

D-010

タグ情報に基づくファイル管理システムと実用性の評価

An Evaluation of tagging file system

阿部 淳也†
Junya Abe出石 大志†
Hiroshi Izuishi堀 幸雄†
Yukio Hori今井 慈郎†
Yoshiro Imai

1. はじめに

近年、ネットワーク環境の充実や HDD の低価格化に伴い、これまで各人の PC 内に保存していたファイルの共有が活発になってきている。各人の研究活動で作成されるファイル群を研究室内において共有することは、研究についての議論・討論の円滑化を促し、研究の引継ぎにおいてもファイル移動の手間や不注意による損失を低減することが期待できる。本研究室ではネットワーク HDD を用いたファイル共有を試みていたが、以下に述べるような問題点のため、効果的なファイル共有が難しい状況だった。

ファイル群は基本的にディレクトリをノードとする木構造形式で管理されるため、ユーザ間の木構造定義の差異により、共有空間内でのファイル保存先が決めづらいという問題点があった。同様の理由から、他のユーザがどのディレクトリにファイルを保存したか分かりづらい場合もあった。また、共有空間をワークスペースではなく保管庫という感覚で捉えるため、完成形の成果物のみが保存される。そのため、作業途中のファイルは共有されず、結果ファイルの共有数も少なく、ユーザにとってファイル共有の意義も低減する。

本稿では、複数のユーザによる効果的なファイル共有とユーザ間の木構造の差異を吸収するためのファイル管理機能の実装について述べる。また、プロトタイプをユーザに試験利用してもらい、ファイル共有の効果が向上するかどうか、アンケートにより評価した。

2. タグ情報を用いたファイル共有

2.1 木構造型管理における問題と解決

従来の一般的なファイルストレージにおいては、木構造型のファイル管理を行っている。トップダウン的な分類でファイルを管理するには適しており、ノード数を増やしていくことによりファイルを詳細に分類することが可能である。また、ノードの親子関係（上下関係）の分類もできる。しかし、複数人が共同利用するストレージにおいては、ユーザ間の木構造の差異（意味的位置付けの違いなど）によりファイル管理が適切とは言えない。

そこで、ファイル管理にキーワードベースの視点を適用することを考える。キーワードを用いたファイル管理については以前から研究が行われており、UNIX マシン上での実装事例などもある[1]。本研究でも、木構造による管理ではなく、キーワードをベースにしたタグ情報を用いてファイル管理を行い、ファイルの効果的な共有環境を提供する。

本手法では、ファイルにタグ情報を付与して管理することにより、ユーザは木構造型のディレクトリ構成を意識せず、ファイルを共有することができる。ファイル分類の詳細化は付与するキーワードを増やすことで可能となる。逆

に、検索時のキーワードを減らすことで、大まかな分類でファイルを開覧することもできる。また、ディレクトリノード間の親子関係は、クライアント PC 上のファイルの保存パスを利用することで表現できる。一方、ユーザ間でこの親子関係が異なりうまく迎れない問題も、親か子のどちらかのキーワードに直接アクセスし、その上下関係をたどることで解決できる。

2.2 タグ情報の構成

共有されるファイルには表1に示すタグ情報が付与される。ファイル検索に用いる情報は、表中の「キーワード」項目のみではなく、他の項目も対象として検索を行う。

表1 タグ情報の構成

名称	内容
ファイル ID	ファイル固有の ID
オーナー情報	ファイル所有者の情報
登録日付	ファイル登録時の日付
キーワード	ファイルの特徴を表すキーワード群
共有パス	共有空間におけるファイル保存パス
ローカルパス	ファイル所有者の PC におけるファイル保存パス
ファイル名	ファイル登録時のオリジナルファイル名
拡張子	ファイル拡張子

2.3 キーワード自動抽出

ファイルを登録するたびに、複数のキーワードからなるタグ情報を手動で入力すると、ユーザの手間が増大する。そこで、ユーザの手間を軽減し、キーワードの入力漏れを防ぐ目的で、以下の2種類の手法でキーワードの自動抽出を行う。

(1) ディレクトリ名からのキーワード抽出

ユーザはクライアント PC 上において、多種多様なファイル群を複数の階層からなる木構造に分類し、ファイルを格納している。その際、各階層の分類を表すディレクトリ名は、ファイルの内容を端的に表している情報の1つと考えることができる。そこで、対象となるファイルに至るまでのパス名をキーワードとして活用する。しかし、利用している OS の性質上、直接ファイルの内容に関係ないパス（ディレクトリ）名が付与されている場合がある。それらはキーワードとして用いないようにする。

(2) ファイル名からのキーワード抽出

ファイル名は、そのファイルの内容を最も良く表している情報と考えられる。そこで、ファイル名をキーワードとして用いる。対象ファイルのファイル名を形態素解析にかけ、得られた名詞をキーワードとして用いる。

