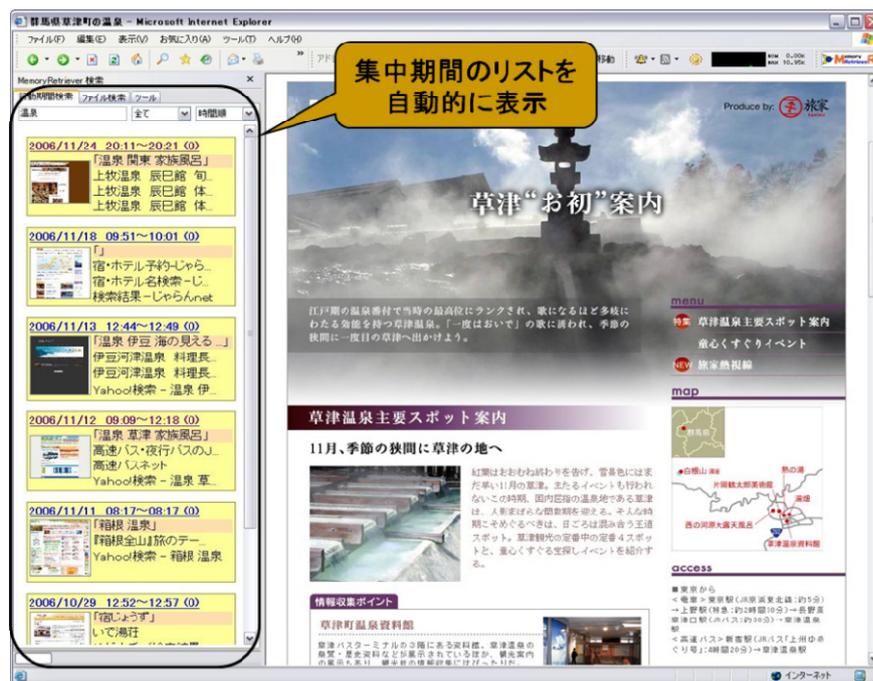


図 4-11 ユーザの負荷なしに集中期間を気付かせるインタフェースの例



図 4-12 気付き機能と、3章および4章で述べたインタフェースとの関係図

## 5. 第三者による WEB 閲覧履歴提示手法と試作システムの実装

### 5.1 対話型オンラインショップ

インターネット上に商品を提示する WEB ページを用意して、無人で 24 時間商品を販売する EC サイトが多く存在する（以降、従来型オンラインショップと呼ぶ）。従来型オンラインショップでは、売り上げを伸ばすために、低価格での販売や幅広い品ぞろえだけでなく、顧客の要望や不満に対応していく必要がある。つまり、顧客満足度を高める必要がある。一方、ブロードバンド回線、その中でも高速な光回線の普及が進んでいる。多くの高帯域を利用したサービスが計画され、実際に導入されている[56][57][58]。そこで、店舗映像の配信システム、テレビ電話システム、および EC サイトがシームレスに融合した新しいタイプのオンラインショップ（以降、対話型オンラインショップと呼ぶ）が普及することを想定する。

#### 5.1.1 対話型オンラインショップの機能

対話型オンラインショップが提供するサービスは、WEB サイトの商品ページによって販売する従来からのサービスのほかに、広帯域回線を利用して、EC サイトをみながら店員とテレビ電話ができるサービス、および実店舗の商品棚をリアルタイム配信するサービスを備えると想定する。

図 5-1 に、対話型オンラインショップの構成を示す。特徴を以下に示す。

- a) メンテナンスコストをかけず、日々変化するリアル店舗内の商品画像を提示する。
- b) テレビ電話会議システムを利用して、店員と顧客がフェイストゥフェイスで会話する。

これらの機能により、下記にあげる従来型オンラインショップの課題を軽減できる

課題-a) 従来型オンラインショップでは、店員がメンテナンスしない限り、陳列する商品を変えられない。よって日々変わる商品をオンラインで販売するのが難しい。

課題-b) 顧客はしばしば商品に対して十分な情報が得られず、不満を感じることもある。その結果、商品の購入に至らない。(たとえば、興味を持った植物が、毎日水をやらなくてはならないのかどうか、特別な肥料が必要か等)

また、対話型オンラインショップの利点として、顧客に店員の顔やリアル店舗の画像をみせることで、親近感や安心感を与えられることがある。また、店員の余剰時間を有効に利用できる店舗もあるであろう。

図 5-2 は、生花を販売する対話型オンラインショップの顧客向け画面例である。画面右側に、リアル店舗の画像がリアルタイムに表示される。画面右上の左側は、鉢植えの陳列棚を一つのカメラで撮影し提示している。顧客は詳しく見たい鉢植えをクリックすると、クリックした部分が拡大して表示され、画像を詳しく見ることができる。画面右下には、同様に店舗の入り口の画像が表示されている。画面左上には、テレビ電話システムのボタンが表示されている。顧客は店員呼び出しボタンを押すと、店員と話をする事ができる。画面左下には、顧客自身の画像が表示されている。

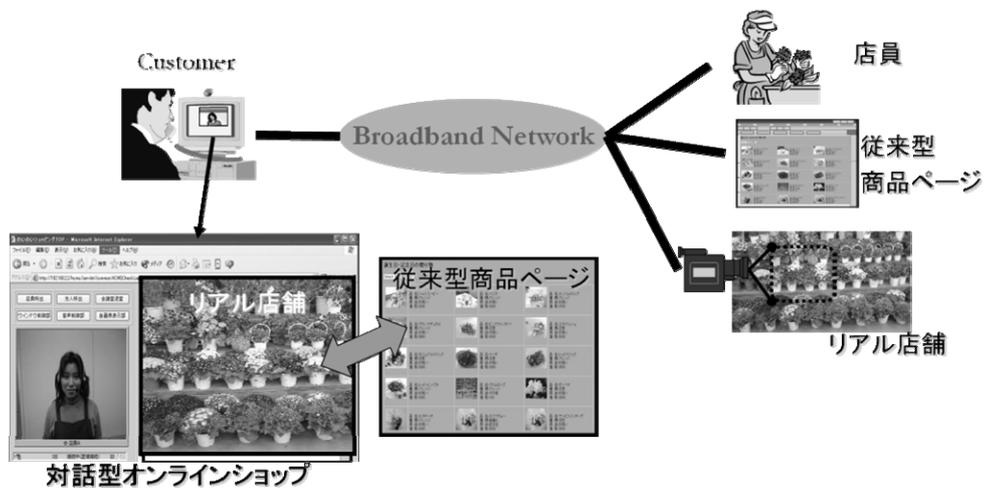


図 5-1 対話型オンラインショップの構成



図 5-2 対話型オンラインショップの顧客向け画面の例

また、対話型オンラインショップの顧客向けの商品提示画面として、次の二つのタイプを想定する。

実店舗陳列棚画像表示型: 顧客は、販売している商品が並ぶ実店舗の陳列棚の写真を閲覧する。陳列棚は、定期的、あるいは日々の入荷状況に合わせて変化する。詳しく見たい陳列棚の特定部分をクリックすると、その場所が拡大して表示され、画像を詳しく見ることができる(図 5-3)。

商品一覧ページ表示型: それぞれの商品がサムネイル画像と短い説明を用いて表示される。詳細な情報が提示される個別ページにリンクされている(図 5-4)。

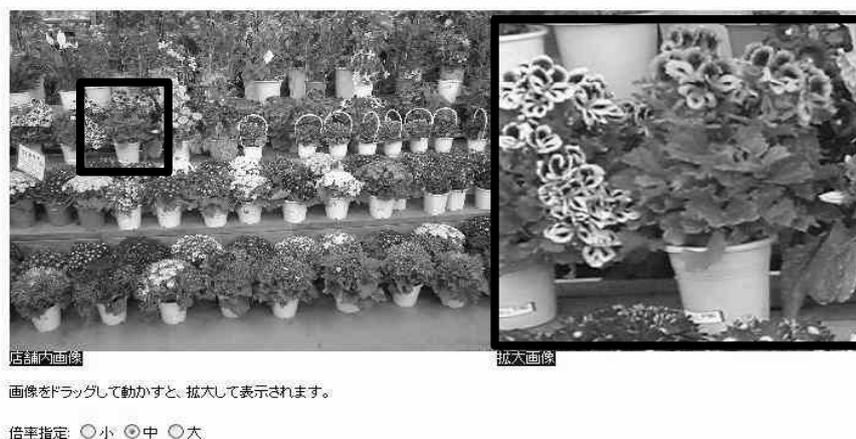


図 5-3 実店舗陳列棚画像表示



図 5-4 商品一覧ページ表示

### 5.1.2 対話型オンラインショップにおける顧客の嗜好情報の必要性

実店舗のように、対面式で会話しながら顧客の相談に乗ることができるのは、従来型オンラインショップと比べた対話型オンラインショップの利点であることを述べた。

しかし、顧客の嗜好がわからなければ、店員は適切なアドバイスを行うことが難しい。逆に、顧客の嗜好が分かれば、店員は顧客に少しの質問をするだけで、顧客満足につながる情報提供を行うことができる。

これを実店舗で考えてみる。実店舗では、会話を行う前に、店員は顧客の店内での行動を観察しており、顧客のなんらかの情報を獲得している。顧客がどのような商品を見ていたのかは店員にとって非常に重要な情報だからである。たとえば、ある顧客が黒のブランド品のスーツを見ていたのであれば、店員は同じようなスーツについての情報を顧客に提供するであろう。もし、店員がなにも顧客に関する情報を知り得なければ、店員は、顧客との会話を繰り返す中で、顧客の嗜好を引き出さなくてはならない。

そして、顧客と会話を行う対話型オンラインショップにとっても、ビジネストークをスムーズに行い、高い顧客満足を獲得するために、顧客の嗜好情報を獲得しておくことは非常に重要であると思われる。

また、対話型オンラインショップには、固定客だけが訪れるのではなく、普段はインターネット上に多数存在する他店舗で商品購入を行っていて、今回初めて自店舗に訪れる顧客が多く訪れる、と想定できる。その場合、店員は、固定客の顧客満足向上だけに偏ったサービスを行うのではなく、初めて訪れた顧客に対してもビジネストークをスムーズに行

