

PCにおけるタスク想起のためのユーザ操作可視化法の提案

松本光弘^{†1} 岡野真一^{†2} 森田哲郎^{†2}
沼尾正行^{†3} 栗原聡^{†3,†4}

近年、個人レベルで大容量の記憶媒体が容易に使えるようになりつつあり、PC上には多くのファイルが蓄積されていく一方、有効に利用されていない。しかし、多くのタスクがPCにて行われる状況では、タスクに依存してファイル間に依存関係があったり、過去のファイルが現在のタスクにおいて重要である場合があるなど、すべてのファイルにはそのファイルに関するコンテキストが存在するはずである。そこで、ファイルに対する一連の操作などの時系列履歴やファイルとWebページ間の類似関係をネットワークとして抽出・可視化してユーザに提示することが、ユーザにとってその時のコンテキストを思い出し、検索対象のファイル等を従来の検索手法に比べて容易に見つけることが可能となることを示す。

Visualization Method of User's File Operation for Recalling Task on Desktop Environment

MITSUHIRO MATSUMOTO,^{†1} SHINICHI OKANO,^{†2}
TETSUO MORITA/AFFIREFSUMITOMO,
MASAYUKI NUMAO^{†3} and SATOSHI KURIHARA^{†3,†4}

Users can use the huge storage of personal computers at a moderate price in late years, and they tend to store a large number of files on their computers but cannot make efficient use of them. On the other hand, when a task is carried out on the computer, some files related to the task may be generated or some old files may be important for the task. Therefore, all files must be associated with a task and some contexts corresponding to a file must exist. Hence, we will show that visualization of "file event network", which is made by both file access or web access history and analogous relationship with files or web pages, allows to recall the contexts according to a file and to retrieve the desired file or web page easier than the traditional approach.

1. はじめに

インターネットの発展に伴うクラウドコンピューティングなどの登場で、もはやパソコン(PC)は単体での存在であると同時にネットワーク端末としての存在でもあり、より多様性のある使い方がされるようになった。仕事からプライベートに至る多種多様なタスク、例えば報告書の作成から日記の執筆、家計簿の管理に至る様々なタスクがPCにて行われるようになり、タスクの処理に応じて生成されるファイルの種類も多様化しつつある。

一方、近年のPCの高性能化と、HDD、SSDなどの二次記憶装置の大容量化・低価格化により、個人レベルで大容量の記憶媒体が容易に使えるようになった。記憶容量が少ない時は必要なファイル以外は消去するしかなかったが、記憶容量が十分な状況となれば無理に消去する必要はなく、結果としてPC上には様々なファイルがひたすら蓄積されていく傾向にある。

通常我々は、あるタスクを処理している間はそのタスクに関するコンテキスト、すなわち、作成中のファイルがどのファイルを参照しているとか、どのファイルのどの部分を編集したなど、そのコンテキストに関わる自分の行動履歴を把握できている。よって、バックアップのために類似したファイル名にてファイルを保存したり、一時的なフォルダを作成し、そこに一時的に重要なファイルを保存していても迷うことはない。

しかし、タスクが終了して時間が経過してしまえば、ほとんどの場合そのタスクに関するコンテキストは忘れられてしまう。例えば、我々はいきなり過去に生成されたファイルを見ても、そのファイルがいつどのタスクにおいて生成され、その生成過程においてどのファイルを参照したとか、どのファイルの一部を編集したとかなどをなかなか思い出すことはできない。しかも、我々がPCで行う作業を見つめ直してみると、新規にゼロから文書の作成を開始し、しかも完成するまで一切他のファイルなどを参照しないような作業は希であり、通常は、過去に作成したいいくつかの文書やウェブページを参照したり、過去の文書の部分的な更新であったりと、多くのタスクは過去のタスクに何かしら関連している。

^{†1} 大阪大学大学院情報科学研究科

Graduate School of Information Science and Technology, Osaka University

^{†2} 住友電気工業株式会社 情報通信研究所

Information and Communications Labs, Sumitomo Electric Industries, Ltd

^{†3} 大阪大学産業科学研究所

The Institute of and Industrial Research, Osaka University

^{†4} JST CREST

よって、我々が新しいタスクへの取り組みを開始する時には、何らかの方法で過去のタスクを参照する必要性が発生するのであるが、現在我々ができる手段はファイル名やキーワードに基づくファイル単位の検索であり、基本的なしくみはインターネット検索と同じである。しかしながら、キーワード検索だけでは十分な検索を行うことが出来ないことが知られている¹⁾。1) では、メールやローカルファイルや Web 検索において、人々がどのような検索を行うかの検証を行っており、特定の情報においては 42%、一般の情報においては 55% がキーワードによって検索されたものであり、残りは別の手段を用いて検索されている。例えば、ある教授室の電話番号を調べる際には、「教授名、電話番号」というキーワードで調べるのではなく、学科のページから該当教授のページに行き、そこから電話番号を得る。これは、キーワードを使って目的のページに飛ぶのではなく、過去に行なったプロセスを思い返しながら、小さなステップを経て目的の情報にたどり着いたひとつの例であり、1) では小さなステップを経ることで徐々に情報を思い出していく過程が検索において重要であることを示している。また、ファイル名での検索は、同一ファイル名や類似するファイル名であるファイルが複数検出されることがあり、一つ一つのファイルがそれぞれどのタスクに関連したものであるかまでは特定できず、日付や内容、保存されているディレクトリなどをいちいち確認しなければ所望のファイルを探すことはできない。

仮にタスク同士が互いに独立した関係で個々のタスクが簡潔なものであれば、このような方法であってもファイルを特定することは比較的容易かもしれない。しかし、実際には、タスク自体も複雑であり、タスクを完了するまでに様々なファイルが生成されていたり、さらに異なるタスク間でファイルを共有していたり、あるタスクに関連するファイルが別のタスクのファイルを参照している場合や、異なるタスクで使用されたにも関わらず、類似した名前を持つファイルが複数存在していたりと、ファイルのみに着目する検索ではなかなか該当するタスクに関連するファイルを見つけることは難しい。しかも、上述するように PC に保存されているファイル数は増加しつつある。

この状況において、もしも個々のファイルに関連するコンテキストを思い出すことができれば、個々のファイルとそれに対応するタスクを結びつけることができ、参照したいタスクとそのタスクに関わる目的とするファイルを見つけることが容易になると期待される。ここで、コンテキストについて考えてみると、コンテキストはタスクを処理する際の一連の行為そのものであり、中でも、ファイルの生成・編集・コピーなど、タスクを処理する過程でユーザが PC を操作した一連の履歴は、コンテキストにおける主要な要素であり、幸い、現在の PC では過去のファイルも消去されずに保存される傾向にある。

