

Memory-Retriever: 体験獲得情報を想起させる行動検索手法

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[†]

『過去に見た』という体験を思い出し、その体験で得られた情報を想起しようとすることは多いのではないだろうか。本論文では、Web ページの内容だけでなく、Web ページ間の関係、および Web ページから得た知識といった Web ページの閲覧体験によって得られた情報を想起させるための、コンピュータ上の過去の行動を検索する手法を提案する。行動を検索するための元データとなるコンピュータ上の行動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ついて着目し、それらの重要度を算出する手法を提案する。さらに、行動を検索するための行動検索インタフェースを示す。また、評価実験により、提案手法は従来手法に比べ、Web ページの閲覧体験によって獲得した重要な知識を、短期間で効率的に想起できることを示す。

Memory-Retriever: Method of Retrieving a User Action History for Reminding Him of Obtained Information

TETSUSHI MORITA,[†] TSUNeko KURA,^{††} TETSUO HIDAKA,[†]
KEIICHIROU OOURA,^{†††} AKIMICHI TANAKA,[†] YASUHISA KATO[†]
and MASAHIRO OKU[†]

Most of us have tried to recollect obtained information by our past experience after we remembered the experience was useful. We propose a method for retrieving user action history. By finding his action history, he can acquire not only passes of web pages but also relation among the web pages and knowledge based on the web pages. We show a technique for collecting user action history which is basic data for the retrieving. We also propose a method for extracting a period when a user intensively takes action related a keyword and a period when he made a webpage shown actively, and a method for estimating the degree of importance of the periods. We also suggest interfaces for retrieving user action history. Finally we show proposed method let user recall knowledge obtained from the web pages viewed in the past more efficiently.

1. はじめに

多くのユーザがコンピュータを使った体験を日々積み重ねてきている。わが国では、パソコン経由でインターネットを利用する者のうち、約 4 割が毎日利用しており、なかでも 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¹⁾。そして、『過去に見た』という体験を思い出し、その体験で得られた情報を想起しようとすることは多いのではないだろうか。

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。そしてデスクトップ検索技術のように、コンピュータ内に保存された Web ページを検索するツールが公開されている^{4),5)}。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

[†] 日本電信電話株式会社 NTT サイバースリユーション研究所
NTT Cyber Solutions Laboratories, NTT Corporation

^{††} 日本電信電話株式会社 NTT サービスインテグレーション基盤
研究所
NTT Service Integration Laboratories, NTT Corpora-
tion

^{†††} 西日本電信電話株式会社技術部研究開発センタ
Technology Department Research and Development
Center, NTT West Corporation

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。

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 1 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、1 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、Web ページを見て得られた知識などの様々な情報があると考えられる。1 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。

本論文では、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、Web ページではなく『行動』を検索する手法を提案する（ここで、本論文では、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）。行動検索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。まず、2 章で関連する研究について述べる。3 章で過去の体験によって得られた情報の中から、本手法の対象とする情報（本論文では『体験獲得情報』と呼ぶ）を整理する。また体験獲得情報をユーザに想起させるための行動検索手法の概要について述べる。4 章で、提案する行動検索手法の詳細を述べる。具体的には、行動履歴を自動的に収集するモジュール、再検索キーワードに関して集中して行動を起こしていた期間を指す『集中期間』と 1 つのウィンドウが 1 つの Web ページをアクティブに表示していた期間を指す『アクティブ期間』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、および行動を検索するインタフェースを提案する。5 章でプロトタイプ（Memory-Retriever）を用いた実験により、提案手法の有効性を議論し、最後にまとめを述べる。

2. 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。MyLifeBits³⁾ では、メタデータをファイルに付加する簡単な手法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良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⁾。SmartBack¹³⁾ 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など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。Retrospector¹⁴⁾ はコンピュータ上の行動履歴を自動収集し、時系列に画面キャプチャを表示/再生する。TimeMachineComputing¹⁵⁾ は、過去のデスクトップの状態を記録し、過去のデスクトップの状態に移動する機能を提供している。俺デスク¹⁶⁾ では、ユーザの行動履歴から Web ページやファイルごとの着目度を算出し、着目度を縦軸、時間を横軸にして過去に閲覧した Web ページなどのサムネイルを表示する手法を提案している。増井ら¹⁷⁾ は、ファイルの更新日時、デレトリ構造、内容など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-organizer¹⁸⁾ 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、行動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、行動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、本手法は行動履歴自体を検索するという点でこれらと異なる。