い、高い顧客満足を獲得して、新規顧客を固定客にしたいと考えるであろう。つまり、固定客だけでなく、初めて店舗を訪れる顧客の嗜好を獲得することも重要である。

### 5.1.3 抽出対象とする顧客の嗜好情報

顧客の嗜好は、短期的嗜好と長期的嗜好の二つに分類できる。長期的嗜好とは複数回以上のオンラインショップにまたがる嗜好であり、短期的嗜好は、一回のオンラインショッピングの間のみの一時的な嗜好である。たとえば、ワインのボトルを買おうと思っているとき、昔から好きな白ワインを探すかもしれない。これは長期的嗜好である。一方、その日の気分で赤ワインを探すかもしれない。これは短期的嗜好である。また、ドレスを買う時に、その年の流行りピンクのドレスに興味を持つかもしれない。ただし、ドレスのブランドは、その年に流行っているブランドではなく、昔から気に入っているブランドを選ぶかもしれない。これは、長期的嗜好と短期的嗜好の両方が融合した嗜好である。

一方、店員は、対話型オンラインショップ会話で顧客と対話するとき、長期的な嗜好を推定して、顧客に情報を提供することは有効である。また、短期的嗜好を推定して情報を提供することも有効である。しかし、商品を購入させるためにもっとも有効なのは、長期的嗜好と短期的嗜好が融合した商品購入時の嗜好を推定して情報を提供することであると思われる。

そこで、本研究では、短期的嗜好と長期的嗜好が同時に表面化する商品購入時の嗜好をターゲットとする(図 5-5)。具体的な嗜好情報としては、商品購入時の嗜好として店員が把握しておきたいと考えられる『商品』、『商品カテゴリ』、『価格帯』を対象とする。

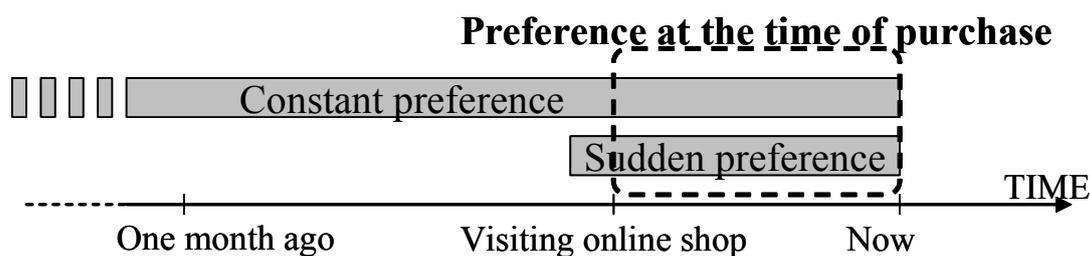


図 5-5 短期的嗜好，長期的嗜好，および商品購入時の嗜好の関係

## 5.2 WEB サーバのアクセスログによる閲覧履歴の収集

対話型オンラインショップを訪れる顧客には、初めての顧客が多く含まれることを前節で述べた。よって、初めての顧客の閲覧履歴を収集する必要がある。しかし初めての顧客

に対し、閲覧履歴を収集するための履歴収集ツールをインストールするなどの手間を要求することは難しい。そこで、顧客には手間を取らせず、対話型オンラインショップ側の事前準備だけで履歴を収集する手法を採用する。具体的には、ブラウザの Cookie 情報を利用して WEB サーバのアクセスログから個人(正確に言うとある OS にログインしているユーザのブラウザ)を特定することにより閲覧履歴を収集する。また、WEB サーバのアクセスログから、WEB サーバに接続されている商品データベースの情報を引き出せるようにする。具体的には、URL に商品を特定できる商品 ID のような属性を含ませ、商品データベースで管理する商品 ID 属性と対応付けることにより、閲覧した WEB ページと商品との関係を正確に把握する。

また、実店舗陳列棚画像表示型の WEB ページを顧客が閲覧した際、陳列棚の全体の画像内で、拡大表示のためにクリックした X, Y 座標と時間を記録することとする。そして、X, Y 座標に置かれている商品はなにであるかというデータベースは存在しないと仮定する。実店舗の陳列棚におかれる商品は、毎日あるいは同日内でしばしば変更されることが予想されるが、商品の配置が変わるたびに、座標と商品の対応付けを店員が手動で行うことは、非常に手間がかかるため、現実的でないと考えられるからである。

### 5.3 対話型オンラインショップの店員に提示するための閲覧履歴の解析

顧客が対話型オンラインショップを訪問してから、店員を呼びだして TV 電話を始めるまでに商品を閲覧した履歴を用いて顧客の嗜好情報を抽出させることを検討する。この履歴を利用することで、すべての顧客に対応できる半面、短いと数分程度のわずかな履歴しか存在にないことが想定される。その条件の中で、閲覧履歴を解析し、履歴の特徴を抽出する必要がある。

閲覧履歴には、顧客の嗜好情報が様々な形で表れる。たとえば、顧客がある商品を長い時間閲覧していたとすると、その商品に興味がある可能性が高い。あるいは、店員に電話をしてくる直前に閲覧していた商品は、興味を持っていた可能性が高いであろう。つまり、店員に顧客の閲覧履歴を表示して顧客の嗜好を理解させるにあたり、EC サイトを閲覧している間の閲覧履歴の詳細な情報を店員に提供できることが重要である。

一方、顧客から対話の要求があった時、店員はすぐに対応しなくてはならない。顧客を待たせることになれば、顧客満足度は上がらない。つまり、店員は顧客の嗜好を理解するために、多くの時間をかけることは好ましくない。

よって、対話型オンラインショップの店員に顧客の嗜好情報を抽出させるための WEB 閲覧履歴提示手法の要件を次のように定める。

- 要件 1: 閲覧履歴のより詳細な情報を把握させなくてはならない
- 要件 2: 一目で顧客の嗜好を抽出させなければならない

この二つの要件を満たすため、次の方針をたて、WEB 閲覧履歴を解析する。

- 要件 1 に対して:

- 商品、価格帯および商品カテゴリごとの閲覧履歴を提示できること

→たとえば、顧客がある特定の価格帯やカテゴリの商品の多くを閲覧している場合、その顧客はそれら価格帯や商品カテゴリに興味があると思われる。そこで、商品一覧ページ表示型の WEB ページの閲覧履歴に対しては、収集した WEB 閲覧履歴の URL が、どの商品を表示していたのか、また、その商品の価格はいくらか、その商品のカテゴリは何に属するかを商品データベースから取得することとする。

実店舗陳列棚画像表示型の WEB ページの閲覧履歴に対しては、陳列棚画像の X, Y 座標に置かれている商品はなにであるかというデータベースは存在しないと仮定するため、陳列棚画像を画像処理して自動的に商品を識別する技術などを利用しない限り、商品、価格帯やカテゴリを取得することはすることは難しい。また、画像処理によって任意の商品を識別できればよいが、識別精度の問題だけでなく、例えばアレンジメントの花などの一時的な商品は、商品自体のデータが商品データベースに存在せず、画像処理を行うため正解データが無いことにより商品の識別が不可能である場合も考えられる。そこで、商品自体を識別することはせず、顧客が拡大して閲覧した領域を抽出するだけにとどめ商品、価格帯および商品カテゴリの識別は、拡大して閲覧した領域が表示されるインタフェース（次節にて説明）を利用する人間（店員）にゆだねることとする。

- どれだけ長い時間、それぞれの商品、あるいは価格帯および商品カテゴリに属している商品群を閲覧していたかを提示できること

→たとえば、顧客がある特定の商品、価格帯、あるいは商品カテゴリに興味を持っている場合、それらをより長い時間閲覧する傾向があると思われる。そこで商品一覧ページ表示型の WEB ページの閲覧履歴に対しては、WEB ページ(説明の都合上、WEB ページ A と呼ぶ)を閲覧していた時間の長さは、次の WEB ページ(説明の都合上、WEB ページ B と呼ぶ)にアクセスするまでの時間の長さであるとして抽出する。次に、WEB ページ A の URL から、表示された商品を抽出する。ここで、厳密に言えば、ある WEB ページ A を見た後で、メールの閲覧やブラウザの他のタブで別サイトの WEB ページを閲覧し、その後に WEB ページ B にアクセスした場合などは、正確な WEB ページ A の閲覧時間とは言えない。メール閲覧やブラウザの他のタブで別サイトを閲覧していた時間も、WEB ページ A が閲覧されていた時間

と判断されるからである。しかし、前節で述べたように、正確な時間を取得するために顧客に履歴収集ツールをインストールさせることは難しい。そこで正確さは下がるが、より多くの利用者の閲覧時間が取得できることを優先することとする。また、価格帯および商品カテゴリに属している商品群の閲覧時間は、それぞれの価格帯および商品カテゴリに属する商品の閲覧時間を積算することで求める。

実店舗陳列棚画像表示型の WEB ページの閲覧履歴に対しては、拡大表示した部分を閲覧していた時間の長さは、次に別の個所を拡大表示するためにクリックする、あるいは別の WEB ページにアクセスするまでの時間の長さであるとして抽出する。この場合も、厳密に言えば、他のアプリケーション等を利用して時間を含まれる場合があるが、時間の正確性よりも利用者の利便性を優先することとする。

- 商品の閲覧順序を提示できること

→顧客の嗜好は閲覧した順序に現れると思われる。そこで、商品一覧ページ表示型の WEB ページの閲覧履歴に対しては、WEB ページにアクセスした時間を利用し、閲覧順序を抽出する。

実店舗陳列棚画像表示型の WEB ページの閲覧履歴に対しては、陳列棚を拡大表示する部分が変化した時間を利用し、閲覧順序を抽出する。

- 要件 2 に対して:

- グラフや写真を利用して閲覧履歴を提示できること

→店員の素早い理解や直感を支援することにより、顧客の嗜好を抽出できると思われる。そこで、商品一覧ページ表示型の WEB ページの閲覧履歴に対しては、WEB ページの URL から抽出した商品 ID により、商品データベースを検索し、提示用の商品画像を取得する。

また、実店舗陳列棚画像表示型の WEB ページの閲覧履歴に対しては、実店舗の陳列棚画像を利用する。拡大した領域を提示するための画像は、実店舗の陳列棚画像全体から拡大した領域を切り出す処理により取得する。

また、実店舗の陳列棚画像のどこを閲覧したかを提示するための画像の作成方法についてであるが、基本的な原理は、まず、実店舗の陳列棚の画像全体に半透明の黒色を重ねることで、画像全体を暗くする。次に、顧客がクリックして拡大表示した場所に半透明の白色を重ね合わせることで、拡大表示した場所を明るくする。

詳細な実装方法としては、Highlight 提示インタフェースで表示する画像を短時間で作成するために二つの画像を用いて作成する。一つは実店舗の陳列棚の写真であり、もうひとつは、クリックした領域を示すために、陳列棚の写真にかぶせるマスク画像である。