そこで、本論では単にファイル名で検索するのではなく、ファイルに対する一連の行為に関する時系列履歴を「ファイルイベントネットワーク」として抽出・可視化してユーザに提示することが、ユーザがその時のコンテキストを思い出し、参照したいタスクと、そのタスクに関連するファイルをより容易なものとすることを示す。

以下第 2 章において、関連研究における本研究の位置づけを述べ、第 3 章にて提案するファイルイベントネットワークを作成する方法について述べる。そして第 4 章では実装したシステムの動作検証と有用性について述べ、第 5 章にて、まとめと今後の課題を述べる。

2. 関連研究

現在の PC でのファイル検索は、上述するようにファイル名やキーワードに基づく手法が利用されており、GoogleDesktop^{*1}等のサービスでは、PC 内ファイルの各テキスト情報を高速に検索できるように予めインデックスを作成してファイルアクセスを支援している。ランキングを更に工夫する手法としては 2),3) などが提案されている。また、ファイルに関するコンテキストに基づいてキーワード検索ランキングを決定する研究も存在する⁴⁾。Windows Vista などの OS においてもインデックスを用いた高速な検索機能があらかじめ用意されている。このようなキーワード検索では適切なキーワードをユーザが選択する必要があり、文章や文書中の単語を覚えておかなければ、検索することはできない。さらに、内容が類似するファイルや編集された時刻などに着目した検出は可能であるが、お互いに内容が類似するファイル同士がどのような経緯で作成されたのか、という作業のコンテキストまでを思い出そうとすることはなかなか困難である。また、検索対象のファイルについて、ユーザが知っているあらゆる情報に対応するため、メタデータやタグ取り付け機能などを用いた、より高度な検索インタフェースの提案もされている⁵⁾。これについても、現状ではメタデータを人手にて付与する必要があり、常に付与されているとは限らず、またファイルの内容によっては的確なメタデータの付与が困難な場合も考えられる。

ファイルのテキスト情報以外の情報も利用することで適切なファイル検索を行う方法としては 6) が挙げられる。メールに添付されたファイルであれば、メールの送り主や件名との関係性を利用し、出版物であれば著者や会議名といったメタデータと Google や Citeseer によるランキングを基に適切なファイルのランキングを求める手法を提案している。また、メタデータにもファイルのコンテキストが含まれていることから、メタデータをオントロ

*1 Google Desktop, <http://desktop.google.com/en/>

ジーを用いてスコアリングすることによって、精度の高いファイル検索が可能であるという立場をとる研究として、セマンティックデスクトップ⁷⁾が挙げられる。しかし、上述するように、メタデータについては、人手によるオントロジーの構築が必要であるため、ユーザの負担が避けられない。たしかに、セマンティックデスクトップが実現すれば、ミーティングやプロジェクトといったカテゴリ別にファイル検索を行えるため、より効率よくファイル検索を行うことができると考えられるが、メタデータをユーザの負担なくして効率的に付与する仕組みの開発が必要不可欠である。これに対し、本研究ではメタレベルなどを付与することは行わないことからその実現が容易である。また、8) では、人々が何を手がかりに文書ファイルを探すのかを検証しており、ユーザはファイル名やファイルが格納されている場所はもちろんのこと、キーワードや利用時刻・文書全体の見た目といったさまざまな要因を思い出すことによって、ファイル検索を行なっていることが示されており、ファイルに関係するイベントは重要な要因であることが述べられている。

目的は異なるものの、本研究と同じく時間を手掛かりとして検索を行う方法も提案されている。メール管理では、受信メールを時刻順に並べて、日時からメールを特定するといった方法が有効である。Ringelらは公的、および私的なイベントを時系列インタフェースに表示することで、ユーザがイベントとの相対的な時間感覚を頼りに効率よくメールを特定できることを示した⁹⁾。大澤らは、ユーザのデータ参照時間や回数などから算出した着目度に基づいた時間軸インタフェースを実装している¹⁰⁾。また、ファイル検索とは目的が異なるが、暦本はコンピュータの作業履歴を蓄積し、時間移動によって過去の作業環境の再現を行うとともに、時間に伴うPC環境の遷移を視覚的に表した¹¹⁾。また、個人の経験に基づいたコンテンツ検索の研究として12)が挙げられる。個人的な経験を経験要素表現に変換し、コンテンツ間で経験要素がどれほど一致しているのかを用いてコンテンツ間にリンクを結ぶことで、そのリンク構造に基づいてコンテンツ検索結果のランク付けを行なっている。これを実現するためには、自動的な経験要素表現の抽出方法が必須である。

PCの利用履歴から個々のタスクを抽出する研究として13)がある。コピーや編集といった履歴を利用して一連の操作を可視化する手法を提案しているが、ファイル間の類似性は利用しておらず、「メールに添付された文書は編集され、再びメールに添付されて送信されることが多い」といった共通性のあるタスクのパターンを抽出することが目的であり、タスクの内容までを考慮したタスクの特定を行うことは出来ない。これらの研究は、ユーザの振る舞いに着目した検索手法を提案している点では本研究と同じ着眼点を持つと言えるが、これらが振る舞いのみに着目しているのに対して、本研究では、個々のファイルの中身についても

考慮する点が大きく異なる。

3. ファイルイベントネットワーク

コンテキストを構成する要素としては、ファイルの生成・編集・参照・コピーから、ファイルの圧縮・解凍・添付・アップロードやマウスの軌跡やメール・ウェブアクセスなど、様々な要素が考えられるが、今回は、それらの中で、コンテキストをより効果的に表現すると考えるファイルの生成・編集・参照・コピー、ウェブ閲覧に着目し、これらに関するユーザがPCを操作した一連の履歴を収集する。無論、他の振る舞い情報を利用することでより詳細なコンテキストの同定が可能となることが推測されるが、まずは、これらに着目してファイルイベントネットワークとして可視化することで、振る舞い情報の可視化がコンテキストの想起に有用であることを示す。

まず、ファイルイベントネットワークを、ファイル f と f がコピーや編集された時刻 t の対 (f, t) をノード V とし、そのファイルに対する操作である「コピー」「リネーム」「ファイルの更新」「ファイルの参照」をリンク E とする有効グラフと定義する。すると、時系列上にファイルイベントネットワークを可視化すると、図2のようなネットワークとして可視化される。例えば、時刻 t_1 に生成されたファイル f_1 が時刻 t_2 に編集され f_2 にリネームされたとすると、ノード (f_1, t_1) からノード (f_2, t_2) にリンクが張られる。また、3.1節にて後述するが、ファイルイベントネットワークの辺は重み付きであり、コピーやリネームはファイルの内容自体は同一であるが、参照関係にあるファイルや、一部がコピーされた関係にあるファイル同士などにおいては、お互いの類似性が高いほど重みが大きくなるように可視化する。加えて、同時刻にアクセスされるファイル同士は同一のタスクに関連する可能性が高いと考え、アクセスされた時刻間隔が小さいほど重みを大きくする。本論文では、ネットワークの辺の重みを時間的近さ \times ファイルの類似度としてネットワークの作成を行なった。

