

関連性を用いた情報活用支援サービスの提案と実現法

Proposal of Information Utilization Support System for Smart File Retrieval Services
by Relativity between Files

山岸 優也*¹ 服部 峻*² 税田 竜一*³ 亀田 弘之*²
Yuya Yamagishi Shun Hattori Ryuichi Saita Hiroyuki Kameda

*¹東京工科大学大学院 バイオ・情報メディア研究科 コンピュータサイエンス専攻
Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology

*²東京工科大学 コンピュータサイエンス学部
School of Computer Science, Tokyo University of Technology

*³日本工学院八王子専門学校 ゲームクリエイター・ゲームソフトウェア科
School of Game Creator, School of Game Software, Nihon Kogakuin College of Hachioji

As performance of personal computers is getting upgraded, various kinds and huge amounts of information is now on the Internet. Even if a person has data on his/her own personal computers, it might be difficult to access it promptly, and as the worst case may be, its existence itself is forgotten.

This paper proposes a new information utilization support service with use of files' meta-data, which adopts a new advanced retrieval method. In this retrieval method, file itself is regarded as a "key object", and information included in or relating to the file, ie., from content to meta-data of files is affirmatively utilized to retrieve heuristically. A prototype system was implemented on a personal computer, and its fundamental performance was confirmed in a constraint setting on the computer.

1. はじめに

パーソナルコンピュータ (PC) の性能の飛躍的な向上に伴い、現在多種多様な情報が溢れている。そしてこれらの情報を個人ですら大量に所持し利用するようになってきている。自らが所持しているデータであっても、「必要になった時」に迅速にアクセスをすることが困難だったり、場合によってはその存在すら忘れてしまうこともある。情報は利用する時に利用できないのであれば「無い」と同義であると筆者は考える。こうした状況の中、多くの人が「情報」を活用し易くすることを目指して研究を行っている。

本稿では、ユーザに対しての想起支援にもつながる、発見的探索として従来の方式に囚われない“オブジェクトベース”な情報活用方式とそれに基づく情報活用支援サービスシステムについて述べる。

2. ファイルの管理 (整理・検索) の現状と問題点

今日のパーソナルコンピュータ (以降 PC と表記する) のハードウェア・ソフトウェアは様々なことが出来るようになってきている。デジタルな「情報」を手軽に扱え、保存できるファイル数が膨大に増えてきている。

- マルチメディア (音楽・動画・画像) を大量に記憶媒体へ保存・管理・再生をする。
- 「ツール」としての利便性の向上。
- 周辺機器の性能向上。

連絡先: 山岸 優也, 東京工科大学大学院 バイオ・情報メディア研究科 コンピュータサイエンス専攻, 東京都八王子市片倉町 1404-1, 042-637-2111, 042-637-2112, yamagishi.cs.ishihara@gmail.com

- インターネットの普及による情報量の増加

上記のように様々なことが出来るようになったのはとても良いが、その分弊害も出てきている。本章ではそれらの問題点について、及び既存研究・本研究のテーマについて述べる。

2.1 「データの洪水」問題と従来の取り組み

PC の記憶媒体には様々な「情報」が保存、活用されているが、未だに出来ないこともある。

現在の検索では PC 内に保存してあるデータを検索する際にはファイル名・ディレクトリ名を探索する形としての「文字列」検索、作成したアプリケーションや形式で探す「拡張子」検索、作成した日や更新日などの日時での「時刻」検索、これらを複合させた検索方法がある。

Windows OS の標準で搭載されている検索機能として、図 1 では上記の「文字列」検索、ファイルサイズを指定した検索、日時指定した検索をすることができるようになっている。これらの条件を複数指定することで目的のデータを見つける手法になっている。「文字列」検索では検索キーワードを「文字列」とし、正規表現のパターンでキーワードと同じ「文字列」があったファイルを検索結果として表示している。

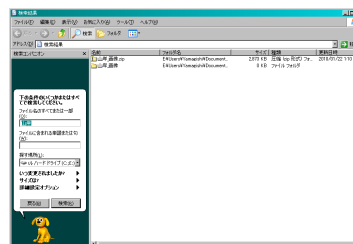


図 1: Windows での検索

もちろん様々な研究者が増大する情報を整理したり、うまく

検索するという願望を解決しようと研究に取り組んでいる。

近年には Google 社が開発したアプリケーション「Google デスクトップ」というインターネットと連携したファイル検索アプリケーションがある。図 2 にアプリケーションを示す。「Google デスクトップ」の特徴は PC 内にあるファイル・データの情報を DB (インデックス) に保存し、取り出せるようにしていることである。テキストファイルなどは書かれている内容も登録されているので、ファイル名には検索キーワードを含まないが、ファイルの内容中にキーワードがあった場合にも検索結果として出力されるようになっている。