ステップ 1) マスク画像を作成する：

マスク画像は、黒色の背景の上に、適切な透明度を設定した白い長方形を、顧客が拡大表示した場所に重ね合わせて作成する。この白い画像を重ね合わせる処理は、すべての拡大表示履歴に対して行う。

ステップ 2) 最終的な表示画像を作成する：

ステップ 1 で作成したマスク画像を適切な透明度を設定して実店舗の陳列棚の画像上に重ね合わせる。

上記透明度を設定するために、アルファ値を利用する。アルファ値とは透明度を表す数値であり最大値の 255 を設定している時は不透明で数値が小さくなるにつれて透明度が上がる。0 を設定した場合には完全な透明となる。

## 5.4 顧客の嗜好情報抽出のための WEB 閲覧履歴提示インタフェース

WEB サーバのアクセスログと商品データベースによって閲覧履歴を収集し、前節で説明した手法で閲覧履歴を解析した結果は、複数のインタフェースを用いて対話型オンラインショップの店員に提示される。店員はこれらインタフェースをぱっと見て、短時間で効率的に顧客の嗜好情報を抽出し、顧客との対話を開始する。

### 5.4.1 Bubble-graph 提示インタフェース

商品一覧ページ表示型の WEB ページのための WEB 閲覧履歴可視化インタフェースを二種類提案するが、そのうちの 하나가 Bubble-graph 提示インタフェースである。本提示インタフェースでは、3 つの属性を表示することに広く利用されているバブルグラフを応用する。

前節で述べた方針に従って構築したが、商品、価格帯および商品カテゴリごとの閲覧履歴を提示できることという方針に対しては、商品、価格帯を提示することとする。本インタフェースを用いて商品カテゴリを表示することも可能であるが、それは次節で述べる Category 提示インタフェースで実現することとする。

本提示インタフェースを図 5-6, 図 5-7 に示す:

- バブルグラフを用いて閲覧履歴を提示
  - 縦軸は価格帯、あるいは商品カテゴリを表示。横軸は閲覧時刻を表示
  - 顧客の閲覧履歴を、商品画像を用いて表示
  - 商品画像の大きさは、商品ごとの閲覧時間の長さに比例
- とする。

本提示インタフェースにより、価格帯、商品ごとの閲覧時間、商品の閲覧順序を提示することができる。商品カテゴリはこのグラフの商品画像をみることで把握することができる。このように詳細な閲覧履歴を表示することにより、店員が顧客の嗜好を抽出することを支援する。

また、本インタフェースはグラフと画像を用いて、商品、価格帯、閲覧した順番、閲覧時間の長さを表示しているため、店員は、一目見て WEB 閲覧履歴の特徴を把握でき、顧客の嗜好の抽出に注力できると思われる。

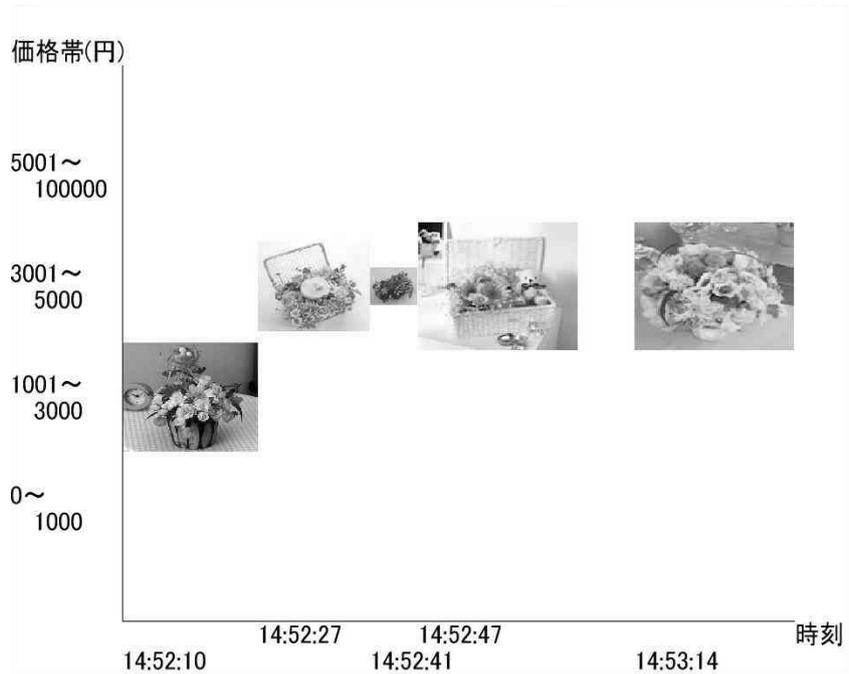


図 5-6 Bubble-graph 提示インタフェース 1

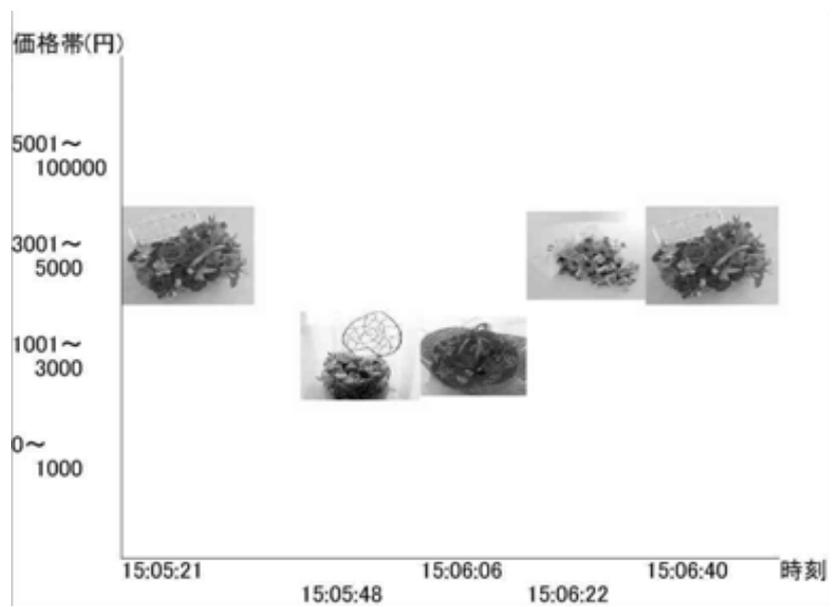


図 5-7 Bubble-graph 提示インタフェース 2

## 5.4.2 Category 提示インタフェース

商品一覧ページ表示型の WEB ページのための WEB 閲覧履歴可視化インタフェースのもう一つが Category 提示インタフェースである。Category 提示インタフェースでは、2 つの属性を表示することに広く利用されている二次元セルを応用する。

前節で述べた方針に従って構築したが、商品、価格帯および商品カテゴリごとの閲覧履歴を提示できることという方針に対しては、価格帯および商品カテゴリを提示することとする。本インタフェースは、商品単体の嗜好を表示することは向いていないが、それは前節で述べた Bubble-graph 提示インタフェースで実現することとする。

本提示インタフェースを図 5-8, 図 5-9 に示す:

- 二次元セルを用いて閲覧履歴を提示
- 縦軸は価格帯、横軸は商品カテゴリを表示
- 閲覧履歴は、商品の代表画像を用いてそれぞれのセルに表示
- 閲覧時間をバー形式でセルの下部分に表示
- それぞれのセルの左上に、閲覧した商品の数を表示。また、フレームの色を商品数によって変化させる。

とする。

また、本提示インタフェースにより、価格帯、商品カテゴリ、価格帯ごとの閲覧時間、商品カテゴリごとの閲覧時間、および価格帯と商品カテゴリに属する閲覧した商品の個数を提示することができる。このように詳細な閲覧履歴を表示することにより、店員が顧客の嗜好を抽出することを支援する。

また、本インタフェースはグラフと画像を用いて、価格帯、商品カテゴリ、それらの閲覧時間の長さを表示しているため、店員は、一目見て WEB 閲覧履歴の特徴を把握でき、顧客の嗜好の抽出に注力できると思われる。

