

分散ファイル群の高度管理を目的としたミドルウェアの提案

Proposal of Middleware for the Advanced Management of Distributed Collections of Files

三森 祐一郎[♡] 森嶋 厚行[◇]

Yuichiro MITSUMORI Atsuyuki MORISHIMA

近年、計算機に格納されるファイル数は飛躍的に増大し、ファイル群の管理が困難になっている。本稿では、分散して管理されたファイル群の管理するためのミドルウェアの提案を行う。特徴は、既存のファイルシステム操作系との互換性を維持すること、および、大量のメタデータを保存・利用することにより高度なファイル群管理の実現を目指すことである。

Recently, the number of files stored in computers is rapidly increasing, and managing files is becoming a very difficult task. This paper proposes a middleware to support the management of distributed collections of files. Its features include the compatibility with the current file systems, and the preservation and utilization of a large amount of metadata for realizing the advanced management of files.

1. はじめに

近年、計算機による情報処理が日常のものになり、計算機が格納するファイル量が飛躍的に増大している。コンピュータネットワークや各種デバイスの発達により、これらのファイルは複数の機械に分散して格納されていることが一般的である [1]。例えば、ある大学の研究室のファイルサーバに格納されているファイル数は約 20 万であり、また、それらのファイルのコピーや関連ファイルはファイルサーバ内に留まらず、研究室の構成員のノート PC をはじめとして様々な機器に分散して格納されている。その結果、例えば、あるプロジェクトに関連するファイルはどこに散らばっているのか、あるファイルの最新バージョンはどれか、バックアップの管理はどうすればよいのか、など、計算機に格納されているファイル群の管理はますます困難になっていると言える。このような状況にもかかわらず、ファイルシステムの基本機能は、階層ファイルシステムの登場以来ほとんど変わっていない。近年、Google デスクトップサーチ [2] などのデスクトップサーチシステムの登場により、ファイルの検索は高速に行なえるようになったが、ファイル群の管理的側面に関してはあまり注目されてこなかった。

本稿では、このように分散管理された大量のファイル群の管理を支援するためのミドルウェアの開発について述べる。本ミドルウェアは、今後、ファイル群管理のために必要とされるコンピューティングリソースとアプリケーション開発コストの増大に対応するために、全てのアプリケーションから利用可能な汎用性の高いソフトウェア基盤を開発しようとするものである。これは、過去に計算機処理においてユーザインターフェースに必要とされるコンピューティングリソースと開発コストの増大に対応するために、各種 GUI ライブラリが開発されたことに類似する動機である。本ミドルウェアの特徴は、(1) 分散管理されている多量のファイル群の管理機能を提供する共通ソフトウェア基盤を目標としているこ

[♡] 学生会員 筑波大学大学院 図書館情報メディア研究科
mitsu@slis.tsukuba.ac.jp

[◇] 正会員 筑波大学大学院 図書館情報メディア研究科
mori@slis.tsukuba.ac.jp



図 1 情報空間の例
Fig. 1 Example Information Space

と、(2) 既存のファイルシステム API と互換性を持ち、既存のアプリケーションでも機能の一部を利用可能なこと、である。

本稿では、このミドルウェアの概要、API 設計、ミドルウェアの実装について説明する。

2. コミュニティ情報空間ガバナンスプロジェクトの概要およびミドルウェア開発の動機

現在、我々が進めているコミュニティ情報空間ガバナンスプロジェクト [4][5][6] では、各クライアント PC やファイルサーバで分散管理されたファイル群の管理を行うフレームワークである、InfoSpace Governor を開発している。

本フレームワークの第一の特徴は、個々のファイルシステムではなく、コミュニティ情報空間 (Community Information Spaces, CIS) と呼ぶ複数のファイルシステム群を含む空間を管理の範囲としていることである。図 1 はある大学の研究室における CIS の例である。一般に、CIS には複数の計算機が含まれており、それぞれが情報を分散して管理している。例えば、この研究室で行われているプロジェクトに関連する情報は、ファイルサーバと参加メンバーの PC に分散されて管理されている。それらに管理されている情報は、お互い関連しているものの、それらはシステムレベルでは明示的には関連づけられていない。例えば、そのプロジェクトに関連するあるファイルの最新バージョンはどれか、といった情報や、そのファイルが参照するファイルはどれか、といった情報をシステムはもっていない。したがって、これらの間にシステムが答えることは出来ない。ネットワークを通じてこれらの情報源が物理的には接続されているにもかかわらず、先に述べたような関連のレベルでは実は接続されていないのである。

本フレームワークの第二の特徴は、これらの関連を明示的に表すメタデータを保持することにより、情報空間の統治の支援を行う事である。具体的には、今説明したような情報資源間の関連を明示的に保持するための InfoSpace Server と呼ぶメタデータ DB を利用し、コミュニティ情報空間の管理を行う (図 2)。このメタデータ DB では、RDF[3] によるラベル付きエッジを持つグラフ構造によってファイル間の関連を保持している (図 4)。ここでは、ファイル、ディレクトリ、プロセス、ユーザを RDF グラフのノードとして表現する (以下では DSL ノードと呼ぶ)。

このようなメタデータは、アプリケーションが直接メタデータ DB と通信することによって登録・利用することも可能であるが、あくまでもデータを保持しているだけであり、このメタデータを利用した機能は個別に実現しなくてはならない。本稿で提案するミドルウェアは、あらかじめこれらのメタデータを利用して高度なファイル管理機能を実現するための機能を実装したライブラリを用意し、アプリケーションから容易に利用できることを目指すものである。