このように、個々のファイルに対するファイルイベントネットワークを構築することで、ファイルが作成されるまでの過程や、あるファイルが起点となって様々なファイルが生成された過程などを可視化することができる。例えば、あるファイル f_i から複数のファイルが生成されていれば、 f_i は、 f_i に関するタスクにおける根幹となったファイルであることを意味していることが分かる。

また、ある文書を作成する過程で、バックアップのためのコピーや内容を修正するための編集を複数回繰り返した経緯もファイルイベントネットワークとして可視化することが出来ることから、どのファイルが最新であり、いつ編集やコピーされたのかも容易に把握するこ

とができる。また、提案法では、ファイルの参照関係もファイルイベントネットワークとして可視化することから、あるファイルが作成された時に、どのファイルを参照したのかも容易に把握することができる。

先行研究として、我々はファイルイベントネットワークの情報を用いて、個人と複数人間のPCのコンテキストを描画する手法を提案している¹⁴⁾。先行研究では、1つのファイル(種ファイル)に対するイベントとそのファイルに関するファイルのイベントを時系列上に表示する手法を提案しているが、この手法では、種ファイルの作成過程を知ることは可能であるが、コピーやリネームによって作成された類似ファイルのイベント情報を同時に閲覧することができないため、最新バージョンのファイルや不要ファイルを検出するといったことが不可能であった。提案法では、コピーやリネームによって作成されたファイルを同一画面上に表示することで、上記のことを可能とし、さらに、ブラウジング機能により、関連ファイルのファイルイベントネットワークも同時に閲覧できるため、1つのタスクに関する個人のPCでの活動内容を詳細に閲覧することができる。

以下の小節でファイルイベントネットワークを構築する手順について述べる。

3.1 コンテキスト抽出

まずは、ユーザがファイルに関する一連の操作の履歴を収集する必要があるため、我々は ManagementCore^{*1}を利用している。実際に ManagementCore で収集されるログを表1に示す。

ファイルに関する事象は以下の5種類が存在し、ファイルに関するユーザの操作を取得できる。

- IO_FIL_MODIFY:ファイルの作成・編集
- IO_FIL_READ:ファイルの閲覧
- IO_FIL_COPY:ファイルのコピー
- IO_FIL_RENAME:ファイルのリネーム
- IO_FIL_DELETE:ファイルの削除

表1の一行目は、“test.txt”ファイルが作成または編集されたことを示しており、2行目は test.txt ファイルが“コピー ~ test.txt”にコピーされたことを示している。

また、ファイルのアクセスログとは別にウェブ閲覧のログを収集する必要があるため、firefox のアドオンを作成し、ウェブ閲覧の履歴を収集した。実際に得られたログを表2に示す。

表2 ウェブのアクセスログ

Table 2 web access log

日時	事象コード	URL	タイトル
2010-11-10-Wed-16-15-28	0	http://www.google.co.jp/	iGoogle - Mozilla Firefox
2010-11-11-Thr-16-25-53	1	http://www.yahoo.co.jp/	Yahoo! JAPAN - Mozilla Firefox

事象コードは以下の3種類が存在し、ウェブに関するユーザのアクセス履歴を取得することができる。

- 0:新規ページの読み込み
- 1:タブの切り替え
- 2:ウェブページを閉じる

次に、内容が類似するファイルやウェブページは同一のタスクに関するリソースとして参照関係にあたり、お互いに依存関係があるタスク同士に関するリソースであったりと、ファイル同士のコンテンツの類似性はファイルイベントネットワークを構築する際に必須であることから、PC上の全ファイルと閲覧したウェブページに対して、重要語の抽出を行う。MS-Office ファイルやPDFファイルのような特有のフォーマットファイルに対しても xdoc2txt などの、バイナリ文書からテキストを抽出するツールを用いてテキストを抽出した。形態素解析には Sen^{*2} を用い、これと松尾らの重要語抽出法¹⁵⁾ を用いて、重要語の抽出を行った。株に関係する文章から重要語を抽出すると、「株主、換金、いつ、性、最大、優待、期待、配当、売買」などの重要語を抽出することができる。これにより、ファイル・ウェブアクセスと重要語から、ファイルの操作過程と関連するリソースを抽出することで、ファイルのコンテキストを獲得する。コンテキストとは人がそのタスクを実行するに際して行った行動履歴であり、そのコンテキストを思い出すことで、所望のファイルやそのファイルに関するタスクを思い出すことができる。このことから、我々はコンテキストを想起する手がかりとして、ファイル・ウェブのアクセス履歴と、その内容の両方に着目する。

本論文で扱わなかった重要なファイルアクセスイベントとして、クリップボードのトレースが考えられる。これは、本論文で用いている ManagementCore がファイルのIOを監視するロガーであり、クリップボードを対象としていないからである。しかしながら、ファイルを作成する際に、ファイルの一部をコピーすることがあるため、クリップボードの情報はファイル作成過程において重要な情報である。クリップボードの情報を用いることは今後の

*1 IT 資産管理システム ManagementCore, <http://mcore.jp/>

*2 Sen, <http://www.mlab.im.dendai.ac.jp/~yamada/ir/MorphologicalAnalyzer/Sen.html>

表 1 ファイルのアクセスログ
Table 1 file access log

日時	事象	事象コード	プロセス ID	スレッド ID	読込サイズ	書込サイズ	ファイル名	先ファイル名	プロセス名
2010/11/16 17:42:29:062	IO_FIL_MODIFY	0x00000040	0x17ac	0x1834	0	6	C:\test.txt		NOTEPAD.EXE
2010/11/16 17:42:41:984	IO_FIL_COPY	0x00000020	0x688	0x1648	0	6	C:\test.txt	C:\コピー ~ test.txt	Mame5.exe

課題である。

3.2 リンクの生成

本論文の提案するファイルイベントネットワークは各リンクに重みのついたネットワークである。リンクの重みはファイル間の類似度とアクセス時刻の差によって決まる。

時刻 t_A にファイル a に対してアクセスイベント A が生じ、時刻 t_B にファイル b に対してアクセスイベント B が生じた場合のリンク強度の計算方法を以下に示す。ファイル間の類似度によるリンク強度の計算方法は以下の通りである。

$$L_{A \rightarrow B}^c = \frac{Keyword_a(t_A) \cap Keyword_b(t_B)}{Keyword_a(t_A)} \quad (1)$$