3. 体験獲得情報を想起させる行動検索

本章では、過去の Web ページの閲覧体験で得られた情報の中から本手法の対象とする情報を整理する。次に、それらの情報をユーザに想起させるための行動検索の概要について述べる。

3.1 体験獲得情報

Web ページの閲覧体験で得られた情報の 1 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、たとえばユーザが何かを調べるとき、1 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、ユーザがたどった 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な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、1 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。我々はコンピュータ上での 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とき、ある 1 つの Web ページの内容や 1 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。たとえばある商品の重さが知りたいときは、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。たと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もと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

3.2 行動検索

過去の体験獲得情報を想起させる方法として、体験獲得情報のメモをユーザに随時記述させる、体験獲得情報を直接推測する、などが考えられる。しかし、メ

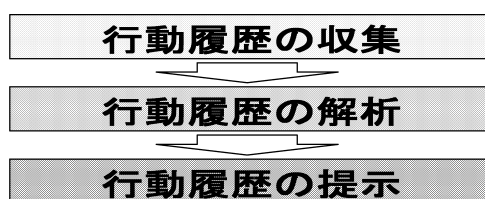


図 1 行動検索のステップ

Fig. 1 Steps for using action history.

モを随時記述させる方法ではユーザに負荷がかかり、体験獲得情報を直接抽出する方法は、抽出精度の向上が容易ではないと思われる。

そこで我々は、ユーザに体験獲得情報を想起させるために、過去の行動を検索する手法を提案する。つまり、体験獲得情報を直接推測するのではなく、行動検索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。

行動を検索するためには、行動履歴を収集し、行動履歴データを解析/抽出して、ユーザに効果的なインタフェースで提示する必要がある(図 1)。次章では、提案する行動検索手法について詳細に述べる。

4. 提案手法

本章では、まず、行動を検索するための元データとなる行動履歴を自動的に収集する手法について述べる。次に、提案する行動検索手法の特徴である『行動期間』について述べ、具体的な行動期間である『アクティブ期間』と『集中期間』の抽出方法、および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方法を述べる。また、集中期間やアクティブ期間を活用し、効率的に体験獲得情報を想起させる行動検索インタフェースについて述べる。

4.1 行動履歴の自動収集

行動履歴を解析/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行動履歴が必要である。我々はコンピュータ上の詳細な行動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てきた(図 2)^{12),22)}。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履歴の取得を可能にした。具体的には、マウス、キーボード、ウィンドウ状態、コピー、印刷などの行動履歴を収集する。また、アプリケーションごとに特化した詳細な行動履歴、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

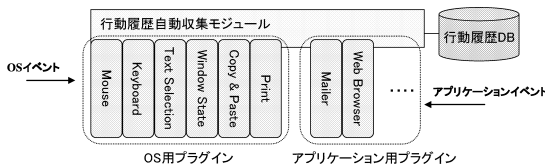


図 2 行動履歴自動収集モジュール

Fig. 2 Automatic logging module of action history.

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ごと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などを取得する。行動履歴は自動的に取得/保存され、ユーザは行動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、閲覧したほとんどすべての 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、行動履歴データを暗号化し、かつローカルのコンピュータ内のみに保存することとした。

4.2 行動期間に着目した行動履歴の解析

体験獲得情報を効率良く想起させるために、時間的に連続した行動を 1 つのグループとして抽出し、理解することは良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 1 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 ページではなく、ユーザが過去に行動を起こした期間（『行動期間』と呼ぶ）に着目する。本節では、具体的な行動期間として、1 つのウィンドウが 1 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した行動を集中して起こした期間（『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述べる。

