

複数PC間ファイルネットワークによるプロジェクト管理のためのインターフェース

松本 光弘[†], 福井 秀徳[†], 森田 哲郎^{††}, 岡野 真一^{††}, 沼尾 正行[‡],
栗原 聡[‡]

[†] 大阪大学大学院情報科学研究科情報数理学専攻 ^{††} 住友電気工業株式会社 [‡] 大阪大学産業科学研究所

産業における情報化が進む中、その情報が漏洩するといった問題が発生し、情報を管理することへの重要性が高まっている。また、構成人数や端末の多様化・複雑化によって、プロジェクト管理が困難になっており、プロジェクトを円滑に進めていくことも重要な課題となっている。こういった状況に対応するため、管理者はグループ内の情報遷移の状況やメンバの活動状況を監視する必要がある。本研究では、機密ファイルの所在と遷移を明らかにするために、複数PC間のファイルネットワークを可視化するインタフェースを作成した。これにより、主要なファイルと関連ファイルの変遷、それに関わった人間などの情報を一度に得ることができることを確認した。

File Network Interface of Multi-PC for Project Management

Mitsuhiro MATSUMOTO[†] Hidenori FUKUI[†] Tetuo MORITA^{††} Shinichi OKANO^{††} Masayuki
NUMAO[‡] Satoshi KURIHARA[‡]

[†] Department of Information and Physical Sciences, Graduate School of Information Science and
Technology, Osaka University

^{††} Sumitomo Electric Industries, Ltd.

[‡] The Institute of Scientific and Industrial Research, Osaka University

Along with the development of information industry, it is more important to control confidential information. Due to the growing diversity and complexity of group members and terminal machines in the organization, it is also more difficult to control project. Therefore, a project manager should get to know more about member's activity. In this paper, we constructed a file network interface of Multi-PC in order to know whereabouts and transitions of confidential information. The Interface can inform us flows of information and activities of group members.

1 はじめに

産業における情報化が進む中、その情報が漏洩するといった問題が発生し、情報を管理することへの重要性が高まっている。企業は、社外からの攻撃に対してファイアウォールやアンチウイルスソフトといったセキュリティソフトを導入し、情報漏洩防止に対応しているが、企業内の人間による機密ファイルのうっかり送信や、外部への持ち出しといった問題には対応できていない。構成人数や端末の多様化、複雑化によって、プロジェクト管理が困難になっており、プロジェクトを円滑に進めていくことも重要な課題となっている。これらの問題は、プロジェクトマネージャが企業内の

個々の人間の活動を全て把握できていないことに起因する。情報の所在を明らかにすることで、情報の所有者に情報の機密性に依じて注意を促すことができ、機密ファイルのうっかり送信を減らすことができる。また、個人の所有している情報や活動状況から適切に仕事を割り振ることもできる。このように、グループ内の情報遷移の状況やメンバの活動状況を把握することは、セキュリティや仕事効率の点で非常に有用である。

本研究では、機密ファイルの所在と遷移を明らかにするために、複数PC間のファイルネットワークを可視化するインタフェースを作成した。これにより、主要なファイルと関連ファイルの変遷、そ

れに関わった人間などの情報を一度に得ることができ、各メンバの活動が容易に把握できることが分かった。

2 関連研究

本研究において提案するシステムはファイルネットワークを用いて複数のコンピュータ間におけるファイルの変遷を視覚化するものである。リンク(関連性に基づいた繋がり)を辿ることで関連性の深いファイルを導き出すという意味では、ファイル検索ツールの側面をもち、情報の流れを視覚化するという意味では、プロジェクト管理システムとしての側面をもつ。ファイル検索において、ファイルのテキスト情報はファイルを特定する際の有用な手掛かりとなる²⁾。また、ユーザのファイルアクセス情報を用いて、時間的な手掛かりからファイルを特定する方法も有用な手段である^{2) 3)}。プロジェクト管理において、大平らはソフトウェア開発データを自動収集・解析するプロジェクト管理ツール Empirical Project Monitor (EPM) を作成している⁴⁾。開発過程をリアルタイムに可視化することは、作業停滞の原因を究明したり、進捗具合を把握することができるため、プロジェクトの流れを見通し、プロジェクトを円滑に遂行するためにも重要なことである。このように、ファイルのテキスト情報とアクセス情報を時系列インターフェースに適用することは、ファイル検索やプロジェクト管理という点で有用であると考えられる。