$L_{A \rightarrow B}^c$ は、ファイル b に対するファイル a の類似度を示す。 $Keyword_a(t_A)$ は時刻 t_A の時点でファイル a が持つ重要語数となる。 $Keyword_a(t_A) \cap Keyword_b(t_B)$ は、時刻 t_A におけるファイル a と時刻 t_B におけるファイル b に、共通して見られる重要語の数である。同様に、アクセス時刻におけるリンク強度の計算方法は以下の通りである。

$$L_{A \rightarrow B}^t = \begin{cases} 1 & |t_A - t_B| \leq T \\ 0 & \text{else} \end{cases} \quad (2)$$

$L_{A \rightarrow B}^t$ はファイルアクセス B に対するファイルアクセス A のアクセス時刻に関する関連度を示す。 T はアクセス時刻の差に関するパラメータで、 T 以内にアクセスされたファイル同士は何らかの関係があると見なされる。

ファイル間の類似度とファイルアクセス時刻から、各ファイルアクセス同士のリンクの強度は以下のように計算される。

$$L_{A \rightarrow B} = L_{A \rightarrow B}^c \times L_{A \rightarrow B}^t \quad (3)$$

$L_{A \rightarrow B}$ はファイルアクセス B に関するファイルアクセス A の関連度であり、ファイルアクセス A からファイルアクセス B へのリンクの強度を示す。

ここで述べた、テキスト情報とアクセス情報の共起によるリンク生成以外にも、アクセス

イベント同士に関連があると判断できる場合に、図 1 のように強度 1 のリンクを張る。具体的には、ManagementCore から得られる、ファイルのコピー、リネームがあった場合には、互いのアクセスイベントには関連があるとしてリンクを張る。図 1 のコピー・リネームに関して、コピーであれば、コピー元のファイルは存在しているため、同一ファイルに関するアクセスを示す縦へのリンクが存在するが、リネームであれば、リネームもとのファイルが消えてしまうため、縦線も消滅する。これにより、コピーとリネームを可視化されたネットワークから判別することができる。

上述したように、ファイルの関係を導くために、本論文ではファイルの中身を用いている。つまり、ファイルの中身が異なるファイル同士が関連するファイルと見なされることはない。もちろん、ファイルの中身が異なるファイル同士であっても関連するファイルである可能性もあるが、多くの場合、無関係なファイルであることのほうが多い。ファイル作成について考えると、あるファイルの中身を参考にファイルを作成することが多々あり、関連するファイルの中身はある程度似た内容になっているものと考えられる。そのため、本論文ではファイルの中身の類似性を用いて同一のタスクかどうかを判断している。

3.3 ファイルイベントネットワークの可視化

作成したファイルイベントネットワークを時系列インタフェースに描画する。時系列インタフェースを用いることで、ファイルが持つ情報の遷移が明らかになる。例えば、あるファイルに着目したとき、そのファイルが保有する情報が、他のファイルに分散していく様子が時系列インターフェースから取得できる。

図 2 に可視化インタフェースを示す。可視化インタフェースの左部分がファイル選択画面になっており、その中の 1 つのファイル（種ファイル）を選択すると、種ファイルに関するファイルイベントネットワークが可視化される。可視化されたファイルイベントネットワークのアイコンの上にマウスポインタを置くと、そのファイルアクセス情報とそのファイルの重要語が画面下部に表示される。アイコンをクリックすると、そのファイルに関するファイルイベントネットワークが新たに表示される。これにより、ファイルイベントネットワークをたどっていけば、所望のファイルにたどり着くことができる。

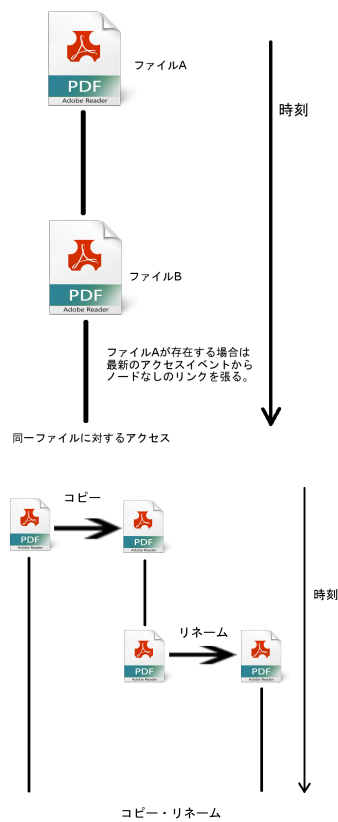


図 1 リンクの生成例

Fig. 1 Examples of link generation

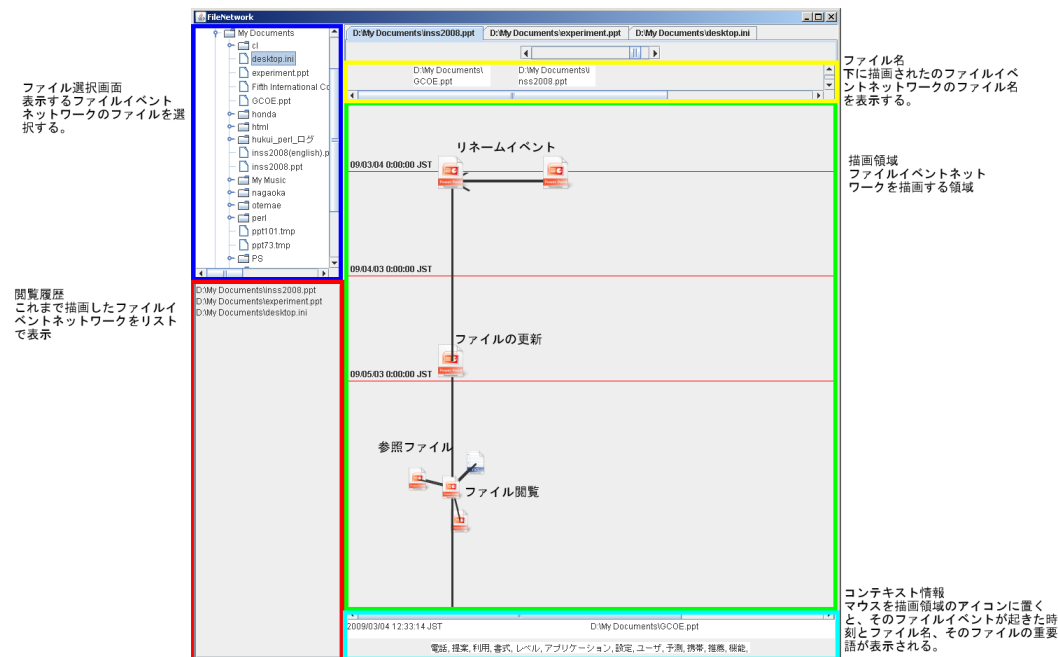


図 2 ファイルイベントネットワークインタフェース

Fig. 2 File event network Interface

4. 実装と有用性の検証

提案法の有用性を示すために、以下の2つの実験を行った。

- (実験1) 長期間データを取得し、過去に作成したファイルに関するファイルイベントネットワークを可視化することによって、得られた知識の具体例を示し、タスク想起への効果を測る。