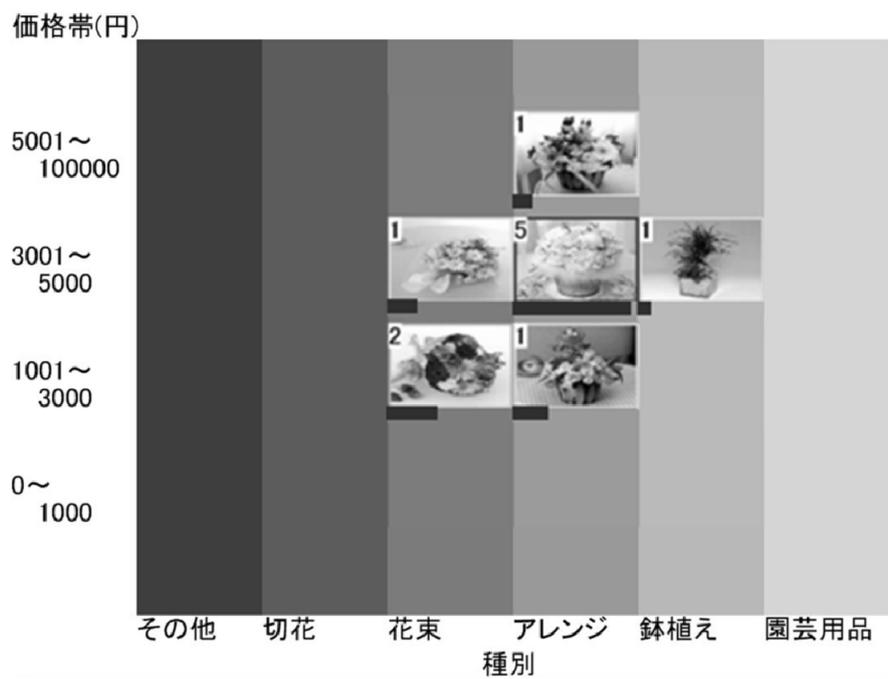


図 5-8 Category 提示インタフェース 1

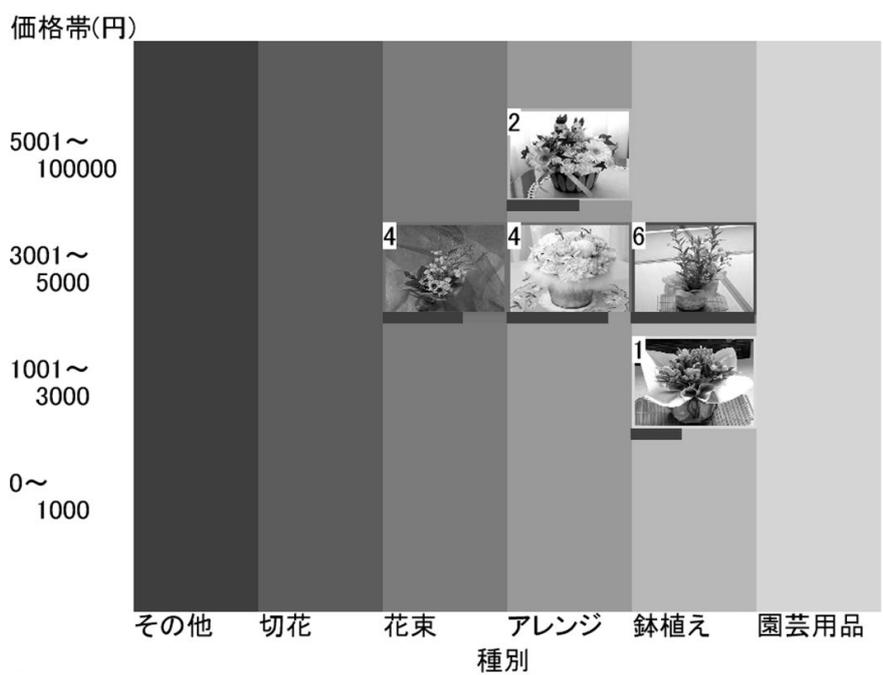


図 5-9 Category 提示インタフェース 2

### 5.4.3 Highlight 提示インターフェース

実店舗陳列棚画像表示型のための WEB 閲覧履歴可視化インターフェースが、顧客が閲覧した部分をハイライトして提示するインターフェース(Highlight 提示インターフェース)である。

前節で述べた方針に従って構築し、商品、価格帯および商品カテゴリごとの閲覧履歴を提示できることという方針に対しては、実店舗陳列棚画像を活かしたまま、拡大して閲覧された部分を一目でわかるようにすることで、顧客が興味を持った可能性がある商品、価格帯および商品カテゴリを提示する。

店員に表示するインターフェースを図 5-10、図 5-11 に示す:

- 拡大して閲覧された部分をハイライトして表示。ハイライトの強さは、閲覧回数に比例させる
- 拡大して閲覧した部分を、全体画像の外に時系列に並べて表示
- 拡大して閲覧した時間の長さを、全体画像の外にバー形式で表示とする。

本提示インターフェースにより、陳列棚が価格帯あるいは商品カテゴリによって分けられていた場合、価格帯あるいは商品カテゴリはハイライトされた場所、および拡大順序を示す表示領域により推定される。拡大操作ごとの閲覧時間、拡大表示された順序が示される。このように詳細な閲覧履歴を表示することにより、店員が顧客の嗜好を抽出することを支援する。

また、本インターフェースは画像とグラフィカルなバーを用いて、拡大した領域、拡大表示した順序、拡大した部分の閲覧時間の長さを表示しているため、店員は、一目見て WEB 閲覧履歴の特徴を把握でき、顧客の嗜好の抽出に注力できると思われる。



図 5-10 Highlight 提示インターフェース 1



図 5-11 Highlight 提示インターフェース 2

## 5.5 第三者による WEB 閲覧履歴提示手法の試作システムの実装

5.2 章で述べた WEB サーバのアクセスログと商品データベースにより WEB 閲覧履歴を収集し、5.3 章で説明した WEB 閲覧履歴の解析を行い、5.4 章で述べたインタフェースで店員に表示する試作システムを実装した。試作システムを構成する機器は、WEB サーバ、店員用の PC、テレビ電話用サーバ、そして顧客 PC である(図 5-12)。

それぞれの機器の機能を次に示すように実装した。

### ● WEB サーバ:

WEB サーバで、顧客のアクセスログ、および商品データをデータベース管理システム(DBMS)に蓄積する。実店舗の商品棚の写真と商品の写真も WEB サーバによって管理する。顧客向けに提示される画面を図 5-2 に示す。画面右側に商品を表示し、画面左側にテレビ電話画面を表示する。

商品表示部 ... 商品は WEB ページの右側に表示される。対話型オンラインショップの一例として、二つの異なるレイアウトの花屋を構築した。一つは実店舗陳列棚画像表示型の店舗であり(図 5-3)、もうひとつは商品一覧ページ表示型の店舗である(図 5-4)。

テレビ電話表示部... 店員の映像、店員呼び出しボタン、コントロールボタンが WEB ページの左上側に表示される。顧客自身の映像が、左下側に表示される。

### ●店員用 PC:

店舗端末は、実店舗の商品棚を撮影するためのビデオカメラ、テレビ電話を行うためのビデオカメラ、ヘッドセット、そして店員用アプリケーションを搭載した PC で構成される。実店舗の商品棚を撮影した写真は自動的にあるいは手動で WEB サーバに転送される。店員は顧客の嗜好情報を店員用アプリケーションを利用して抽出する。(図 5-13)。

顧客の閲覧履歴表示部... 顧客の WEB 閲覧履歴がスクリーンの左側に表示される。閲覧履歴は Highlight 提示インタフェースあるいは Category 提示インタフェースで表示される。店員は、Category 提示インタフェースと Bubble-graph 提示インタフェースを切り替えることができる。

顧客リスト... 店舗にログインしている顧客のリストがスクリーンの左側に表示される。店員は会話する顧客を選択することができる。

テレビ電話表示部... 顧客の映像とテレビ電話コントロールボタンがスクリーンの左上側に表示される。店員自身の映像が画面の左下側に表示される。

### ● TV 電話サーバ:

テレビ電話サーバは、3人以上の複数ユーザ(e.g., 店員, 顧客, 顧客の友人)と会話できる機能を有すソフトウェア[57][58]をインストールして利用した。

● 顧客 PC:

一般的に市販されているカメラとヘッドセットを PC に接続した。OS は WindowsXP であり, WEB ブラウザは Internet Explorer を利用した。

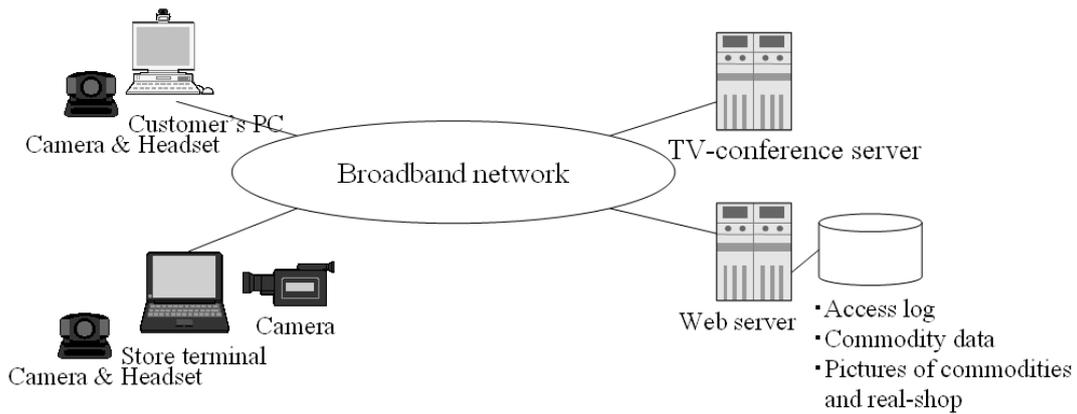


図 5-12 試作システムの構成図



図 5-13 店員向けアプリケーションの画面例

## 5.6 第三者による嗜好抽出手法に関する関連研究

Electronic Customer Relationship Management (E-CRM) やデータマイニングが WEB サーバに訪れた顧客の嗜好情報を抽出，利用する実用的なサービスとして研究・開発されてきた[51][52]．

E-CRM は，カスタマイズされたホームページを，登録された顧客の情報を用いて作成する，あるいは，顧客の購入履歴を同じような購入履歴をもつ顧客と比較することにより，顧客ごとのお勧めの商品を推薦する[53][54]等，情報システムを応用してインターネット上の店舗などが顧客と長期的な関係を築く手法である．顧客データベースを作成し，商品の売買，問い合わせ，クレームへの対応や保守サービスなど，個々の顧客とのやり取りを一貫して管理することにより実現する．それにより，顧客の要求に個別に対応することで，顧客の満足度を高め，顧客と一度だけの売買の関係ではなく，常連として長期に囲い込んで収益を上げようとする営みである[49]．これらのツールは，店員や EC システムに顧客の嗜好を理解させ，活用させる．しかし，E-CRM は長期に顧客との関係を築くことを目標としているため，本手法のように一度だけの購入時の嗜好を抽出することに注力した研究は少ない．

データマイニングは大量のデータの中からある程度の時間を許容して顧客の嗜好を抽出する[55][56]．数ページの WEB 閲覧履歴というわずかな履歴を用いて短時間で顧客の嗜好を抽出するためには，これらデータマイニングの技術をそのまま利用するのは難しい．

一方，顧客の嗜好を抽出するために，顧客が訪問した WEB ページに対してなんらかの方法で評価付けを行い，さらに WEB ページのコンテンツを自然言語処理等を用いて解析することで，顧客の嗜好を抽出する手法が数多く研究されている．

Foltz らの手法[50]では，閲覧者に対して閲覧した WEB ページに明示的に評価づけをさせることで，閲覧者の嗜好情報を抽出している．

森田らはネットニュースの記事の閲覧時間を観測し，閲覧時間と記事への興味には相関があることを示している[64]．閲覧時間による重みづけを行って適合性フィードバックした手法は再現率 20%の場合精度が 60%弱であり，この精度は Foltz らの明示的な精度と大きく差はないことから，閲覧時間による重みづけは効果的であることが分かる．逆に，記事の長さ，記事の読みやすさ，未読の記事の数などは閲覧時間とは相関が無いと述べている．

マウスの動きを用いて閲覧者の嗜好を抽出する手法も提案されている．Kamba らの研究[66]では，WEB 上のニュース記事に対するクリック動作やウィンドウ拡大/縮小操作から閲覧者が興味を持ったかを推定している．土方らの手法[65]では，なぞり読み，リンククリック，リンクポインティング，およびテキスト選択により，ユーザが WEB ページ内で

興味を持ったテキストを推定している。これらの操作を行った場所にあるテキストを文、または行として抽出することにより、TF-IDF よりも高い精度でキーワードを抽出できる文章があることを述べている。