他にも Web での検索履歴、電子メールの情報まですべてをまとめてくれるため、データ以外にも様々な情報を検索することができる。

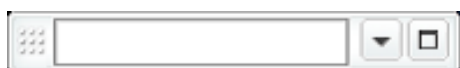


図 2: Google デスクトップ

上記のように Google デスクトップのような手法や、ユーザのスケジュールに基づいたファイル整理の方式 [松原 2007] やアクセス履歴を利用した管理機構 [小林 2006] など提案されている。

しかしこれらの既存手法では以下のような場合に十分に対応出来ていない。

- あの時使ったあのファイルってどこに置いたかたを忘れた
- 検索したいが、何というファイル名にしたか忘れた
- 今よりも簡単にファイルを整理することができない

こうした問題点を、どう解決していくのが今後は必要になってくるはずである。「データの洪水」[野口 1993] という言葉で、この手軽に扱えて便利だが情報増加している事を筆者らは呼んでいる。

「検索」とはそのまの意図ならば、必要な事柄を探し出してくること、というのが辞書での意味となる。現在の PC での検索とは「記憶媒体に記録されている情報を見つけ出してくる」と定義することができると思う。

ただし IT 用語での検索では人がモノを探す検索とは少し違う意味を持っている。デジタルデータは 0 と 1 のビットパターンにより構成されているので、検索というとこの同じビットパターンを探し出すということというのが「検索」となっている。

ここに人間の「検索」と PC での「検索」の差が出てきてしまい、目的のものを探そうと思ったときに素早く見つけられなくなる。最悪の場合見つからないという事も出てくる。

- ファイル名・キーワードが思いつかず探せない
- 保存した場所の忘却
- 保存時のミス (名前の綴りミス・意図しない場所への保存)
- システム不良による事故

2.2 オブジェクトベース検索方式

従来の情報検索では主としてファイル名や頻出単語などをキーワードとして文字列検索する方式 (キーワード検索) が多く用いられている。本研究では「単語」だけでなく、例えば「ファイル」そのものを検索のキーとする検索方式 (オブジェクトベース検索方式) を提案する。オブジェクトベース検索とは、現在主流であるキーワード検索とは異なる検索手法の呼び名であり、ファイルを「キー」として検索をかけることのできるものである。

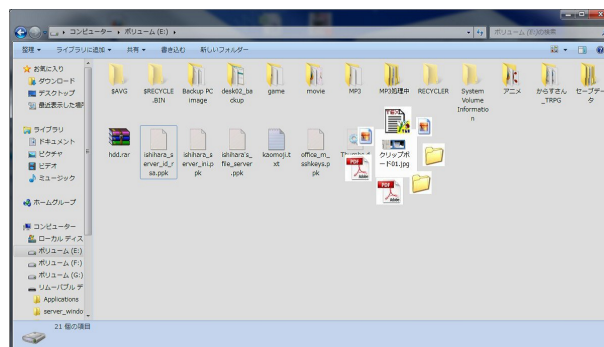


図 3: オブジェクトベース検索方式のイメージ

図 3 は「クリップボード 01.jpg」というファイルをキーにして検索をかけ、そのアイコン周りに検索結果を表示するというオブジェクトベース検索の利用イメージである。

3. 情報活用支援サービスの提案

本章は前章で述べたオブジェクトベース検索方式を含めた情報活用支援サービスについて述べる。

本サービスの位置付けは OS のファイル管理機構の上に覆いかぶさる形を目指している。

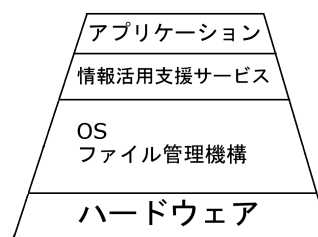


図 4: 情報活用支援サービスの位置付け

利用者がデジタル情報を OS の管理機構や様々なフリーソフトを活用して独自に整理・検索をしているが、本サービスは整理・検索・表示などを一括で行う事が出来、且つ今までの OS のディレクトリ構造などを変えずに活用していくことが出来るものである。これを「情報活用支援サービス (IUSS: Information Utilization Support System)」と筆者らは呼ぶ。