3 テキスト情報とアクセス情報に関する検証

ここでは、テキスト情報とアクセス情報について、簡単な検証を交えた上で両者の特性を挙げ、組み合わせることの有用性について述べる。

3.1 テキスト情報から得られるファイル相関性の検証

図1は筆者がPC内で作業を行う際に利用するディレクトリ内のデータに対して、テキスト情報を用いて作成したファイルネットワークである。各テキストから、重要語を抽出し、重要語が一つ以上共起しているファイル同士にリンクを張った。テキスト解析の対象としたのは、一般的なユーザが使用する機会の多いテキスト情報を含むフォーマット(txt, html, doc, ppt, xls, pdf, tex等)であ

る。グラフの左側中央、右側中央、左側下部の三箇所において、類似性の高いファイルの集まりが見られる。これらのファイルは比較的内容が近いファイルの集合となっていることから、テキスト情報のみで、大まかな分類が可能であることがわかった。しかしながら、グラフの中には、筆者にとって直感的に分かり辛いファイルの繋がりも存在する。これは、ユーザ自身が、関連があると自覚していないファイル同士から共通のキーワードが抽出されてしまうことが原因である。このように、テキスト解析は高い精度でファイルの相関性を抽出するが、ユーザが意図しないファイル同士が関連のあるファイルとして抽出される恐れがあることが分かった。また、ここで得られた結果のように、テキスト解析で得られる結果は一意的であり、ユーザの個性や特徴に応じた結果は期待できない。

3.2 アクセス情報から得られるファイル相関性の検証

図1の平面グラフに対して新たに第3の軸としてファイルの最終更新時刻を加え、3次元空間上にノードをプロットしたのが図2である。新たに時間情報が与えられたことで、就職活動関連ファイルのような時間に強い依存性を持ったファイル群が新たなまとまりとして抽出されている。ユーザはなんらかの目的のためにファイル操作を繰り返しており、ここで見られた時間的に近い関係にあるファイルは共通の目的のために使用されたファイルであると考えられる。ユーザは自身のファイル操作をイベントの前後関係によって記憶に留めることが多いため、近い時刻にアクセスされたファイル同士は、ユーザの記憶の想起を促す効果も期待できる。時間情報のみを用いた解析では高い精度の関係抽出は難しいが、テキスト解析の結果と組み合わせることで、ユーザに応じたファイル相関性の抽出が可能になることが明らかになった。

4 ファイル相関ネットワーク視覚化システムの提案と実装

本研究では、ファイルのテキスト情報が類似しており、またアクセスされた時間が近いファイル同士にリンクを張ることでファイル相関ネットワークを作成する。



Fig. 1 キーワード共起によるファイルネットワーク

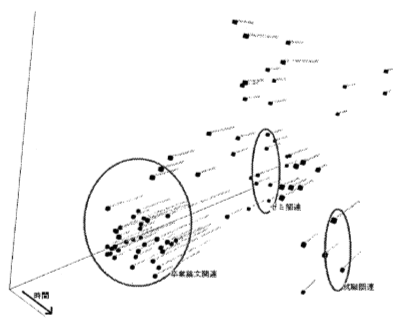


Fig. 2 図 1 に時間軸を加えた 3 次元図

4.1 提案

まず、1つのファイルに着目し、そのファイルがアクセスされた時刻の周辺で生じた他のファイルを抽出する。ファイル同士のテキスト情報からファイルの類似度を測定し、ファイルアクセスイベントをノード、情報の類似度をリンク強度としたファイルネットワークを作成する。さらに、作成されたファイルネットワークを図 4.5 のような時系列インタフェースに描画する。これによりファイルが持つ情報の遷移が明らかになる。

4.2 実装

本システムの構成を図 3 に示す。本システムでは現在、クライアント・サーバ型のネットワーク環境を構築し、ファイルやファイルアクセスログといった必要な情報をサーバで集中管理している。ファイル同士の類似を測定するために、予めファイルに関する重要語を抽出する。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 Namaze⁵⁾ を用いて、プレーンテキストへと変換した後、形態素解析システムである茶室⁶⁾ と単名詞の連結に基づいて重要語抽出を行ってくれる TermExtract⁷⁾ を用いて重要を抽出した。情報の類似度を表すリンク強度は以下の式によって求められる。

$$L_{A \rightarrow B} = \frac{KeyWord_a(t_A) \cap KeyWord_b(t_B)}{KeyWord_a(t_A)} \quad (1)$$

$L_{A \rightarrow B}$ はファイルアクセス B に対してファイルアクセス A が持つ関係値である。 $KeyWord_a(t_A)$ は時刻 t_A の時点でファイル a が持つ重要語数とな