3. システム設計

3.1 全体構成

本システムは、図1のように、情報、ファイルなどを管

† 香川大学

理するサーバ群と各ユーザに配布される操作用クライアントプログラムから構成される。

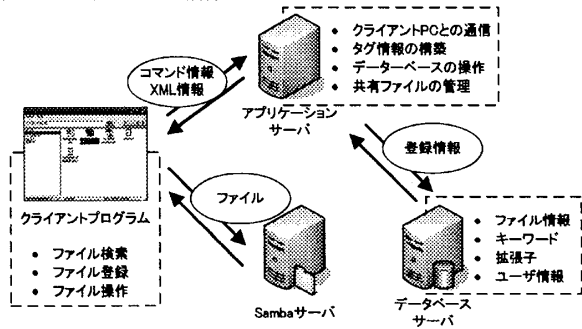


図1 システム全体構成

3.2 サーバ群

サーバ群は、アプリケーションサーバ、Sambaサーバ、データベースサーバから構成される。

(1) アプリケーションサーバ

サーバ側の各サーバ群を統括し、細部を隠蔽する役割を果たす。また、クライアントプログラムとの通信も本サーバが行う。TCPソケットを用いた通信を行い、独自に定義したプロトコルにより命令や情報をやり取りする。

(2) Sambaサーバ

本システムでは、サーバ側のファイルストレージとしてSambaサーバを用いる。これにより、クライアントプログラムを通してクライアントPC上のアプリケーションからサーバ側のファイルを直接読み書きすることが可能となる。ファイル転送などもSambaを通して行う。

(3) データベースサーバ

データベースには、タグ情報を構築するためのファイル情報やキーワード、拡張子、ユーザ情報などが登録されている。

3.3 クライアントプログラム

ユーザには、図2に示すGUI機能を有するクライアントプログラムを提供する。ユーザは、クライアントプログラムを通して、ファイルの共有環境への登録やキーワードによる検索などを行う。また、検索結果として表示されたファイルは、クライアントPC上において関連付けられたアプリケーションから直接実行することができる。

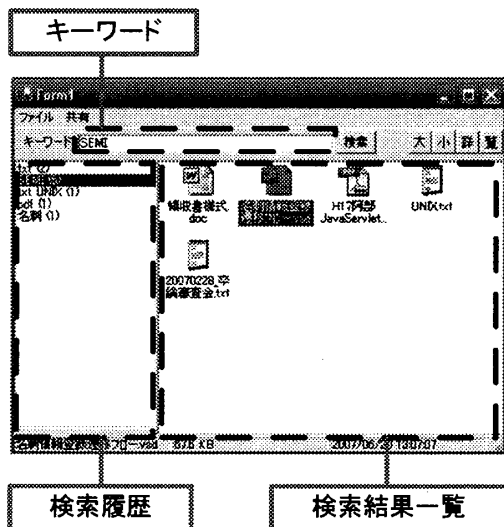


図2 クライアントプログラム

4. ユーザ評価

4.1 評価内容

本システムの実用性に関する評価として、実際にユーザに利用してもらいアンケートを取った。内容は、従来方式(木構造ベースのExplorer)と提案方式(タグ情報ベースの本システム)の比較が主である。直感的にどちらの操作が分かりやすいか(操作理解性)、どちらが操作しやすいか(操作容易性)、どちらが目的のファイルまで辿り着きやすいか(ファイル到達性)の3点について、5段階での比較を行った。また、利用した感想なども自由記述でアンケートとして収集した。

評価対象とした本システムの現行バージョンでは、ファイル名およびディレクトリ名から自動抽出したキーワードのみを検索に用いており、キーワードの完全一致による検索とAND検索の機能を実装している。

4.2 評価結果

アンケートに基づく評価比較結果を図3に示す。

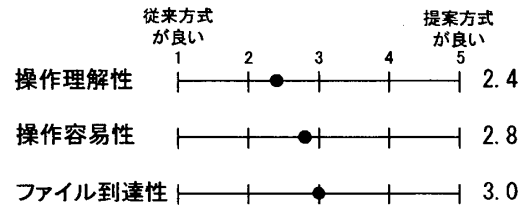


図3 比較結果

4.3 考察

最近ではウェブ検索などが頻繁に行われ、キーワードによる検索方式にも慣れていていると思われるが、提案方式は受け入れられにくい結果となった。

操作理解性および操作容易性については、被験者がExplorerの操作に慣れているための結果と考えられる。本システムを使うためには、予めある程度のシステム概念、操作説明を行う必要がある。また、操作補助となる機能についても検討し、実装しなければ実用としては難しいと考えられる。

ファイル到達性については、「キーワードさえ分かるならどこにフォルダがあってもファイルが見つかる」という感想もあり、キーワードに関する補助が重要であることが浮き彫りになった。

5. おわりに

本稿では、タグ情報に基づくファイル管理システムを提案し、設計を行った。プロトタイプを実装し、ユーザに試験利用をしてもらい、簡易的な評価を取った。

今後は、クライアントプログラムの操作面の実装を詰めるとともに、キーワードに関する補助機能を検討し、実装および評価を行っていく。具体的には、対訳検索やキーワードの意味ネットワークの構築、活用などを考えている。

参考文献

[1] David K Gifford, Pierre Jouvelot, Mark A. Sheldon and James W. O'Toole, Jr., "Semantic File System", 13th ACM Symposium on Operating Systems Principles, October 1991.