ファイル間に存在する様々な「関連性」を積極的に用いるようにし、このシステムの一層の重要な部分となる。

3.1 関連性とは

図 5 は「関連性」の例を示しており、黒い線で結ばれているのは従来のディレクトリ構造の繋がり、点線で繋がれている線が本研究で取り扱う新たな「関連性」である。

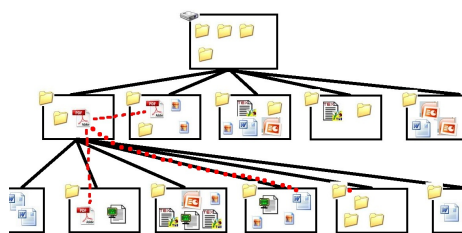


図 5: 「関連性」のイメージ

この「関連性」はデジタルデータに付加されている「属性値」を利用して構築している。今日の PC で使われる OS のファイル管理機構には属性値をデジタルデータに付加しているため、利用しやすく、さらに利用者が新たに属性値をつけたりする事が出来るという利用者の利用状況を取得しやすいもののため本研究で採用した。

本研究で用いる属性値は以下の 6 項目である。

1. ファイル名 (ファイル名のみ・拡張子含み)
2. ディレクトリ名
3. 保存されているパス (相対パス・絶対パス)
4. 時刻 (作成日時・更新日時)
5. ファイルタイプ (拡張子)
6. サイズ

例えば検索対象「/home/yamagishi/docs/siei_doc.tex」の属性値は以下ようになる。

表 1: 属性値例

属性名	値
ファイル名	siei_doc.tex
ディレクトリ名	docs
保存されているパス	/home/yamagishi/docs/
ファイルタイプ	tex

こうした上記の属性値から例えばだが、更新日・作成日が近いものというのは作業時に開いて利用していた、つまり何らかの「繋がり」があったと見ることが出来る。

3.2 オブジェクトベース検索の実現方法

デジタルデータには様々な情報が記録されている。その情報を属性値 (プロパティ値) と呼ぶが、本研究ではこの「属性値」を活用し「関連性」を見つけ出す。「関連性」を見つけるにあたり拡張子、時間 (作成・更新日)、ディレクトリパス情報、ファイル名、ディレクトリ名、拡張子、サイズの「属性値」を利用した。この「属性値」から、従来のシステムでは見つかることの出来ない繋がりが構築され、これを利用することにより、目的の情報を見つけるだけでなく、ユーザ自身が忘れてしまっているファイルでも再発見をすることも可能になる。

4. 情報活用支援サービスシステムの実装

図 6 は情報活用支援サービスの概要図である。本来のサービスは整理・検索などを行うサービスなのだが、本プロトタイプは「検索」を行うことだけに機能を絞っている。

ユーザが GUI を介してファイルを指定すると、関連性フィルタがそのファイルの属性値を取り出し、データベースに照会する。その結果、「関連性」があるファイルやディレクトリが検索結果として GUI に表示される。なお、PC 状態監視機能とファイル取得機能は事前に関連情報を取得するための機能である。開発には以下のような環境で行った。

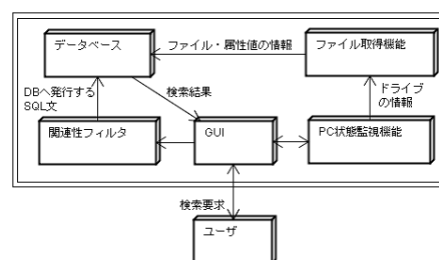


図 6: 概要図

- 開発 OS : Microsoft WindowsXP Professional Service Pack 3
- 開発言語 : JRuby 1.1.5 (ruby 1.8.6 patchlevel 114) (2008-11-03 rev 7996) [x86-java] and Java JDK 1.6.0_11
- DB : SQLite3
- 統合開発環境 : NetBeans IDE 6.1
- CPU : Intel Core2Duo T8300 2.4GHz
- メモリー : 3072MB
- HDD : 160GB

4.1 GUI

ユーザへ情報提供及び操作部の提供をするモジュール。検索結果の表示形式はリスト表示となっている。すべて表示してしまうと情報の量が多くなってしまったため、分けて表示する方式を採用した。

GUI 部を考えた際、わかりやすく・見やすいことを念頭に入れ考えた。キーワード入力部、結果表示部として分け構築した。構築には Java の Swing, Awt を利用し, JRuby 上から呼び出している。

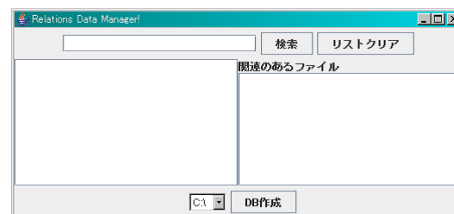


図 7: GUI の構築

4.2 データベース

記憶媒体に保存されているデジタルデータの属性値が保存されている。データベースには SQLite3 を利用している。

4.3 関連性フィルタ

本プロトタイプシステムではSQLite3をデータベースとして利用しているため、SQL文を発行するモジュール・時間であれば、選択されたファイルの作成日・更新日を取得して、前後1時間以内のファイルを探すようにするSQL文を選択してDBへ入力している。