- (実験2) 被験者に複数のファイルを引用した新しいファイルを作成してもらい、3週間後に新しいファイル(種ファイル)を基に、引用されたファイルの検索やウェブ検索を行ない、検索システムとしての有用性を測る。

4.1 パラメータと前処理

提案法は、ファイルに含まれる重要語とファイルのアクセス時刻を用いてファイルイベントネットワークを描画する。重要語については、得られる上位10個の単語を重要語とし、重要語が1つでも重複すれば関連するファイルとみなした。ファイルアクセスについては、ファイルアクセス時間差のパラメータ $T = 30min$ とし、前後30分以内であれば、同時刻に利用していたのとみなし、その時間内に関連ファイルがアクセスされていれば、参照ファイルとしてファイルイベントネットワークに追加した。本論文で30分以内のアクセスに限って、同時刻に利用されたと見なしたのは、本論文で用いているログ収集ソフトでファイルを閉じるというイベントを取得することができず、ファイルが参照されていた時間を取得することができないためである。そのため、一度開かれたファイルは30分程度は開かれているだろうという仮定の元に本論文では30分前後という値を用いている。ファイルを閉じるというイベントを取得することができれば、30分という前提を使う必要はなくなるが、30分が適切であるかどうかについてはコンテキストに依存する部分でもあり、検討する必要がある。

4.2 実験1:可視化による効果

長期間データを取得し、普段のデスクトップ上での作業を取得するために、2台のPCを用意し、ファイルイベントネットワークを実際に構築し可視化する実験を行った。ユーザA: CPU Pentium(R) 4 3.60GHz, メモリ 2GB, OS Microsoft Windows XP と、ユーザB: CPU Core 2 Duo 3.06GHz, メモリ 4GB, OS Microsoft Windows XP である。ユーザAは7ヶ月間、ユーザBは約2ヶ月間に渡りログを収集した。ユーザAのログにおいて、解析の対象となったファイル数は2440個である。

4.2.1 可視化例1

図3に被験者のPCにおけるファイルイベントネットワークの一例を示す。

図1に示すように、縦線は同一ファイルへのアクセスを示し、横の矢印はコピー・リネームを示す。図3はある発表資料であるfile_Bのファイルイベントネットワークの全体図である。file_Bが最も古く作成されたファイルであるため、file_Bによって複数のファイルが作成されたことは分かるが、どのような手順で複製されたのかが分からない。そのため、コピーイベント部分を拡大することによって、file_Bの複製過程を詳細に示す。コピーイベン

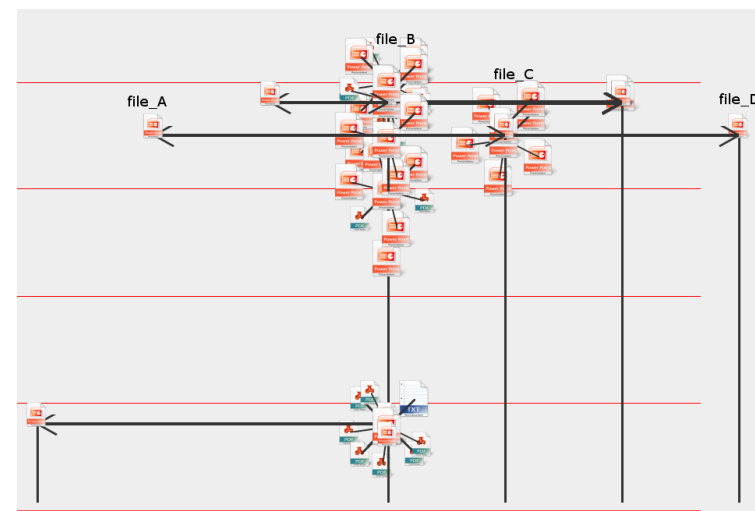


図3 可視化例1
 Fig.3 Example of visualization 1

ト部分を拡大した図を図4に示す。

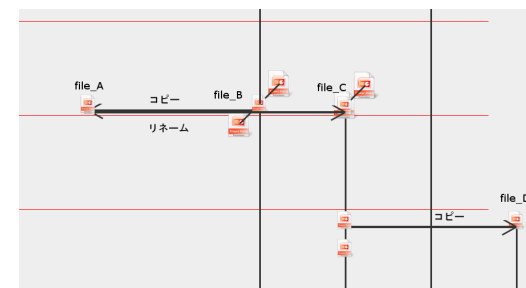


図4 可視化例1の拡大図(コピーイベント)
 Fig.4 Expanded figure for example operation 1(the process of coping file_B)

図4より、file_Bはfile_Aにコピーされ、file_Aはfile_Cというファイルに変更されており、file_Cは後にfile_Dにコピーされている様子が分かる。つまり、file_B → file_C → file_Dという流れでファイルが作成されていったことが分かる。

また、file_B が作成された過程を見るために、ファイル作成部分を拡大したものを図 5 に示す。

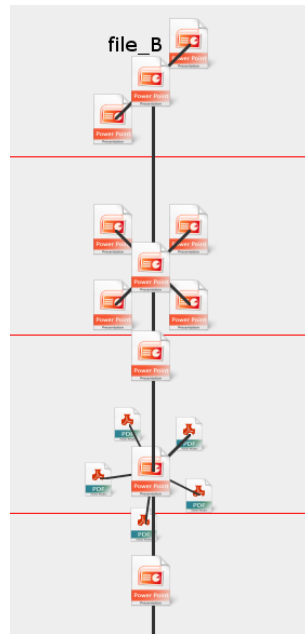


図 5 可視化例 1 の拡大図 (file_B の作成過程)

Fig. 5 Expanded figure for example operation 1(the process of making file_B)

図 5 より、file_B は様々なファイルを参考にしながら、徐々に作成されていったことが分かる。また、図 3 より、file_B は作られた時期が最も古いにもかかわらず、編集された時期が他の派生ファイルに比べて最も新しいため、file_C や file_D よりも新しい情報が付け加えられた、つまり新しいバージョンのファイルである可能性が高い。このように、ファイルの編集時期を派生ファイルと簡単に比較することができるため、提案法はファイルのバージョン管理を行なううえでも有用であると考えられる。