4.2.1 アクティブ期間

本項では、行動をグループ化する最小期間として『アクティブ期間』を定義し、ユーザの注目の強さを表すアクティブ期間の注目度、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を算出する手法について述べる。

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 ページが見つかったら、Web ブラウザ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。

そこで、1 つのウィンドウが 1 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、たとえば、Web ブラウザに表示している Web ページを変更するたびに、新しいアクティブ期間が発生する。

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、様々な行動を起こす。そこで、ユーザが多くの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表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け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た（式 (1)）。さらに、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とを乗ずることで、式 (2) のように求めた。

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

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

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

$Att(ap)$: 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 中表示されている 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

表 1 HTML タグによる重み
Table 1 Weight based on HTML tag.

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

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 IF-IDF 値⁷⁾を用いた (式 (3))。 tf 値は、HTML タグによる重み付けを行った (表 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 1 日ごとなど定期的、あるいはユーザ指定時に最新状態に更新した。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、文字列選択を考慮した。

4.2.2 集中期間

本項では、任意の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 (集中期間) を抽出し、再検索キーワードに対するそれぞれの集中期間の重要度を算出する手法についてのべる。

ある再検索キーワードに関連する Web ページを集中して閲覧していた場合、閲覧した多くの Web ページに再検索キーワードが含まれていたと考えられる。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 ページの閲覧に逸れ、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

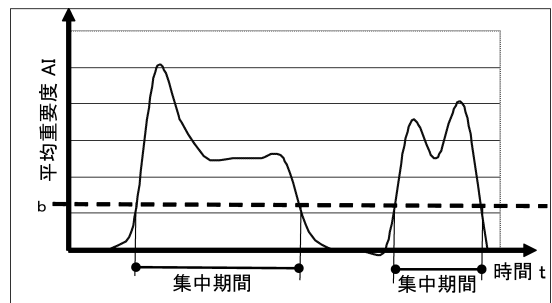


図 3 集中期間の抽出

Fig. 3 Extraction of intensive periods.

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 (式 (4))。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見なす (図 3)。いい換えると、継続判定関数 $B(k, t)$ が 1 以上である連続した期間を、1 つの集中期間として抽出する (式 (5))。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 1 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見なすことができる。たとえば $(a, b) = (300, 0)$ 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見なすことになる。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、1 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 1 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。

$$I(k, t) = IAP(k, ap) / (ap_{et} - ap_{st}) \quad (4)$$

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

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

$apst$: 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 に対する重要度を式 (6) 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

4.2.3 Web ページの重要度

本項では、指定した期間における再検索キーワードに対する Web ページの重要度を算出する手法について述べる．この指標は次節に述べる行動検索インタフェースで利用される．

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した (式 (7))．

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．

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

$IW(k, url, st, et)$: 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

4.3 行動検索インタフェース

本節では、再検索キーワードに関連する集中期間を検索する集中期間検索インタフェース、集中期間内の行動を詳細に提示する行動シーケンス表示インタフェース、および Web ページ検索インタフェースについて述べる．

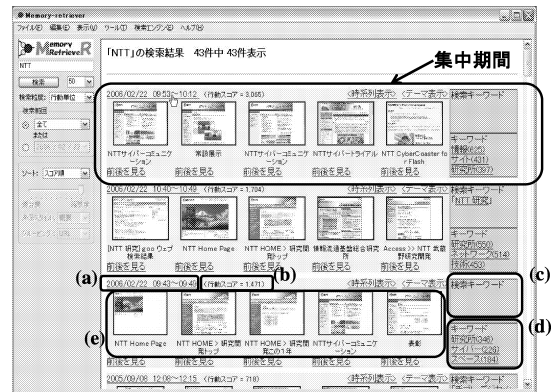


図 4 集中期間検索インタフェース

Fig. 4 Interface for retrieving intensive periods.