4.4 PC 状態監視機能

接続されている記憶媒体を取得するモジュール・Javaに備わっている「File」クラスに存在するドライブ情報を取得するメソッドを利用して、検索する記憶媒体の情報を取得する。

4.5 ファイル・ディレクトリ取得機能

デジタルデータに付加されている属性値を取得するモジュール・JRubyに備わっている機能を用いて、3.で挙げた属性値を取得をする。

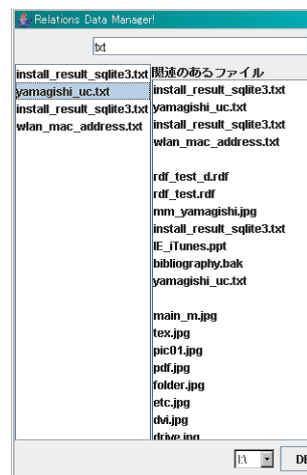


図 8: キーワード 1 の関連ファイル

5. プロトタイプシステムの動作例

上述した考えに基づき、構成したプロトタイプシステムの動作例を示す。検索条件は以下の表 2 の環境となる。

表 2: 検索条件

検索する場所	USB メモリー (256MB)
全ファイル数	329
全ディレクトリ数	32

USBメモリーのトップディレクトリには「test」「school」「music」の3つディレクトリを作成した。そして各ディレクトリにも数個から100前後のファイル・ディレクトリが保存しており、ファイル名・ディレクトリ名はすべて英数記号にて構成している。

プロトタイプシステムでは、2回に分けての操作で「関連性」の高いファイルをリストへ表示するようにしている。最初にキーワード検索にて「キー」にしたいファイルを左のリストから探し、次に「キー」にするファイルを選択すると「関連性」の高いファイルを検索することが出来るようになっている。

検索キーワードとして「txt」を入力。左のリスト部に「txt」の含まれているファイルが表示される。左リスト中の「キー」にしたいファイルをクリックすることで「関連性」の高いファイルが出力される。表示は、同形式ファイル、同ディレクトリファイル、更新時刻が近いファイル、作成時刻が近いファイルの順で出力される。「yamagishi_uc.txt」をクリックした結果は図8のようになっている。

キーワードとしてファイル名・ディレクトリ名・拡張子と全てに入りそうな単語にて検索した。そして出てきたファイルと関連しているファイルを表示することができた。

6. 今後の展開

プロトタイプシステムの作成までを行ったが、改良の余地が残っている。

・UI (ユーザーインターフェース)

検索をする際に「オブジェクトベース」の重要なGUIでの検索が出来ていないが、これは技術的な問題点のため早急にご利用ができるように作成をする。

・関連性の保存方法及びリアルタイムでの監視

現在はSQLite3を利用した環境だが、変更があった場合などにカラムへの書き換え作業が必要となってしまうリアルタイム

での整理・検索には向かないため、新しくリアルタイムで使える保存方法を考案する必要がある。

・日本語に対応

JRubyが日本語に対応していないため、この問題も早めに解決しないとイケない。

7. おわりに

本稿では関連性を用いた情報活用支援サービスの提案と実装について述べた。

参考文献

- [野口 1993] 野口 悠紀雄, “「超」整理法”, (1993)
- [松原 2007] 松原 靖子, 小林 一郎, “ユーザーのスケジュール情報とPC内データ間の相関関係を利用したデスクトップ検索システムの開発”, (2007.7)
- [小林 2006] 小林 大, 田口 亮, 横田 治夫 “アクセス履歴とデータライフサイクル情報を組み合わせたストレージ管理機構”, (2006.7)

発表文献

- [山岸 2010] 山岸優也, 亀田弘之, “関連性を用いた情報活用支援サービスの提案と実装”, 電子情報通信学会, 電子情報通信学会 全国大会 ポスターセッション (発表者: 山岸) 2010年3月18日発表
- [山岸 (2) 2009] Yuya YAMAGISHI, Chiaki KUBOMURA, Ryuichi SAITA, Hiroyuki KAMEDA, “Proposal of Information Utilization Support System for Smart File Retrieval Services by Relativity between Files”, ITCA2009 (発表者: 山岸) 2009年12月12日発表
- [山岸 (1) 2009] 山岸優也, 税田竜一, 久保村千明, 亀田弘之, “アナロジーを用いた情報活用手法の提案と実装”, 電子情報通信学会, TL2008-59, 思考と言語, pp51-54 (発表者: 山岸) 2009年2月6日発表