また、複数のユーザの閲覧履歴を解析し、似たような閲覧行動を行っているユーザの履歴を活用して情報を提示する手法も広く研究されている[53][67]。これらの手法は、閲覧した WEB ページの内容、あるいは複数人のアクセス履歴を解析することにより直接顧客の嗜好や目的とする情報を抽出しているが、本手法は履歴をビジュアルに提示することにより店員に顧客の嗜好情報を抽出させることが異なる。

また、WEB サーバの閲覧履歴を可視化する研究としては、WEB ページの解析技術[69][70] や WEB サーバ管理者支援のための利用動向可視化システム等がある[68]。これらは対象とする WEB サーバの利用動向を抽出することを目的としており、特定の個人の嗜好情報を抽出する本手法とは目的が異なる。

## 6. 第三者による WEB 閲覧履歴提示手法の評価

5章で述べた第三者に対する WEB 閲覧履歴提示手法の評価実験を行う。評価実験に利用するシステムは、5.5 節で述べた試作システムを利用する。対話型オンラインショップの店員が顧客の嗜好を正しく抽出できているかどうかを評価し、評価指標としては店員の抽出結果の正解率を選択する。

### 6.1 評価実験の手順

次に示す手順で評価する。

- (1) 顧客が商品ページを閲覧する。閲覧する店舗は、実店舗陳列棚画像表示型の店舗、および商品一覧ページ表示型の店舗のいずれかを指示する。
- (2) 顧客は興味の高い商品を 3 つまでと、興味の高いカテゴリか価格帯を 3 つまで（カテゴリと価格帯の両方が混在していても良い）を報告する。さらに、EC サイト閲覧中に嗜好が変化した場合は報告する。
- (3) 3 つの提案インタフェースのうち一つを利用し、店員は顧客の嗜好として、3 つの商品、3 つのカテゴリか価格帯を推測する。また、店員は商品閲覧中の顧客の嗜好の変化を推測する。店員が推測に要する時間は、10 秒間と、時間無制限との両方の場合で行う。
- (4) 従来手法として“リスト提示インタフェース”（後述）を利用し、上記と別の店員が、顧客の嗜好として 3 つの商品、3 つのカテゴリか価格帯を推測する。また、店員は商品閲覧中の顧客の嗜好の変化を推測する。ここで、“リスト提示インタフェース”とは、広く利用されている表計算ソフトで閲覧履歴を表示する方法であり、具体的には、閲覧開始時間、閲覧終了時間、商品名、商品説明、価格、および商品カテゴリをリストで表示し、ボタンのクリックでそれぞれの項目ごとに並べ替えることができる手法である（図 6-1）（図 6-2）。ここで、リスト提示インタフェースは、商品一覧ページ表示型の店舗を閲覧した履歴に対してのみ用意し、実店舗陳列棚画像表示型の店舗に対しては用意しなかった。なぜなら、実店舗陳列棚表示型の店舗を閲覧したログは、顧客がクリックした画像中の  $x$ 、 $y$  座標になるため、これらをリスト提示インタフェースで表示しても、商品の名前やカテゴリを店員が推定するのは困難であるからである。
- (5) 最終的に、店員が顧客の嗜好を正しく抽出できているかどうかを求める。具体的には、店員が推測した 3 つの商品、3 つのカテゴリか価格帯に、顧客が報告した嗜好が含まれていれば正解とし、正解率を求める。顧客の嗜好としては、最も興味の高い商品、二番目に興味の高い商品、最も興味の高い商品カテゴリか価格帯、二番目に興味の高い商品カテゴリか価格帯について調べる。

Start time, Stop time, Commodity name,  
commodity's explanation, Price, Category

図 6-1 リスト提示インタフェースで表示する項目

Start time	Stop time	Commodity name	Explanation	Price	Category
2003-08-2216:21:18	2003-08-2216:21:18	シャイニーピー	ナチュラルに持ち	3200	arrangement
2003-08-2216:21:18	2003-08-2216:21:18	ランナチュラル	爽やかさと上品さ	2500	bouquets
2003-08-2216:16:59	2003-08-2216:17:00	ピンクファンタジー	空間をしなやかな	1500	bouquets
2003-08-2216:16:42.6	2003-08-2216:16:43	スタイリッシュ	コンパクトながらに	2000	Potted plants
2003-08-2216:18:10	2003-08-2216:18:11	カジュアルマジック	籐製花器アイアン	3000	arrangement
2003-08-2216:19:30	2003-08-2216:17:00	メルカート	オトボケ顔のハチ	3000	arrangement
2003-08-2216:17:37	2003-08-2216:17:38	サンダーソニア	まるで花畑に舞い	2500	bouquets
2003-08-2216:20:03	2003-08-2216:20:04	ドルチェ	明るくナチュラルな	5000	arrangement
2003-08-2216:21:18	2003-08-2216:21:19	スウィング	アレンジ中央に配	4000	arrangement
2003-08-2216:17:47	2003-08-2216:21:18	ピンクファンタジー	空間をしなやかな	1500	bouquets
2003-08-2216:21:18	2003-08-2216:20:00	スタイリッシュ	コンパクトながらに	2000	Potted plants
2003-08-2216:17:47	2003-08-2216:21:18	サンダーソニア	まるで花畑に舞い	2500	bouquets

図 6-2 リスト提示インタフェースの画面例

## 6.2 評価実験のパラメータ

商品一覧ページ表示型の店舗には、50 個の商品を表示した。商品のカテゴリは、花束、アレンジメント、および鉢植えとした。商品価格帯としては、0-1000 円、1001-3000 円、3001-5000 円、5001-10000 円とした。

実店舗陳列棚画像表示型の店舗には、70 個の商品を提示した。商品のカテゴリは、7 つの色（赤、白、ピンク、黄色、紫、バイオレット、複数の混合色）とした。すべての商品は、一つの価格帯に属するとした。

Highlight 提示インタフェースで利用するマスク画像を作成するためのアルファ値は、予備実験から 168/255 とした。マスク画像を実店舗の陳列棚の画像に重ねるときのアルファ値は、予備実験から 80/255 とした。

顧客は 20 代から 40 代の男性 5 名、女性 1 名が行い、実店舗陳列棚画像表示型および商品一覧ページ表示型の店舗それぞれに対して 20 回の商品閲覧を行った。また、店員は 20 代から 40 代の男性 5 名で行った。

一つの提示手法ごとに、上記 20 回の閲覧履歴を利用して、5 人の店員のからランダムに選んだ 3 人の店員が顧客の嗜好を抽出した。よって、インタフェースごとに 60 回分の嗜好抽出を行った。

また、実験に参加した顧客がより実際に近い閲覧行動を行うように努めた。具体的には顧客の商品閲覧ごとに、商品を購入する背景を指定した。例えば、"会社同期の異性の誕生日。実はひそかに恋心を抱いている。歳は 32 歳。仲良し 4 人でご飯を食べに行くが、そのときにお花をプレゼントしてびっくりさせたい。女性の容姿性格はおとなしめの本庄まなみ。"などである。指定した状況リストを表 6-1 に示す。

表 6-1 実験参加者に指示した顧客の状況リストの例

#####

状況リスト:(下記のうち1つにシーンを指示)

#####

シーン0 70歳のおじいちゃんの退院祝いのお食事会。実はおじいちゃんはガンで余命3ヶ月。かわいがってくれたお礼のお花をプレゼントしたい。

病室は4人部屋で普通の大きさ。

シーン1 口うるさい姑が家に遊びに来るので、リビングに飾る花を想买いたい。

シーン2 母の日がもうすぐなので、ちょっと変わった花をプレゼントしてびっくりさせたい

お母さんは、結構派手好み。手入れもできる。

シーン3 会社同期の異性の誕生日。実はひそかに恋心を抱いている。

年は32歳。仲良し4人でご飯を食べに行くが、そのときにお花をプレゼントしてびっくりさせたい。女性の容姿性格はおとなしめの本庄まなみ。

シーン4 大先輩の派遣さんが退職することになった。会社で一番お世話になった人だ。

年は48歳。容姿性格は、背の低い和田あきこ。

シーン5 なんとなく部屋が殺風景。気分も暗い。

花を飾ってリフレッシュすることにした。さていい花ないかな。

シーン6 おばあちゃんの米寿祝がある。

親戚20人ぐらいでフランス料理を食べに行く。実は私はおしゃれな人で通っている。

我が家族(妻、子供2人)からもお祝いの花を贈りたい。

おばあちゃんをよくしゃべる元気ばあちゃん。

シーン7 日ごろお世話になっている親しい親戚にお中元を贈ることになった。

親戚は花に詳しく、へたな花だといやみを言われかねない。自分らしい贈り物をしたい。

シーン8 大学の同期に子供が生まれた。お祝いの花を贈りたい。

シーン9 悪夢の交通事故で家族4人のうち友人だけが奇跡的に助かった。ちょうど彼の誕生日なので、励ましの意味を込めて花を贈りたい

### 6.3 評価実験の結果と考察

対話型オンラインショップの顧客が商品閲覧した時間の長さは次のようになった。

#### 実店舗陳列棚画像表示型

平均閲覧時間： 1分34秒

最大閲覧時間： 3分2秒

#### 商品一覧ページ表示型

平均閲覧時間： 3分28秒

最大閲覧時間： 8分42秒

平均閲覧時間は、実際の商品購入時に比べ比較的短いと思われる。しかし、本実験で収集すべき閲覧履歴は、店員と会話をする前の商品閲覧行動の履歴であることを考慮すると、実際の状況と大きく差はないと考えられる。

これらの WEB 閲覧履歴を利用して顧客の嗜好情報を抽出し、正解データと比較して、統計的な有意差があるかどうかを調べた。

#### 6.3.1 Bubble-graph 提示インタフェースの評価結果と考察

Bubble-graph 提示インタフェースあるいはリスト提示インタフェースを用いた場合の商品の嗜好の正解率を図 6-3 に示す。10 秒以内に顧客の嗜好を推測する場合、二番目に興味の高い商品の正解率は Bubble-graph 提示インタフェースのほうがリスト提示インタフェースよりも高い。一方、一番興味を持った商品および、時間無制限で顧客の嗜好を推測した場合、両者の正解率は変わらなかった。

Bubble-graph 提示インタフェースがリスト提示インタフェースより良い結果を示した原因としては、店員は顧客の閲覧時間および閲覧順序の両方を利用して顧客の嗜好を推測したからだと思われる。