4.3.1 集中期間検索インタフェース

図 4 に集中期間を検索するインタフェースを示す．ユーザはまずこのインタフェースを利用することを想定している．

ユーザが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、集中期間が表示される．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が提示され、どのような行動をした集中期間かを理解できるようになっている．具体的には、集中期間の開始時刻と終了時刻 (図 4(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フェース (図 5、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 4(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 4(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 4(d)) が表示される．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が表示される (図 4(e))．

本インタ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する多くの行動が行われている．よって、次項で説明する行動シーケンス表示インタ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．

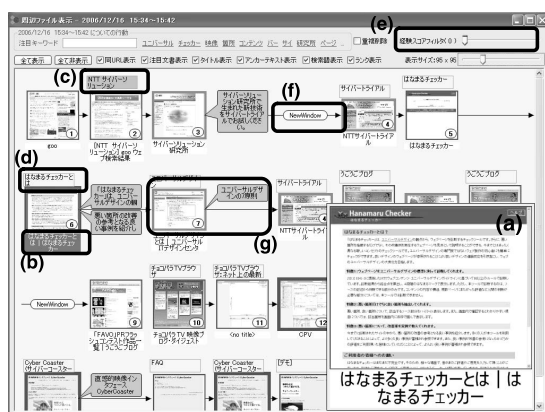


図 5 行動シーケンス表示インタフェース

Fig. 5 Interface for presenting action sequence.

4.3.2 行動シーケンス表示インタフェース

図 5 に行動シーケンス表示インタフェースを示す。集中期間検索インタフェースで選択された集中期間内の詳細な行動を提示する。

具体的に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 5(a))。サムネイルを右クリックするとさらに画面いっぱい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。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 5(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 5(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 5(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 ページはこの詳細ページの商品を推薦しているといった 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 つは赤色、次の 5 つは橙色の枠によって強調表示される。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 5(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 ごとの番号を提示し、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 5(f))、ウィンドウを切り替えた場合は、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。

ここで、本を読んでいるときに重要な箇所に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 5(g)) は、選択したテキスト情報を表している。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 6)。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。

4.3.3 Web ページ検索インタフェース

図 7 に Web ページ検索インタフェースを示す。本インタフェースは、集中期間やアクティブ期間ではなく Web ページを検索する。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。

ユーザが再検索キーワードを入力すると、行動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。さらに、閲覧回数、閲覧時間、コピーや印刷



図 6 注目箇所表示

Fig. 6 Highlighting sentences that the user previously selected.



図 7 Web ページ検索インターフェース

Fig. 7 Interface for retrieving Web pages.

したかどうかなどの行動履歴を表示する(図 7(a)). Web ページごとに付加されたリンク(図 7(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ーフェースで表示する.

5. 評価

提案手法により, 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した. また, 重要な体験獲得情報が短期間で多く想起できているか分析した.

5.1 評価対象

本実験では, 行動シーケンス表示インターフェースおよび注目箇所表示が, 体験獲得情報を想起させているかを定量的に評価した.

想起させる体験獲得情報として, Web ページから



図 8 従来手法の画面イメージ

Fig. 8 Interface of a conventional method.

表 2 被験者の属性

Table 2 Details about subjects.

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

得た知識を選択した. これは, Web ページの閲覧によってユーザが最終的に得た体験獲得情報であり, 重要性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ーフェースの有効性について, 定性的に分析した.

5.2 実験条件

(1) 比較対象

本実験の比較対象として, Web ページを単純に時系列に表示する手法(従来手法)と比較した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 8).

(2) 被験者

インターネットでの情報検索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した. 被験者の属性を表 2 に示す.

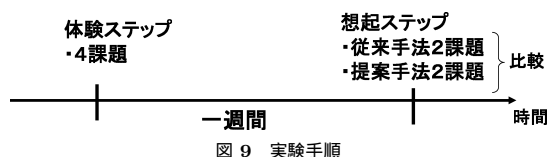
(3) アクティブ期間のパラメータ

アクティブ期間の注目度を算出するため, 行動種類ごとの重み付け係数 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

表 3 注目度算出パラメータ

Table 3 Value of weighting factor to calculate user's attention.

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

図 9 実験手順
Fig. 9 Procedure of the experiment.

