

ユーザ動作を基にしたデータ間関連度とデータ着目度算出機構 — 履歴情報検索ツール「俺デスク」 —

1. 背景

PC ユーザは文書や画像、映像、音楽などのデータを日々PC 上で参照するが、過去に参照したデータを再度参照するのは思ったより困難である。OS やブラウザには過去に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。

2. 目的

本ソフトウェアの目的は、ユーザに普段特別な手動操作を要求しない手法で、ユーザの履歴データ検索を効率化することである。そこでローカルディスクに保存された文書や画像、映像、音楽などのデータや Web 上にあるデータを検索可能にするソフトウェアを実装した。

3. 開発の内容

本プロジェクトではユーザが Windows 上で行った操作履歴を保持し、それらの履歴からデータを想起可能にするソフトウェア、俺デスクを実装した。俺デスクはデータ間関連度とデータ着目度を算出し、関連検索ツールとタイムラインビューアをユーザに提供する。

3. 1 関連検索ツール



1. 検索フォーム, 2. 関連データ一覧, 3. 属性アイコン

図 1

関連検索ツールを図1に示す。関連検索ツールとはあるデータに対して関連のあるデータを検索できるツールである。俺デスクはユーザ操作の履歴情報を基にデータ間関連度を算出する。具体的には参照時刻、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。ユーザ操作を基にした関連度を用いることで、検索システムはユーザに対して参照記憶を基にした検索を提供できる。例えば、「Word を使って企画書を作成していたときに参照していた Web ページを検索」といった要求ができる。

また、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。キーワードとしてはアクセス日時、ユーザの特定操作を入力できる。また、Google Desktop SDK を利用し、ファイル内に含まれている文字列も並列して、検索にかけることができる。具体的には「ファイル A に関連があって、地図という語句が含まれているファイルを検索」といった要求が可能になる。この機能は Google Desktop 本体がインストールされている PC のみ有効になる。属性アイコンとはそれぞれのデータが、過去のユーザ操作から、どのような特徴をもっているかを示すアイコンである。俺デスクでは過去のユーザ操作から属性アイコンを自動付与する。

3. 2 タイムラインビューア



図 2

タイムラインビューアを図 2 に示す。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したものである。ユーザはカレンダーから日付と日付幅を指定し(図 2 の左)、指定した期間のタイムラインから目的のデータを探す(図 2 の右)。タイムラインビューアは、横軸に時間を取り、データ参照時間をラインで表示する。白いラインはデータのあるアプリケーションで開いていた時間、赤いラインは実際に前面に来てアクティブになっていた時間、その中央のサムネイルはユーザが実際に参照した Web ページやドキュメントを表す。タイムラインビューアを使うことで「昨日の夜中に参照していたレストランガイドをもう一回見たい」という探し方が可能になる。縦軸はデータ着目度を表しており、データ着目度が高いデータほど上位に表示される。データ着目度とはユーザのデータに対する注目の度合いを示した指標である。データ着目度は、データの参照時間や参照回数、選択文字列の反転回数などの計測結果から算出され、これらが大きいほど着目度は高くなる。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。

3. 3 動作環境

表 1

対応 OS	Windows 2000/XP
実装言語	Visual C++
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, マウス, キーボード, ファイルアクセス, スクリーンセーバ
イベント監視対象 (アプリケーション)	オフィス (Open Office, MS Office), ブラウザ (Firefox, Internet Explorer), メッセンジャ (MSN Messenger), VoIP (Skype)
使用ライブラリ	SQLite, Google Desktop SDK

俺デスクの動作環境を表 1 に示す。俺デスクは詳細なアプリケーションのイベントを監視するために、アプリケーションごとにプラグインを提供する。

4. 従来の技術（または機能）との相違

既存の検索手法にはデータ内容を解析し、データの分類や重み付けをする手法が多いが、本ソフトウェアではユーザ操作の履歴情報を基にデータの様々な分析を行っている点で独自性がある。また、ユーザ操作を基にデータ着目度とデータ間関連度という独自の指標を算出している。これら 2 つの指標を用いて、ユーザがより効果的に履歴情報を検索できるようなユーザインターフェースを提供している。

5. 期待される効果

2004 年後半に Google Desktop の β 版がリリースされて以降各ベンダがデスクトップ検索をリリースしてきた。そして現在は Google パーソナライズドホームに代表されるように個人に合わせ、検索結果が変化する実装が進められている。このような背景において、本ソフトウェアは検索の個人カスタマイズ分野においてユニークな手段としてユーザに選択肢を提供できると考えている。

6. 普及（または活用）の見通し

Web サイトにてソフトウェアの配布を行う。そして、フィードバックを受け開発を続けていく。今後は学会やシンポジウムなどを通して、より多くの人に使ってもらえるようにする。

7. 開発者名（所属）

大澤 亮（慶應義塾大学大学院政策・メディア研究科）

（参考）<http://oredesk.net/>