残念ながら、今回実験を行った数は十分ではなかった。よって、統計的に Bubble-graph 提示インタフェースがリスト提示インタフェースよりも 10 秒間でより多くの顧客の嗜好を抽出できることを示すことはできなかった。

また、顧客の閲覧履歴と正解データを個別に比較した。すべての場合に当てはまるわけではないが、最後に見たのは 1 位にはいる場合が多い、長く見たものは上位に入る、迷っているときは最後の方が重要度高い、という傾向があった。これらは Bubble-graph 提示インタフェースによって短時間で把握しやすい項目であり、それが正解率の向上につながったと考えられる。

さらに、なぜ一番興味を持った商品は二番目に興味を持った商品に比べて正解率が低かったのか分析した。その結果、顧客はしばしば最も興味を持った商品を一度か二度だけ閲覧しており、二番目に興味を持った商品に比べ短い時間しか閲覧していなかった。この理由の一つは、顧客はまず直感的に商品を気に入り、その比較対象を長い時間かけて閲覧することが考えられる。顧客がこの行為を行った場合、Bubble-graph 提示インタフェースを利用して顧客の嗜好情報を抽出することは難しい。そのような直感的な嗜好を抽出するためには、テキストチャットやテレビ電話会議で顧客とインタラクティブに会話することで抽出することが適切であると思われる。

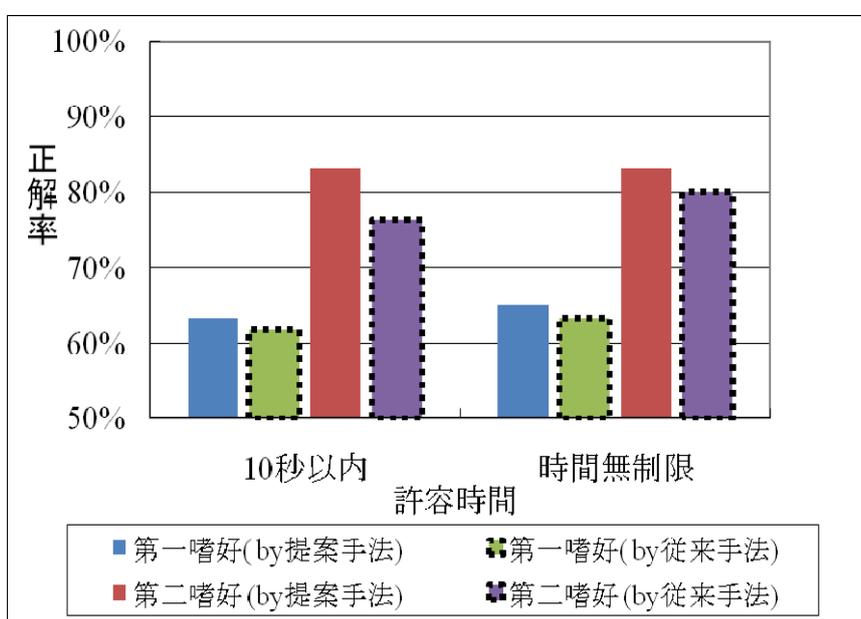


図 6-3 Bubble-graph 提示インタフェースの正解率

### 6.3.2 Category 提示インタフェースの評価結果と考察

Category 提示インタフェースあるいはリスト提示インタフェースをもちいた場合の、商品カテゴリあるいは価格帯の嗜好の正解率を図 6-4 に示す。10 秒以内に、Category 提示インタフェースを用いて一番興味の高い商品カテゴリあるいは価格帯を推測した場合、正解率は約 90%以上であり、リスト提示インタフェースと比べて高い結果を示した。一方、時間無制限で顧客のカテゴリあるいは価格帯を推測した場合、両者の正解率は変わらなかった。

Category 提示インタフェースがリスト提示インタフェースより良い結果を示した原因としては、店員は顧客が見た商品のカテゴリあるいは価格帯に属している商品の数、および閲覧時間を一目で把握して顧客の嗜好を推測したからだと思われる。

10 秒間で Category 提示インタフェースがリスト提示インタフェースに比べ高精度に顧客の嗜好情報を抽出していることを統計的に確認するため、二つの比率を比較するための統計的手法を利用した[63]。帰無仮説として、Category 提示インタフェースを用いた嗜好抽出の正解率は、リスト提示インタフェースを用いた嗜好抽出の正解率に比べて高くない、とした。対立仮説として、Category 提示インタフェースを用いた嗜好抽出の正解率は、リスト提示インタフェースを用いた嗜好抽出の正解率に比べて高い、とした。 $p_1 - p_2$  は常に正となると想定されるため片側検定を行った。二つの比率を比較するための統計的手法[63]によると、検定統計量  $Z_0 > 1.645$  であれば、帰無仮説は有意水準 5%で棄却される。検定統計量  $Z_0$  の計算式を下記に記す。この結果より、10 秒間で行う Category 提示インタフェースを用いた嗜好抽出の正解率は、リスト提示インタフェースを用いた嗜好抽出の正解率に比べて高いことが統計的に示された。

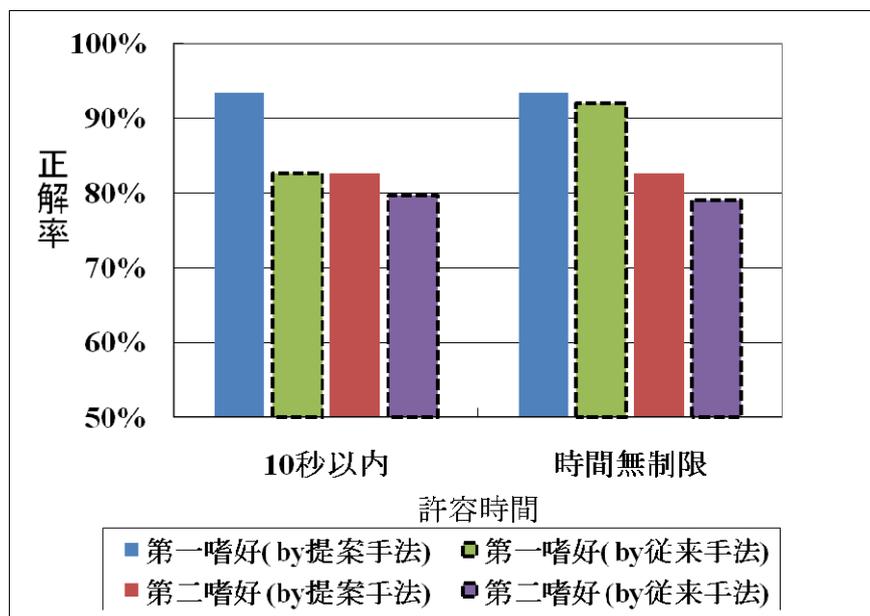


図 6-4 Category 提示インタフェースの正解率

### 二つの比率を比較するための仮説検定:

“Category 提示インタフェース”

Sample size:  $n_1 = 75$ , Number of success:  $x_1 = 70$

Population proportion:  $p_1$

Sample proportion:  $\hat{p}_1 = x_1 / n_1 = 0.933$

“List 提示インタフェース”

Sample size:  $n_2 = 75$ , Number of success:  $x_2 = 62$

Population proportion:  $p_2$

Sample proportion:  $\hat{p}_2 = x_2 / n_2 = 0.827$

The null hypothesis is  $H_0 : p_1 = p_2$

The alternative hypothesis:  $H_a : p_1 > p_2$  (one-side test)

The test statistic:

$$\begin{aligned} Z_0 &= \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}} \\ &= 2.01 > 1.645 \end{aligned}$$

(significance level is 5%)

where  $\hat{p} = (x_1 + x_2) / (n_1 + n_2)$

### 6.3.3 Highlight 提示インタフェースの評価結果と考察

Highlight 提示インタフェースを用いた場合の商品カテゴリあるいは価格帯の嗜好の正解率を 95% 信頼区間で表したグラフを図 6-5 に示す .10 秒以内に顧客の嗜好を推測する場合, Highlight 提示インタフェースを用いて一番興味を持った商品カテゴリあるいは価格帯を推測した場合, 正解率は約 80% である. この値は十分高いと思われる. なぜなら, 店員が WEB サーバに残されたクリック座標のアクセスログを見て顧客の嗜好を推測した場合は非常に正解率が低くなることはあきらかであり, また, 80% という数字は, 対話型オンラインショップにおいて顧客と会話するための事前情報としては有用であると考えられるからである. 複数の店員は, 何度も閲覧した商品と一度しか閲覧していない商品との区別が, ハイライトの程度があまり変わらないため分らないとの報告があった. 提示インタフ

エースのチューニングを行い，本問題を解決すれば，正解率がさらに高まることが期待できる．

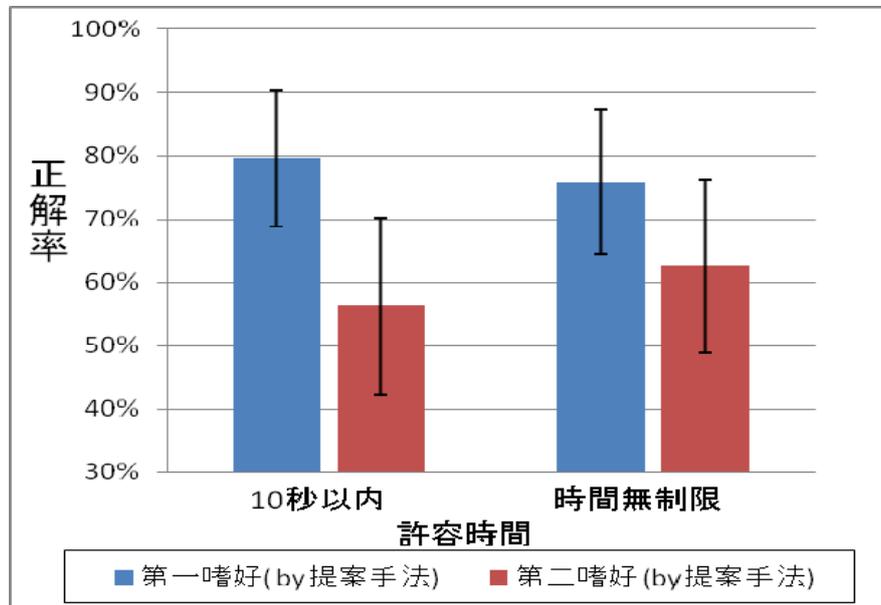


図 6-5 Highlight 提示インタフェースの正解率

## 7. 結論

本論文では、自分自身あるいは第三者に過去に得た知識や嗜好情報を抽出させるため、WEB 閲覧履歴を自動的に収集して記録し、主に時間に着目して閲覧履歴を解析して、自分自身あるいは第三者にビジュアルに検索・提示させる手法について述べた。また、評価実験によりその有効性を示した。