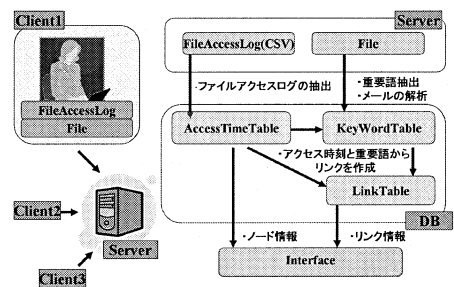


Fig. 3 システム構成

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。最後に、ファイルアクセス情報とリンク強度を用いてネットワークを作成し、インターフェースに描画する。

5 動作実験

本システムを実際に適用した。解析対象のクライアント PC の内の 2 台は学生所有の研究用デスクトップ型 PC であり、残りの 1 台は教員所有のモバイル型 PC である。

5.1 結果 1: グループ内の情報遷移

本システムにおいて、実際にファイルネットワークを描画した際のスクリーンショットを図 4 に示す。図 4 は各クライアントがメールにより情報を伝達する様子を描画している。複数のエリアを横断する形で張られた水平のリンクが複数見られる

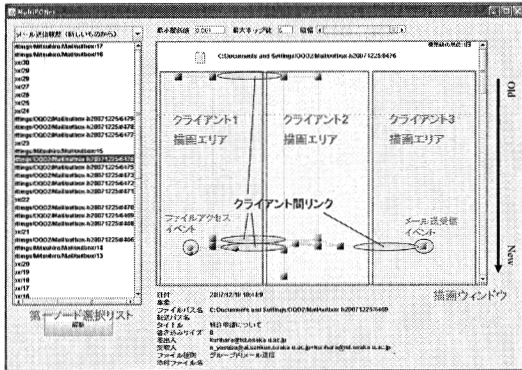


Fig. 4 システムインタフェース 1

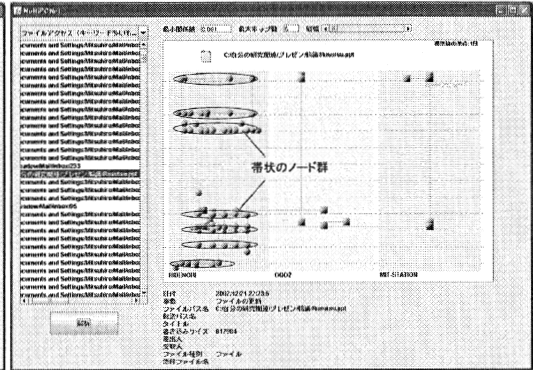


Fig. 5 システムインタフェース 2

が、これらはクライアント間でのメールの送受信を示している。またそれらのメールに付随する形で関係の深いファイルの存在が明らかになっている。このように、ファイル同士の情報の引継ぎや、クライアント間での情報伝達といった情報の遷移を時系列インタフェースから容易に把握できる。

5.2 結果 2:メンバの PC 内活動

図 5 において、左側に示されたクライアントに着目すると、定期的に帯状のノード群を見ることができる。図 5 で対象となったファイルは研究の発表資料であり、帯状のノード群の中には新たに作成された発表資料が存在した。これより、このユーザは過去のファイルを元に、新しい区発表資料を作成していることが分かった。本システムでは、ファイルネットワークから、各メンバの PC 内における活動を確認することができた。

6 まとめと今後の課題

複数のクライアントマシンについてファイルネットワークを作成し、それを時系列インタフェースに適用した。これにより、組織内の情報の流れや、メンバの活動などを容易に把握でき、プロジェクトマネージャは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ができる。これは、プロジェクトを円滑且つ確実に遂行していく上で大変有用である。また、機密情報を含んだファイルを参考に作成されたファイルにも機密情報が含まれている可能性が高く、本システムではそのようなファイルの所在も明らかにすることができる。さらに、

関連するファイルの所在を明らかにしてくれる本システムは、ファイル検索にも応用することが可能であると考えられる。

今後は、より多くのデータを収集するために、被験者の数とデータ収集期間を増やし、本システムの有用性をより実証していきたいと考えている。

参考文献

- 1) : Google Desktop, <http://desktop.google.com/en/>.
- 2) Ringel, M. et al.: Milestones in time: The value of landmarks in retrieving information from personal stores, *Proceedings of Interact 2003*, pp. 184–191 (2003).
- 3) 大澤亮他：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).
- 4) 大平雅雄他:ソフトウェア開発プロジェクトのリアルタイム管理を目的とした支援システム, 電子情報通信学会論文誌, No. 2, pp. 228–239 (2005).
- 5) : 全文検索システム Namazu, <http://www.namazu.org/index.html.ja>.
- 6) : 形態素解析システム茶釜, <http://chasen.naist.jp/hiki/ChaSen/>.
- 7) : TermExtract, <http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html>.