

# 俺デスク:ユーザ操作に基づく参照履歴検索ソフトウェア

今枝 卓也<sup>1</sup> 鈴木 慧<sup>1</sup> 大澤 亮<sup>2</sup> 神武 直彦<sup>2</sup> 高汐 一紀<sup>1</sup> 徳田 英幸<sup>1,2</sup>

<sup>1</sup> 慶應義塾大学環境情報学部 <sup>2</sup> 慶應義塾大学大学院政策・メディア研究科<sup>i</sup>

## 1 はじめに

PCユーザは、PC上で文書やWebページなどのデータを日々参照する。ユーザがPC上で過去に参照したデータを見つける場合、履歴一覧を参照しデータを検索できる。しかし、履歴の量が膨大になると特定のデータを探し出すのは困難になる。例えば、OSの中には最近参照したデータの履歴を一覧表示できるものがある。しかし、ファイル名やアイコンなど限られた情報しか表示されず、効率よくデータを見つけられない。この問題を解決するために、本研究はユーザに特別な手動操作を要求しない手法で、履歴の検索を効率化する。本研究では、PC上でのユーザ操作を基に過去に参照していたデータの履歴を検索できるソフトウェア、「俺デスク」を提案する。俺デスクはデータ着目度とデータ間関連度を算出する。

データ着目度とはユーザのデータに対する注目度合いを示した指標である。データ間関連度とはデータ間の関連の強さを表す指標である。そして俺デスクはそれぞれを利用したタイムラインビューアと関連検索ツールをユーザに提供する。タイムラインビューアは、ユーザのデータ参照状況を視覚化する。関連検索ツールとは、あるデータと関連のあるデータを検索できるツールである。想定されるシナリオとして、ユーザはタイムラインビューアに対し、週末に編集していた文書をもう一度参照したいという要求ができる。さらに関連検索ツールに対し、その文書と同時刻に参照していたWebページを検索といった要求ができる。

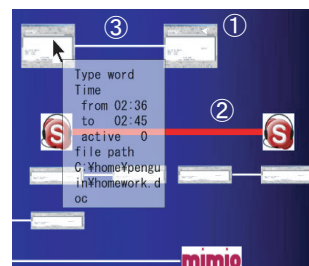
## 2 ユーザ操作履歴に基づく情報検索ソフトウェア

### 2.1 タイムラインビューア

タイムラインビューアを図1に示す。画面上では、ユーザが実際に参照したWebページやドキュ

メントのサムネイルが表示される。横軸に時間を取り、データ参照の開始時間と終了時間が白いラインで結ばれ、実際に画面の前面に来てアクティブになっていた時間が赤いラインで結ばれる。ユーザはカレンダーから日付を指定し、その日参照していたデータを時系列に沿って検索できる。また、縦軸は着目度になっており、着目度が高いデータほど上位に表示される。これを利用して、ユーザが過去に深く参照していたと思ったデータを探す場合は上位を探すという使い方ができる。

着目度は、データの参照時間や参照回数、選択文字列の反転などの要素から算出され、これらが多いほど着目度は高くなる。また俺デスクは付加機能として、各要素の重みを動的に調整する機能を持つ。各要素の重みを静的に決定してしまうと、ユーザにとっての着目度と一致しない可能性がある。そこで、俺デスクはデータに対する着目度をユーザからのフィードバックとして得る。この着目度とユーザ操作履歴との相関を求め、この相関から、データ着目度を求める際の各要素の重みを決定する。



- 1: サムネイル
- 2: アクティブ時間
- 3: オープン時間

図1: タイムラインビューア

### 2.2 関連検索ツール

関連検索ツールを図2に示す。ユーザはファイル名や文中のキーワードなどから目的のデータを検索できる。図2の画面上では、検索結果が検索対象データと関連の強い順に表示される。

<sup>i</sup> Ore Desk, Takuya Imaeda, Kei Suzuk, Ryo Oh-sawa, Naohiko Kohtake, Kazunori Takashio, Hideyuki Tokuda, Faculty of Environmental Information Keio University, Graduate School of Media and Governance Keio University