WEB 閲覧履歴の活用で必要となる処理を、WEB 閲覧履歴をどのように収集するかという履歴収集ステップ、収集した WEB 閲覧履歴を解析することでユーザに提示する情報を抽出する履歴解析ステップ、そして解析した履歴の特徴を利用して、利用者に情報を提示する情報提示ステップの3つに分類し、それぞれのステップの特徴を整理した。履歴収集ステップは、履歴を収集する場所によって大きく二つに分類される。一つは、WEB ページを提供する WEB サーバで履歴を収集するもの、もうひとつは、WEB ページを閲覧するローカルの PC 上で履歴を収集するものである。本論文において、自分自身による WEB 閲覧履歴検索・提示手法では、すべてのサイトのより詳細な閲覧履歴を取得するためローカル PC 上で履歴を収集する手法を探求し、第三者による WEB 履歴提示手法では、ユーザへの負担を軽減するために WEB サーバで閲覧履歴を収集する手法を採用した。履歴解析ステップおよび情報提示ステップは、人間の介在の有無によって、大きく二つに分類した。ひとつは、目的とする情報を抽出するために、人間の介在を前提とする手法であり、人間の能力、特に暗黙知を活用できるという利点を述べた。一方で人間の能力を利用するため、一人の人間が万人分の嗜好情報を抽出するといったことは難しく、効率的に抽出できる嗜好情報の量には限界があることを述べた。もうひとつの手法は、目的とする情報を抽出するために、人間の介在を必要としない手法である。目的とする情報そのものを提示するため、利用者は提示された情報から新たに何かを推測あるいは想起する必要はないが、人間の能力を活用できない問題点があることを述べた。

3 章および 4 章では、過去の WEB ページの閲覧体験で得られた WEB ページの内容、WEB ページ間の関係、および WEB ページから得た知識を効率的に想起させるため、WEB ページだけではなく、WEB ページを閲覧した期間を抽出することで、効果的に WEB 閲覧履歴を検索・提示する手法を提案した。閲覧時間、コピー、印刷といった行動を含む WEB 閲覧履歴を収集する手法を示すとともに、再検索キーワードに関連した行動を集中して起こした期間、および一つのウィンドウあるいはブラウザタブがある WEB ページをアクティブに表示している期間という新たな概念を提唱し、再検索キーワードに対するそれらの期間の重要度を算出するアルゴリズムを提案した。WEB ページ単位で閲覧履歴を扱うので

はなく、集中期間単位で閲覧履歴を提示することにより、提示する履歴を効率的に絞り込むことができる。また集中期間を検索するインタフェース、およびアクティブ期間や詳細なイベントを用いてユーザの閲覧履歴を提示するインタフェースを提案し、一つの WEB ページだけでなく多くの体験獲得情報を利用者が抽出できるようにした。また、試作システムを作成して実行可能であることを示した。評価実験により、行動シーケンス表示インタフェースによってアクティブ期間等の WEB 閲覧履歴の特徴を表示する手法は、従来の WEB ページを時系列に提示する手法に比べ、WEB ページから得られた知識を、効率的に想起できることを示した。また、行動シーケンス表示インタフェースの個々の機能の有効性について考察した。また、試作システムを一般公開して数百人規模で WEB アンケートを行い、WEB 閲覧履歴の収集に対する抵抗はそれほど高くないこと、集中期間検索インタフェースと行動シーケンス表示インタフェースを用いた集中期間の検索と WEB 閲覧履歴の提示が新しい検索方法として有効であること、および再検索キーワードに対するアクティブ期間の重要度が 60%以上のユーザの満足を獲得したことを示した。また、今後の展望として、過去の WEB 閲覧履歴の存在をユーザに気付かせ、提示する手法について述べた。

5 章および 6 章では、ブロードバンドが普及に伴い出現する新しい EC 店舗の形態として、テレビ電話システムと WEB サイトがシームレスに融合した対話型オンラインショップを想定した。また、対話型オンラインショップは、実店舗の商品棚をリアルタイムに提示して、時間とともに変わる商品を販売できる WEB ページを提供すると想定した。対話型オンラインショップでは、店員が顧客に的確なアドバイスを与えるために、店員が顧客と対話する直前に顧客の嗜好を把握することは重要であることを述べた。そして、店員が顧客の嗜好を抽出するための履歴解析手法および 3 つの表示手法、具体的には 3 つの属性を提示するのに適した“Bubble-graph 提示インタフェース”、二つの属性を提示するのに適した“Category 提示インタフェース”、そして実店舗の商品棚の画像の一部を顧客が拡大表示した履歴をビジュアルに表示する“Highlight 提示インタフェース”を提案した。また、それらの提示手法によって顧客の嗜好がどれだけ抽出できるかを、正解率を指標として評価し、“Category 提示インタフェース”の統計的な有意性、“Highlight 提示インタフェース”の有用性を示した。

## 謝辞

本論文をまとめるにあたり，終始暖かい激励とご指導，ご鞭撻を頂いた筑波大学大学院システム情報工学研究科の田中二郎教授に心より感謝申し上げます．田中二郎教授には，本研究を進め，まとめる上での適切な助言を数多く頂きました．また，学位論文作成にあたり，貴重なご助言を頂いた筑波大学大学院システム情報工学研究科の志築文太郎講師に心より感謝申し上げます．学位論文審査において，貴重なご助言・指導を頂いた筑波大学大学院システム情報工学研究科の北川博之教授，山本幹雄教授，そして三末和男准教授に深く感謝いたします．筑波大学大学院システム情報工学研究科の田中研究室の皆さんには，合宿やゼミにおいて基調なご意見をいただいたことを心より感謝申し上げます．NTT サイバーソリューション研究所の田中明通氏，日高哲雄氏，倉恒子氏，大浦啓一郎氏，加藤泰久氏，奥雅博氏，中村知彦氏，平川豊氏，近藤光正氏，Tian Luo 氏の熱心な協力と数多くのご助言なくしては，本研究は不可能であったことを記すとともに，深甚の謝意を表します．本研究の公開実験では，多大なるご協力，ご助言をいただいた NTT レゾナントの岡本哲氏，石澤悟郎氏に深く感謝いたします．また，研究を進めるにあたり，ご支援，ご協力を頂きながら，ここにお名前を記すことが出来なかった多くの方々に心より感謝申し上げます．

最後になりますが，いつも心の支えになってくれた京都と北海道に住む父，母，姉，弟，そして東京に住む義母に心から感謝します．そして，すてきな笑顔で大きな力をくれた娘，子育てで大変な時期にも全力でサポートしてくれた素晴らしい妻に心から感謝します．

## 参考文献

- [1]総務省,"通信利用動向調査報告書," 2009.
- [2]Paul Chirita, Rita Ghita Alexandru Gavriiloaie, Stefania, Wolfgang Nejdl, Paiu Raluca, "Activity Based Metadata for Semantic Desktop Search," 2nd European Semantic WEB Conference (ESWC05), 2005.
- [3]Gim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker, and Curtis Wong, "Mylifebits: Fulfilling the memex vision," In ACM Multimedia'02 Proceedings, pp. 235-238, 2002.
- [4]"About Google Desktop", <http://desktop.google.com/about.html>
- [5]"Copernic: Software to search, find, and manage information," <http://www.copernic.com/index.html>
- [6]尾川浩一, 松山佐和, 岩崎晴美, "ウィンドウシステム入門," 朝倉書店, 1996.
- [7]北研二, "情報検索アルゴリズム," 共立出版, 2002.
- [8]大浦啓一郎, 日高哲雄, 森田哲之, 加藤泰久, 奥雅博, "NaMiSEN: ウェブ閲覧履歴への個人嗜好情報付加・抽出ツールの開発," 電子情報通信学会技術研究報告, OIS2005-14, 2005.
- [9]森 敏昭, 吉田寿夫, "心理学のためのデータ解析テクニカルブック," 北大路書房, 1990.
- [10]江口浩二, 大山敬三, 石田栄美, 神門典子, 栗山和子, "NTCIR-3 WEB: WEB 検索のための評価ワークショップ," NII Journal, No.6, pp.31-56, 2003.
- [11]阿部 匡史, 豊田 正史, 喜連川 優, "アンカーテキストとリンク構造解析を用いた WEB 情報検索の改善," 第 14 回データ工学ワークショップ(DEWS2003), 2003.
- [12]日高哲雄, 大浦啓一郎, 森田哲之, 倉 恒子, 田中 明通, 加藤泰久, "MemoryArchiver: 忘却機能を考慮した PC 操作履歴蓄積システムの提案," 電子情報通信学会技術研究報告, OIS2006-42, 2006.
- [13]Natasia Milic-Frayling, Rachel Jones, Kerry Rodden, Gavin Smyth, Alan Blackwell, Ralph Sommerer, "SmartBack: Supporting Users in Back Navigation," 13th International World Wide WEB Conference(WWW2004), New York City NY, 2004.
- [14]近藤秀樹, 三宅芳雄, "計算機上での活動履歴を利用する記憶の拡張システムの評価," 13th Workshop on Interactive Systems and Software (WISS2005), 2005.
- [15]Jun Rekimoto, "Time-Machine Computing: A Time-centric Approach for the Information Environment," ACM UIST'99, 1999.
- [16]大澤亮, 高汐一紀, 徳田英幸, "俺デスク: ユーザ操作履歴に基づく情報想起支援ツール," 情報処理学会 第 47 回プログラミング・シンポジウム, 2006.
- [17]増井俊之, 塚田浩二, 高林哲, "近傍関係にもとづく情報検索システム," 11th Workshop on Interactive Systems and Software (WISS2003), 2003.