4.2.2 可視化例 2

図 6 は提出資料 file_H を作成した際のファイルイベントネットワークを表している。file_H

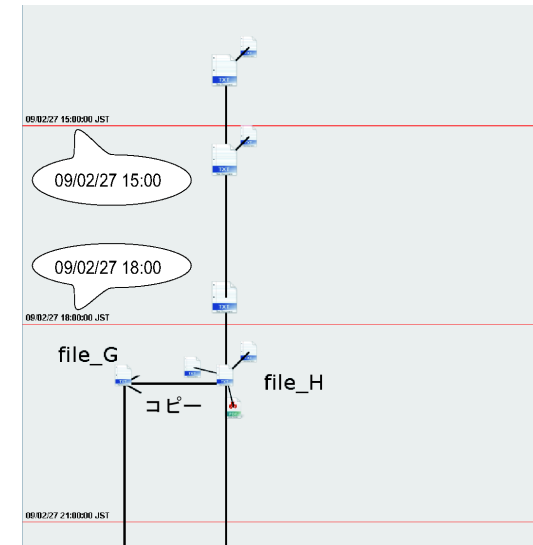


図 6 可視化例 2

Fig. 6 Example of visualization 2

は 1 つのファイルを参考に作成されており、後に file_G にコピーされている。作成日時から数時間という短期間に作成された資料であることがわかる。また、作成されてから後に参照されていないことから、最近では利用されていないファイルであることが分かる。また、file_G にコピーされた後、file_G が編集されていないため、同一のファイルが PC 内に 2 つ存在していることが分かる。これにより、file_G がバックアップのためのファイルでないとすれば、file_G は不要なファイルであることが分かる。

4.3 実験 2: 検索システムとしての評価

実際にファイルイベントネットワークを可視化することでユーザがタスクを思い出し、所望のファイルやウェブページを効率的に検索することができるか検証を行った。被験者は研究室の学生 9 名である。

4.3.1 実験 2-1: ファイル検索の評価

ファイルを保管する際、適切なフォルダに割り振って保管することが多いが、ファイルの作成時期やファイルの内容によって、1 つの適切なフォルダにファイルを割り振ることができないことがある(例えば、論文内に登場する実験結果の画像ファイルを学会フォルダに保

管するか、実験データフォルダに保管するか迷う場合がある。)このような場合、同じファイルをコピーして別々のフォルダに保存したり、研究関連フォルダといった浅い階層のフォルダにまとめて保存しておくといったファイル管理をすることがある。また、一時的にデスクトップ上にファイルを置くようなこともあるため、1つのフォルダに複数のファイルが存在している状況が考えられる。我々は、このような状況下で、適切にファイル検索を行なうことができるかを検証した。実験のために、77個のファイルを用意し、1つのフォルダの中に格納した。77個のファイルの中身は株(10)、法律(13)、ネットワーク(14)、ニュース(12)、テクノロジー(15)、スポーツ(13)に関する記事で、カッコ内はその内容に関するファイル数である。ファイル名については、ファイルに書かれてある最初の一文(句読点まで)をファイル名とした。例えば、株に関するファイルであれば、「株式投資で成功するかどうか」や「株式投資には魅力もあるがリスクもある」があり、お互いに内容の似たファイルも存在する。また、法律関係では「裁判員制度の紹介」や「貸金返還請求事件」といったファイルが存在する。また、ファイルによっては、「昨年」「現在」といったファイル名だけではファイルの中身が分からないファイルも存在する。実験の流れを以下に示す。

- (1) 77個のファイルから、筆者が指定した株に関する4つのファイル(元ファイル)を参考に、1つのファイル(種ファイル)を作成してもらおう。4つのファイル名を以下に示す。
 - ドルコスト平均法で買い付け
 - 株式投資を始めて株主になるメリット
 - 金融商品を選ぶには
 - 連続したローソク足で株価の動きを見てみよう
- (2) 数日後、種ファイルのバックアップを作成してもらい、さらに、元ファイルとは別のファイル(ファイル名は「アナリストの銘柄診断は要チェック」)を用いて、種ファイルに内容を追加してもらおう。
- (3) ファイル作成から3週間後、種ファイルを参考に4つの元ファイル全て探し出してもらおう(制限時間5分)

元ファイルの中で見つけることができたファイルの数や検索に要したクリック数と時間から提案法の有用性を検証する。検索方法は以下の3つである。

- フォルダ内のファイルをファイル名を手がかりに検索
- キーワードによる検索(Windowsのファイル検索機能)
- ファイルイベントネットワークインタフェースを用いて検索

表3 見つけることができたファイル数(見つけるまでに要したクリック数と時間(秒))

Table 3 No. of files that was able to be bound(No. of clicks and time(seconds) which was required by finding the files)

被験者	キーワード	提案法	
A	3(22, 95)	3(55, 300)	4(18, 58)
B	3(35, 163)	3(25, 97)	3(30, 131)
C	2(60, 300)	3(78, 300)	4(22, 73)
D	3(45, 182)	2(40, 105)	2(30, 60)
E	4(19, 91)	3(90, 300)	4(26, 83)
F	4(46, 180)	2(70, 300)	4(36, 205)
G	3(47, 208)	4(30, 113)	4(21, 258)
H	4(63, 275)	4(100, 245)	4(20, 50)
I	4(50, 216)	3(100, 300)	4(35, 47)

実験結果を表3に示す。

表3に示された数値は元ファイルの中で、被験者が見つけることができたファイルの数であり、括弧内の数値は右がファイルを見つけるまでに要したクリック数と左が時間(秒)である。太字の数値は3つの検索方法のなかで、最も良い結果を示した数値である。表3から、提案法がもっとも正確にファイル検索を行えることが分かる。被験者BとDは、すべての元ファイルを見つけることができなかったが、これは、元ファイルに含まれるキーワードが作成されたファイルの中に存在しなかったためである。また、被験者Bにおいて、提案法による検索に時間がかかった原因は、種ファイルのイベント間隔が短かったために、目的のファイルアイコンをなかなか見つけることが出来なかったためである。インタフェースとしての改良は必要であるが、ファイルイベントネットワークを示すことによって、ファイル作成時のコンテキストを思い出させることができ、所望のファイルを検索できたことが実験により示された。

被験者Aのファイルイベントネットワークを図7に示す。

図7より、ファイルが作成された際に、複数のファイルが参照されていることが分かる。また、ファイルが作成された後に、バックアップのためにファイルがコピーされていることもわかり、バックアップファイルを探すのも容易であることが分かる。

4.3.2 実験 2-2:ウェブ検索の評価

文書ファイルを作成する際、ローカルファイル以外にもウェブページを参考に文書を作成することがあるが、そのようなウェブページはブックマークされることなく、閲覧後閉じられることが多い。そのため、文書を編集したり、似たようなファイルを作成する際に、参考