データ間関連度は参照時刻，テキスト内検索，クリップボードの利用履歴などから算出し，これらが関連しあっているデータ同士ほど関連度が高くなる．参照時刻の場合，近い時刻に参照していたデータは関連が深いとする．具体的には，ユーザがデータ A にアクセスした時刻が

$$(a_1, a_2, \dots, a_n) \quad (n \text{ は正の整数})$$

であり，データ B にアクセスした時刻が

$$(b_1, b_2, \dots, b_m) \quad (m \text{ は正の整数})$$

のとき，データ A, B 間の関連度  $R_{ab}$  は

$$R_{ab} = \frac{\sum_{k=1}^n \min(|a_k - b_1|, |a_k - b_2|, \dots, |a_k - b_m|)}{n}$$

で算出する．この式により，A のアクセス時刻と，それに最も近い B のアクセス時刻との差の平均が取れる．従って，同時刻に参照していたデータほど，関連度が高くできる．また，同語句を検索したデータ同士は関連が深いとする．さらに，データ A でクリップボードにコピーし，データ B にペーストした場合，データ A と B は関連が深いと補正する．これら複数の要素を組み合わせると，最終的な関連度を算出する．

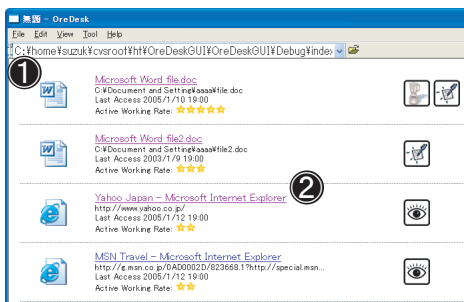


図 2: 関連検索ツール

### 3 実装環境

俺デスクの実装環境を表 1 にまとめる．俺デスクは Windows 上であればバイナリファイルのインストールのみで動作する．

今回のデモンストレーションでは，来場者に表 1 に書かれたアプリケーションで実際に作業してもらおう．その上で，作業に使用したデータがタイムラインビューアと関連検索ツールにより検索でき

ることを体験してもらおう．また，古い履歴も検索できることを示すために，想定されたシナリオに応じた作業を事前に行っておき，そのときに使ったデータを検索できることを示す．

表 1: 実装環境

|                     |  |
|---------------------|--|
| 対応 OS               | Windows 2000/XP  |
| 実装言語                | C++  |
| イベント監視対象 (アプリケーション) | Microsoft Office, OpenOffice, Internet Explorer, Firefox, Skype, MSN Messenger |
| イベント監視対象 (OS)       | ウィンドウフォーカス, クリップボード, キーボード, マウス, ファイルアクセス                                      |
| 使用ライブラリ             | Google Desktop SDK   |

### 4 関連研究

履歴情報を検索するツールとして暦本氏の Time-Machine Computing[1] が挙げられる．Time-Machine Computing はデスクトップにおかれたファイルの生成から削除までの生存期間を時系列で表示する．これに対し，俺デスクは，ファイルが開かれていた時間を記録している点や，データ間関連度やデータ着目度を用いた検索が可能になっている点で異なっている．

PC 上のイベントを保存していく手法として，近藤氏による NecoLogger[2] がある．NecoLogger は Windows 上の様々なイベントを記録していき，過去の情報を記録から取り出すソフトウェアである．俺デスクが NecoLogger と異なる点として，データ着目度やデータ間関連度といった独自の指標を設けている点や，アプリケーション固有のイベントを取得できる点が挙げられる．

### 5 まとめ

本研究では，PC 上で行ったユーザ操作を基に，ユーザが過去に参照したデータの履歴を検索できるソフトウェア，俺デスクを提案した．本研究は IPA 情報処理推進機構未踏ソフトウェア創造事業未踏コースの下で行われている．

#### 参考文献

- [1] J. Rekimoto. Time-machine computing: a time-centric approach for the information environment. In *UIST '99*, 1999.
- [2] 近藤, 三宅. 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. *WISS 2005*.