- [18]Harumi Murakami and Takashi Hirata, "A System for Generating User's Chronological Interest Space from WEB Browsing History," International Journal of Knowledge-Based and Intelligent Engineering Systems, Vol.8, No.3, pp.149-160, 2004.
- [19]山田 和明, 中小路 久美代, 上田 完次, "WEBユーザの行動履歴解析のためのデータマイニング," 第3回 WI2-技術研究報告, 2004.
- [20]MSN Newsbot, <http://uk.newsbot.msn.com/>
- [21]河合由起子, 官上大輔, 田中克己, "個人の嗜好に基づく複数ニュースサイトの記事収集・閲覧システム," 情報処理学会論文誌 データベース, Vol.46, No.SIG8(TOD26), pp.14-25, 2005.
- [22]"goo", <http://www.goo.ne.jp/>
- [23]"Yahoo!", <http://m.www.yahoo.com/>
- [24]MATSUMOTO, Y. et al, "Japanese. Morphological Analysis System ChaSen version. 2.2.1," <http://chasen.aist-nara.ac.jp/chasen/doc/chasen-2.2.1.pdf>
- [25]"Welcome to Lucene!", <http://lucene.apache.org/>
- [26]"ImageMagick: Convert, Edit, and Compose Images", <http://www.imagemagick.org/script/index.php>
- [27]"Camellia What's New", <http://info.isl.ntt.co.jp/crypt/eng/camellia/index.html>
- [28]"CodeProject: Extending Explorer with Band Objects using .NET and Windows Forms. Free source code and programming help", <http://www.codeproject.com/KB/shell/dotnetbandobjects.aspx>
- [29]吉田大我, 中村聡史, 田中克己, "ブラウジングと検索の融合：閲覧履歴からの関連情報の検索・自動提示にもとづくウェブ閲覧," Journal of the DBSJ, Vol.7, No.1, pp.133-138, 2008.
- [30]Tetsushi Morita, Tetsuo Hidaka, Akimichi Tanaka, Yasuhisa Kato, "System for reminding a user of information obtained through a web browsing experience," International World Wide Web Conference(WWW2007), pp.1327-1328, 2007.
- [31]Audience Targeting, Audience Targeting Advertising, Behavioural Targeting, Audience Targeting Companies : AudienceScience, <http://www.audiencescience.com/>
- [32]Chen, L. and Sycara, K., "WebMate: A Personal. Agent for Browsing and Searching," in Proc. of the 2nd International Conference on Autonomous Agent, pp.132-139, 1998.
- [33]Chen, C. C., Chen, M. C., and Sun, Y., "A Web. Document Personalization User Model and System," in Proc. 8th International Conference on User Modelling, 2001.
- [34]Crabtree, I. and Soltysiak, S., "Identifying and Tracking Changing Interests," International Journal of Digital Library, Vol.4, pp.38-53, 1998.
- [35]土方嘉徳, 青木義則, 古井陽之助, 中島周, "マウス挙動に基づくテキスト部分抽出方式と抽出キーワードの有効性に関する検証," 情報処理学会論文誌, Vol.43, No.2, pp.566-576, 2002.

- [36] 大野健彦, "IMPACT: 視線情報の再利用に基づくブラウジング支援法," in Proc. of the 8th Workshop on Interactive Systems and Software, pp.137-146, 2000.
- [37] 松尾豊, 福田隼人, 石塚満, "ユーザ個人の閲覧履歴からのキーワード抽出によるブラウジング支援," 人工知能学会論文誌, Vol.18, No.4, pp.203-211, 2003.
- [38] Jia Li and Osmar R. Zaiane, "Combining Usage, Content, and Structure Data to Improve Web Site Recommendation," In EC-Web, pp.305-315, 2004.
- [39] 高須賀清隆, 丸山一貴, 寺田実, "閲覧履歴を利用した協調フィルタリングによる Web ページ推薦とその評価," DBWS2007, vol.107, No.131, pp.115-120, 2007.
- [40] 森田 哲之, 日高 哲雄, 倉 恒子, 藤本 強, 加藤 泰久, "履歴を利用したローカルファイル検索システム," IEICE2005 総合大会講演論文集, D-8-11, 2005.
- [41] 松本光弘, 清原良三, 沼尾正行, 栗原聡, "携帯電話における時空間的利用履歴を基にしたアプリケーション推薦システム," 電子情報通信学会技術研究報告, 2008-BIO-015, 2008.
- [42] Daniel Ashbrook and Thad Starner, "Learning significant locations and predicting user movement with GPS," In Proceedings of IEEE Sixth International Symposium on Wearable Computing (ISWC02), 2002.
- [43] Paramvir Bahl and Venkata N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," In INFOCOM (2), pp.775-784, 2000.
- [44] T. Tanaka, et al., "Life Microscope: Continuous Daily-Activity Recording System with Tiny Wireless Sensor," Proc. 5th International Conference on Networked Sensing Systems, 2008.
- [45] K. Ara, et al., "Sensible Organization: Changing Our Business and Work Styles through Sensor Data," IPSJ Journal Vol.49, No.4, pp.1625-1636, 2008.
- [46] K. Kitamura, T. Yamasaki, K. Aiazawa, "Food Log by Analyzing Food Images," ACM international conference on Multimedia, pp.999-1000, 2008.
- [47] Anthony LaMarca, Yatin Chawathe, Sunny Consolvo, Jeffrey Hightower, Ian Smith, James Scott, Timothy Sohn, James Howard, Jeff Hughes, Fred Potter, Jason Tabert, Pauline Powledge, Gaetano Borriello, and Bill Schilit, "Place Lab: Device positioning using radio beacons in the wild," In Proceedings of Pervasive 2005, 2005.
- [48] Tian Luo, Tetsushi Morita, Yasuhisa Kato, "Desktop Search Using Context and Categorization," 電子情報通信学会 2006 年総合大会講演論文集, 2006.
- [49] 金子武久, "Positioning the Role of Data Mining in Internet CRM [in Japanese]," 創価経営論集, Vol.29(1/2), pp.57-70, 2004.
- [50] Foltz, P. W. and Dumais, S. T., "Personalized Information Delivery: An Analysis on Information Filtering Methods," Comm. of the ACM, Vol.35, No.12, pp.51-60, 1992.

- [51] M.K. Brohman, R.T. Watson, G. Piccoli, and A. Parasurama, "Data completeness: a key to effective net-based customer service systems," *Communications of the ACM*, vol.46, No.6, pp.47-51, June 2003.
- [52] R.S. Swift, *Accelerating customer relationships: using CRM and relationship technologies*, Prentice Hall, Upper Saddle River NJ, 2000.
- [53] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp.175-186. Chapel Hill, NC, Oct. 1994.
- [54] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating 'word of mouth'," *Proceedings of the SIGCHI conference on human factors in computing systems*, pp.210-217, Denver CO, May 1995.
- [55] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules," *Proceedings of International Conference on Very Large Data Bases*, pp.487-499, Santiago, Chile, 1994.
- [56] R. Srikant, R. Agrawal, "Mining generalized association rules," *Proceedings of the 21st International Conference on Very Large Data Bases*, pp.407-419, Zurich, Switzerland, Sep. 1995.
- [57] S. Honda, T. Yasuno, "Multi-location collaboration support system NetOfficeHIKARI," *NTT Technical Review*, Vol.1, No.1, pp.90-93, Apr. 2003.
- [58] MeetingPlaza Distance Conference Service Over View,  
[http://www.meetingplaza.com/e/mp\\_2d\\_ASP\\_eng.html](http://www.meetingplaza.com/e/mp_2d_ASP_eng.html)
- [59] G.P Premkumar, "Alternate distribution strategies for digital music," *Communications of the ACM*, vol.46, No.9, pp.89-95, Sep. 2003.
- [60] U. Varshney, A. Snow, M. McGivern, C. Howard, "Voice over IP," *Communications of the ACM*, vol.45, No.1, pp.89-96, Jan. 2002.
- [61] Movies.com: Guide to new & upcoming movies, <http://movies.go.com>
- [62] Neil A. Weiss, *Introductory Statistics -6th Ed*, Addison-Wesley, Boston, 2001.
- [63] David S. Moore, George P. McCabe, *Introduction to the Practice of Statistics (Introduction to the Practice of Statistics, 4th Ed)*, W.H.Freeman & Co, New York, 2002.
- [64] Morita, M. and Shinoda, Y., "Information Filtering Based on User Behavior Analysis and Best MatchText Retrieval," in *Proc. of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 272-281, 1994.
- [65] 土方嘉徳, 青木義則, 古井陽之助, 中島周, "マウス挙動に基づくテキスト部分抽出方式と抽出キーワードの有効性に関する検証," *情報処理学会論文誌*, Vol.43, No.2, pp.566-576, 2002.

- [66]Tomonari Kamba, Hidekazu Sakagami, and Yoshiyuki Koseki, "Anatagonomy: A personalized newspaper on the world wide web," International Journal of Human-Computer Studies, Vol.46, No.6, pp.789-803, 1997.
- [67]Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms", Proc.of the 10th International World. Wide Web Conference, 2001.
- [68]戸川聡, 金西計英, 矢野米雄, "WAVISABI : Web 閲覧特性に基づく管理者支援のための利用動向可視化システム," 情報処理学会論文誌, Vol.46, No.4, pp.985-994, 2005 .
- [69]石井研二, アクセスログ解析の教科書, 翔泳社, 2004.
- [70]Itoh T., Yamaguchi Y., Ikehata Y., and Kajinaga Y., "Hierarchical Data Visualization Using a Fast Rectangle-Packing Algorithm," IEEE Transactions on Visualization and Computer Graphics, Vol. 10, No. 3, pp. 302-313, 2004.

## 著者論文リスト

### 【本研究に関する論文】

- [1] 森田哲之, 倉恒子, 日高哲雄, 大浦啓一郎, 田中明通, 加藤泰久, 奥雅博 :  
Memory-Retriever : 体験獲得情報を想起させる行動検索手法, 情報処理学会論文誌,  
Vol.48, No.3, pp. 1197-1208 (2007).
- [2] Tetsushi MORITA, Tetsuo HIDAHA, Tomohiko NAKAMURA, Morihide OINUMA,  
Yutaka HIRAKAWA : Visual Customer Relationship Management System that Supports  
Broadband Network E-Commerce, IEICE transactions on communications, Vol.E87-B, No.7,  
pp. 1789-1796 (2004)
- [3] Tetsushi MORITA, Tetsuo HIDAHA, Tsuneko KURA, Keiichiro OOURA, and Yasuhisa  
KATO, "Desktop search system based on the action-oriented algorism," Proceedings of the  
6th Asia-Pacific Symposium on Information and Telecommunication  
Technologies(APSITT2005), pages 204-207 (2005)
- [4] Tetsushi MORITA, Tetsuo HIDAHA, Tomohiko NAKAMURA, Morihide OINUMA,  
Yutaka HIRAKAWA, "Visual Customer Relationship Management service," Proceedings of  
the 5th Asia-Pacific Symposium on Information and Telecommunication Technologies ,  
pages 31-36 (APSITT2003)