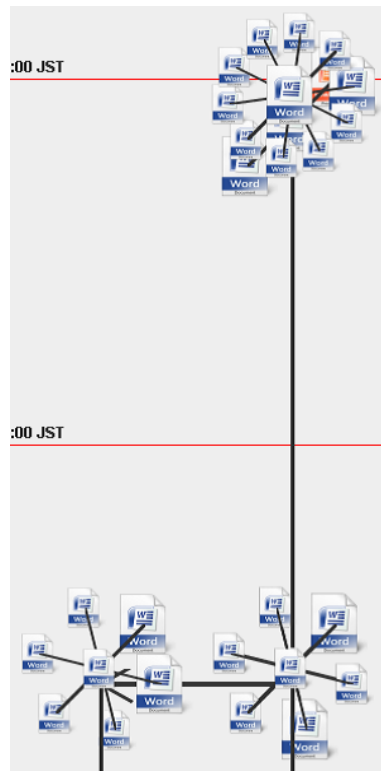


図 7 被験者 A のファイルイベントネットワーク
 Fig.7 file event network for examinee A

にしたウェブページを閲覧しなければならぬ場合、キーワード検索を用いて再度検索を行なうことがある。しかしながら、時間が経過している場合、簡単なキーワード検索だけでは目的のウェブページを見つけることができないこともあり、そのようなウェブページに対して、複数または特定のワードを入れて検索したり、ウェブの閲覧履歴から探したりするが、いずれの場合も多くの時間を費やすことが考えられる。そこで、ファイルに関するウェブページをファイルイベントネットワークに表示することで、ウェブページを参考に作成されたファイルから、参考にしたウェブページを効率的に検索することができるか検証を行なった。実験の流れを以下に示す。

表 4 検索できたウェブページ数/参考にしたウェブページ数 (見つけるまでに要したクリック数と時間 (秒))
 Table 4 No. of Web pages that was able to be bound/No. of Web pages that was referred(No. of clicks and time(seconds) which was required by finding the files)

被験者	キーワード	提案法
A	2/2(12, 39)	2/2(9, 17)
B	1/1(10, 34)	1/1(4, 20)
C	3/3(30, 125)	3/3(12, 38)
D	1/1(12, 38)	1/1(4, 7)
E	4/4(65, 215)	4/4(40, 114)
F	4/5(44, 300)	5/5(34, 123)
G	2/2(12, 52)	2/2(14, 40)
H	3/3(55, 182)	3/3(40, 167)
I	2/2(30, 57)	2/2(16, 37)

- (1) 「先物市場」とは何かをウェブページを参考にファイルにまとめてもらう。(参考にしたウェブページは被験者の任意であり、参考にしたウェブページ数は被験者によって異なる.)
- (2) ファイル作成から 3 週間後、種ファイルを元に参考にしたウェブページを探し出してもらう (制限時間 5 分)

参考にしたウェブページの中で、見つけることができたウェブページ数や検索に要したクリック数と時間から提案法の有用性を検証する。検索方法は以下の 2 つである。

- キーワードによる検索 (被験者が普段使い慣れている検索エンジンを使用する)
- ファイルイベントネットワークインタフェースを用いて検索

実験結果を表 4 に示す。

表 4 に示された数値は検索できたウェブページ数/参考にしたウェブページ数であり、括弧内はクリック数と時間 (秒) である。ほとんどの被験者がどちらの手法でも全てのウェブページを見つけることができた。これは、ほとんどの被験者が Wikipedia や Yahoo 知恵袋といった検索ランキング 10 位以内のウェブページを参考にファイルを作成したためである。検索時間を比較すると、より多くのウェブページを参考にした被験者に対して、提案法は有効であることが示された。これは、キーワード検索が一つ一つのウェブページに対して、適切なキーワードを選んで逐一検索する必要があるのに対して、提案法では、1 つのファイルイベントネットワークの中に参考にしたウェブページ候補が全て挙げられているため、簡単に検索できたのだと考えられる。また、被験者 F において、キーワード検索で見つけることができなかったファイルを、提案法では検索することができている。

4.3.3 実験 2-3: アンケートによる評価

被験者 9 名に提案法の検索システムについて、無記名のアンケートに答えてもらった。アンケート内容と回答は以下のとおりである。

- (1) 3つの検索方法でファイル検索を行いました。簡単にファイルを探せた順に並べてください。
 - フォルダ内のファイルを手探りで検索
 - 1位 1名
 - 2位 4名
 - 3位 4名
 - キーワードによる検索
 - 1位 0名
 - 2位 4名
 - 3位 5名
 - 提案法による検索
 - 1位 8名
 - 2位 1名
 - 3位 0名
- (2) ファイルイベントネットワークを見ることで、ファイル作成の流れを思い出すことが出来ましたか？
 - 全く思い出せなかった。0名
 - 少しは思い出せた。2名
 - かなり思い出すことができた。7名
- (3) ファイルイベントネットワークインタフェースは、検索システムとして、有用だと思いますか？
 - 有用だとは思わない。1名
 - 少しは有用である。2名
 - かなり有用である。6名

ファイル検索を行った際に、提案法が最も簡単にファイルを探すことができた理由について、参考にしたファイルが絞られていたため、簡単にファイルを検索することができたという回答が最も多かった。提案法を2位にあげた被験者は、ファイル名を覚えており、フォルダ内を探すほうが簡単だったと回答している。インタフェースについては、操作性の面での

改良が必要だという意見を多く頂いたが、直感的に操作できれば、有用であるという回答を得ることができた。

4.4 ブラウジングインタフェース

被験者実験により、ファイルイベントネットワークを可視化することによって、ユーザがタスクのコンテキストを思い出し、所望のファイルを検索することが可能であることが示された。しかしながら、提案法は種ファイルに関するファイルイベントネットワークを作成するものであり、種ファイルを探すことが出来なければ、所望のファイルを見つけることができない。そこで、我々は関連のありそうなファイルのファイルイベントネットワークをたどることによって、所望のファイルを検索することができるブラウジング機能をインタフェースに追加した(図8参照)ブラウジング機能により、個々のファイルアイコンをクリックすることで、そのファイルを中心とするネットワークが随時展開される。

図8のように、file_A についてのファイルイベントネットワークを可視化することによって、file_A は file_B を参照して作成されたファイルであることが分かる。さらに、file_B のアイコンをクリックすることによって、同一のインタフェース上に file_B のファイルイベントネットワークを追加することができる。このように、関係のあるファイルについて、ファイルイベントネットワークを展開していくことで、各ファイルのコンテキストが明らかになる。これにより、ユーザは各ファイルの関係性とその操作内容を同時に見ることができ、より鮮明にユーザのタスクを思い出すことができる。また、ネットワークを展開していくことで、所望のファイルにたどり着くことが可能となる。