いうように、ヒューリスティックに決定した(表 3)。

5.3 実験手順

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 段階で行った(図 9)。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した。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。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。ユーザの行動履歴は行動履歴自動収集モジュ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 にタスク例、被験者の記述例、および 1 つの課題に対して 1 人の被験者が記述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 1 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(つまり 30 分間)の行動を、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する。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無いようにした。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

表 4 タスク例および被験者の記述例

Table 4 Examples of given tasks and reported knowledge.

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は困難
・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)

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。

5.4 実験結果

どれだけ体験獲得情報を想起できたかを測定するため、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ない体験獲得情報の 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

それぞれの手法について、全被験者の獲得スコアの平均を図 10 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分かる。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 11 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あたりの獲得スコアの平均を示す。この表において提案手法は従来手法に比べ、想起開始から 7 分までにおいて特に高い値を示している。

図 12 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

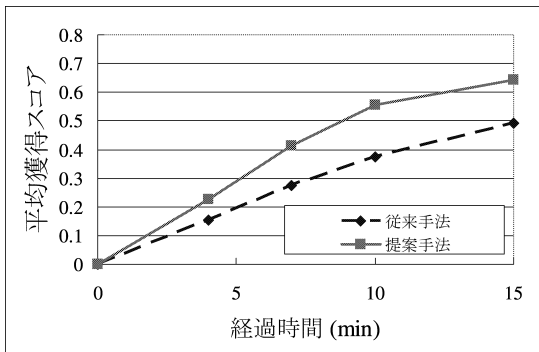


図 10 全被験者の平均獲得スコア

Fig. 10 Average score of conventional and proposed methods.

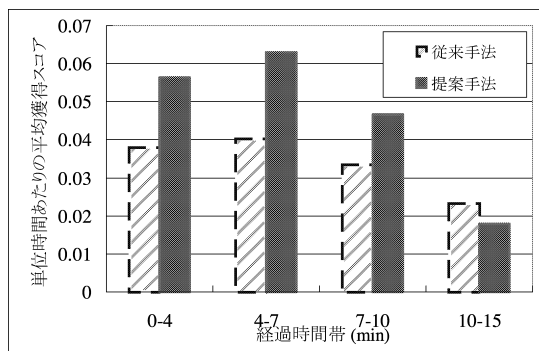


図 11 全被験者の単位時間当たりの平均獲得スコア

Fig. 11 Average score per minute.

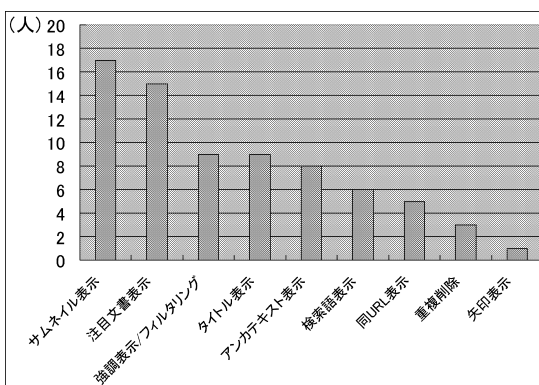


図 12 アンケート結果 (役立つ機能)

Fig. 12 Result of questionnaire (useful function).

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。

5.5 考 察

全被験者の獲得スコアの平均 (図 10) により、サーチ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分かった。

全被験者の単位時間あたりの平均獲得スコア (図 11) 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分かった。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 30 分の期間を費やすとは考えられない。より短い間に効率良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 1 つで拡大表示できる機能は有効であった。多くの Web ページを短時間で理解させ、また、どのような順で Web ページを閲覧したかを短時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時間で発見できた。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。短時間で知識を得た 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ないページの

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な情報であったからではないか。

6. 結論および今後の予定

本論文では、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 ページではなく、行動した期間を抽出することで、過去の行動を検索する手法を提案した。詳細には、行動を検索するための元データとなる行動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 1 つのウィンドウがある Web ページをアクティブに表示している期間を抽出し、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方法などを提案した。また行動期間を用いてユーザの行動を検索する行動検索インタフェースを提案した。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などを表示する手法は、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。