5. ま と め

本論文では、従来のファイル名やキーワードによる検索と異なり、ファイルに対する一連の操作に関する時系列履歴を「ファイルイベントネットワーク」として可視化する方法を提案した。さらに、被験者実験を行なうことによって「ファイルイベントネットワーク」をユーザに提示することで、ファイルに関するコンテキストを思い出し、効率的にファイルやウェブ検索を行なうことができることを示した。現在は単一の PC のファイルのみを対象としているが、複数 PC や複数人への拡張も可能である。たとえば、PC 間でのファイルの移動を把握したり、メールに添付されることで人から人にファイルが移動するイベントを把握することも可能である。これにより、組織内の情報の流れや、メンバーの活動などを把握でき、例えばシステム開発におけるプロジェクトの進捗状況をリアルタイムに把握することも可能である。また、¹³⁾では、ファイル操作履歴とウェブアクセスに加え、メール送受信情報も

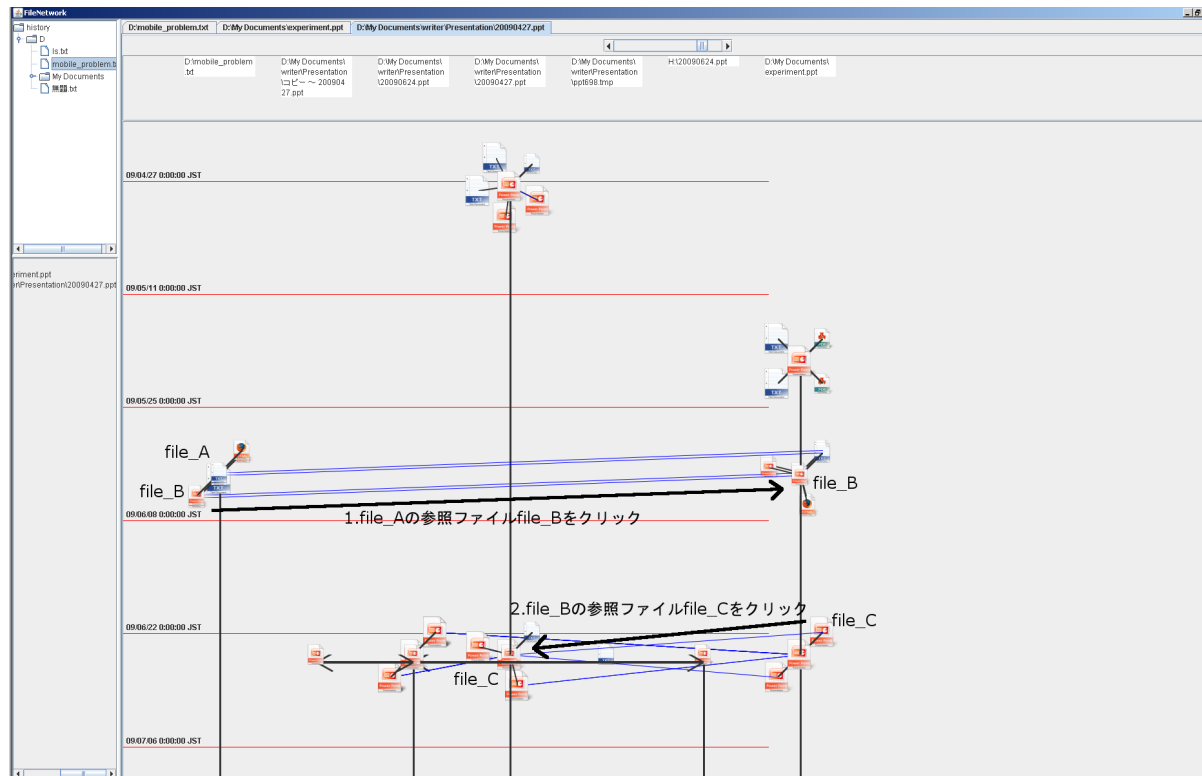


図 8 ブラウジングインタフェース

Fig. 8 browsing interface

取り入れられており、本研究においてもこれらの履歴を早急に取り込む計画である。

参考文献

- 1) Teevan, J., Alvarado, C., Ackerman, M., Karger, D.,: The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search, *In CHI '04: Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 415–422 (2006).
- 2) Cohen, S., Domshlak C. and Zwerdling, N.: On ranking techniques for desktop search, *Transactions on Information Systems (TOIS)*, Volume 26 Issue 2, pp.11:1–11:24 (2008).
- 3) Zobel, J. and Moffat, A.: Inverted Files for Text Search Engines, *ACM Computing Surveys*, Volume 38 Issue 2, pp.1–56 (2006).
- 4) Shah, S., Soules, C., Ganger, G. and Noble, B.: Using Provenance to Aid in Personal File Search, *In Proceedings of USENIX Annual Technical Conference (USENIX 2007)*, pp.1–14 (2007).

- 5) Cutrell, E., Robbins, D., Dumais, S. and Sarin, R.: Fast, flexible filtering with PHLAT Personal search and organization made easy , *In CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp.261–270 (2006).
- 6) Nejd, W., and Paiu, R.: I know I stored it somewhere - Contextual Information and Ranking on our Desktop, *8th International Workshop of the EU DELOS Network of Excellence on Future Digital Library Management Systems*, pp.1–5 (2005).
- 7) Sauermann, L., Bernardi, A., and Dengel, A.: Overview and Outlook on the Semantic Desktop, *Proceedings of the 1st Workshop on the Semantic Desktop at the ISWC 2005 Conference*, pp.1–18 (2005).
- 8) Blanc-Brude, T. and Scapin, D.:What do people recall about their documents?: implications for desktop search tools, *IUI '07 Proceedings of the 12th international conference on Intelligent user interfaces*, pp.102–111 (2007).
- 9) Ringel, M., Cutrell, E., Dumais, S. and Horvitz, E.: Milestones in time: The value of landmarks in retrieving information from personal stores, *Proceedings of Interact 2003*, pp.184–191 (2003).
- 10) 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 (2005).
- 11) Rekimoto, J.: Time-Machine Computing : A Time-centric Approach for the Information Environment , *In UIST '99: Proc. of the ACM Symposium on User Interface Software and Technology*, ACM, pp.45–54 (1999).
- 12) 牛尼 剛聡, 渡邊 豊英: 経験コンテンツ検索における意味的リンク構造を利用したランキングに関する考察, 日本データベース学会 Letters, Vol.6, No. 2, pp.73-76 (2007)
- 13) Shen, J., Fitzhenry, E. and Dietterich, T.: Discovering frequent work procedures from resource connections, *The 13th international conference on Intelligent user interfaces (IUI '09)*, pp.277–285 (2009).
- 14) 福井 秀徳, 森田 哲郎, 岡野 真一, 沼尾 正行, 栗原 聡: ファイルネットワークに基づいた情報の抽出と可視化, 人工知能学会全国大会 (第 22 回) 論文集, CD-ROM(2008).
- 15) 松尾 豊, 石塚 満: 語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム, 人工知能学会誌, Vol.17, No.3, pp.213–227 (2002)