今後は、集中期間の抽出手法、集中期間検索インタフェース、および Web ページ検索インタフェースに対する評価実験を行う予定である。また、集中期間の抽出、およびアクティブ期間の注目度算出のための最適なパラメータ値について、ユーザの利用状況に応じた学習やパラメータを調整できるユーザインタフェースなどを検討していく。過去の複数の時点で体験ステップを実施することで、本手法の有効性を評価することは興味深い課題である。また、本論文での提案手法の対象は Web ページのみだが、行動履歴自動収集モジュールはメールや画像ファイルを利用した履歴も収集している。集中期間およびアクティブ期間を用いた検索を Web ページ以外のコンテンツを利用した期間にも適用可能にする予定である。

参 考 文 献

- 1) 平成 17 年度通信利用動向調査，総務省 (2006).

- 2) Chirita, P., Ghita, R., Gavrilaoie, A., Nejdl, S.W. and Raluca, P.: Activity Based Metadata for Semantic Desktop Search, *2nd European Semantic Web Conference (ESWC'05)* (2005).
- 3) Gemmell, G., Bell, G., Lueder, R., Drucker, S. and Wong, C.: Mylifebits: Fulfilling the memex vision, *Proc. ACM Multimedia'02*, 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) Milic-Frayling, N., Jones, R., Rodden, K., Smyth, G., Blackwell, A. and Sommerer, R.: 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) Rekimoto, J.: 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) Murakami, H. and Hirata, T.: 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: 個人の選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) 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).

(平成 18 年 6 月 22 日受付)

(平成 18 年 12 月 7 日採録)



森田 哲之 (正会員)

1998 年京都大学大学院工学研究科修了。同年 NTT ソフトウェア研究所入所。e-commerce システム, ブロードバンド CRM システムの研究開発に従事。現在, NTT サイバーソリューション研究所において, プロファイリング技術および情報検索の研究に従事。



倉 恒子

1988 年名古屋大学理学部卒業。同年日本電信電話株式会社 (現 NTT) 入社。現在, NTT サービスインテグレーション基盤研究所に勤務。電子情報通信学会員。



日高 哲雄 (正会員)

1993 年慶應義塾大学理工学部計測工学科卒業。1995 年同大学院理工学研究科計測工学専攻修士課程修了。同年日本電信電話株式会社 (現 NTT) に入社。2004 年より NTT サイバーソリューション研究所。情報可視化, 情報検索の研究に従事。電子情報通信学会各会員。



大浦 啓一郎

1999 年広島大学大学院工学研究科修了。同年 NTT 入社。現在, NTT 西日本技術部研究開発センタに勤務。電子情報通信学会員。



田中 明通

1987 年東京大学大学院工学系研究科精密機械工学専攻修士課程修了。同年日本電信電話株式会社入社。現在, サイバーソリューション研究所所属。パターン認識, 人工知能, 学習科学, 次世代ポータルの研究開発に従事。電子情報通信学会, IEEE 各会員。



加藤 泰久

1988 年京都大学工学部数理工学科卒業。1990 年同大学大学院工学研究科応用システム科学専攻修士課程修了。同年日本電信電話株式会社入社。音声応用システム, 教育システム, 検索システム等の研究開発に従事。人工知能学会会員。



奥 雅博

1984 年大阪府立大学大学院工学研究科博士前期課程修了。同年日本電信電話公社 (現 NTT) に入社し, 日英機械翻訳システム, オペレータレス電話番号案内システム等における自然言語処理技術, 特に日本語処理技術の研究開発に従事。新事業開発の企画, 光マーケットクリエーション活動等を経て, 現在 NTT サイバーソリューション研究所において, 検索をはじめとするブロードバンドインターネットサービスに関する研究開発に従事。実験サイト goo ラボ (<http://labs.goo.ne.jp/>) で様々な技術を公開実験中。博士 (工学)。電子情報通信学会, 言語処理学会各会員。