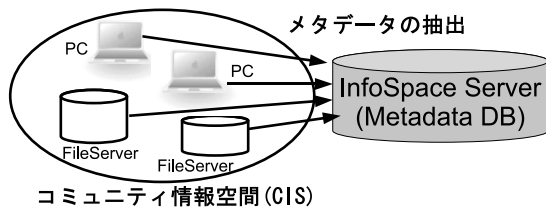


図 2 InfoSpace Governor のアーキテクチャ
Fig. 2 Architecture of InfoSpace Governor

Existing Applications	New Applications
InfoSpace Layer	
Metadata DB	File Layer

図 3 提案ミドルウェアの位置付け
Fig. 3 Position of the proposed middleware

3. 提案ミドルウェアの概要

本ミドルウェアは、前章で説明したメタデータ DB とファイルシステムの上に、新たなレイヤを構築する (図 3)。本ミドルウェア・アーキテクチャの大きなポイントの一つは、ファイル操作関係の既存 API と互換性のある形で実現を行うことである。これにより、既存のアプリケーションからでもミドルウェアの機能の一部を利用可能にする。

ミドルウェアの機能は次の 3 つである。(1) メタデータの入手: アプリケーションや利用者によるファイル操作を監視し、ファイル操作の履歴などのメタデータをメタデータ DB に格納する。具体的にはファイル、プロセス、ユーザに対応する DSL ノードを RDF のリソースとして表現し、それらの関連やプロパティなどを保存する。例えば、あるプロセスでどのファイルが読み込まれ、書き出されたかという情報が保存される (図 4)。これらに関しては 6 章でより詳細に説明する。(2) メタデータへの問合せ: メタデータ DB に格納された各種メタデータへの問合せを行う。大量のメタデータへの問合せを高速に処理するため、専用の問合せ処理エンジン [4] を利用する。(3) 高度ファイル操作・管理機能: メタデータ DB 中のデータを利用して高度なファイル操作・管理機能を提供する。例えば、最新バージョンファイルの発見や、ファイル自動内容更新、異種形式ファイルのビュー作成などの機能である。

これらの機能を利用する方法は複数ある。(1) アプリケーションからの利用: 既存のライブラリを拡張した形のインターフェースを利用する。例えば、Java 言語における java.io パッケージ等の代わりに、互換性のある infospace.* パッケージを利用する。(2) OS での対話型ファイル操作からの利用: OS のコマンドラインインターフェースや GUI 型のインターフェースについては、ミドルウェアと通信するソフトウェアを常駐させ、それを通じて操作のフックの実行やミドルウェアの機能の利用を行なう。

4. 高度ファイル操作・管理の機能例

提案ミドルウェアの利用に関する具体的なイメージを示すため、幾つかの具体的な機能例を紹介する。現段階では必ずしも全ての機能は実装できていないが、本提案ミドルウェアを利用することによって実現可能になると考えられる機能である。

(1) 関連ファイル、コピーファイル、最新バージョンなどの発見: メタデータ DB に格納されている関連を利用して、あるファイルに関係するファイルを発見する機能である。より具体的な例としては、既に削除してしまったファイルが必要になったとき、情報空間に含まれる他の人のノート PC にそのファイルのコピーがあることを発見すること、などがある。また、ファイルをオープンするときに、実はそのファイルにはさらに新しいバージョンが存在する場合、それを報告する機能が実現できる。これらの機能は、OS におけるインタラクティブな操作や新しく開発するアプリケー

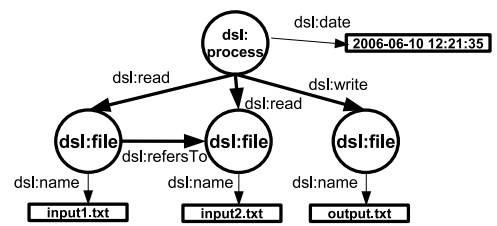


図 4 プロセスとファイルの関係
Fig. 4 A relationship among processes and files



図 5 WindowsXP でのファイルオープン
Fig. 5 Open file dialog in WindowsXP

ションだけではなく、既存のアプリケーションからも一部利用できる。この利用は、ファイルオープン用のウィンドウ (図 5) を差し替えた拡張ファイルオープンウィンドウを通じて行う。

(2) セマンティックバックアップ: ファイルのバックアップはディレクトリ単位で行われることが通常であり、アプリケーションが利用しているファイル群の単位とはミスマッチしている。本ミドルウェアを利用することにより、アプリケーションの利用において関連するファイルの記録が残るため、これらを単位としたバックアップが出来る。もう一つの利用は、アプリケーションを超えたバックアップの再利用である。通常、アプリケーションが自動バックアップ機能などによってファイルのバックアップを行った場合、その情報は異なるアプリケーションでは共有されない。そのため、あるファイルの以前のバージョンのバックアップが存在したとしても、それをユーザが知らなければ他のアプリケーションからはそのようなファイルを利用できない。本ミドルウェアを利用することにより、他のアプリケーションからもほかのアプリケーションが作成したバックアップを参照できる。

(3) 自動更新コピー: コピーの際に、内容同期に関する制約を指定出来る。具体的には、マスタとスレーブの関係指定することにより、スレーブと指定されたファイルの内容がマスタファイルの内容に同期するように設定される。これは、事務書類のフォーマットが年度ごとに変更されるような場合に有用である。

(4) ファイルの関連に基づく警告: 例えば、あるファイルを、ファイルサーバから個人の PC にコピーをおこなう操作を行うと、それ以外のファイルをコピーする必要がないかどうかを調べ、必要なファイルのコピーを推薦する。また、ファイルを削除する場合、それを参照する別ファイルが存在する場合には削除の警告を行う。

(5) 異種形式ファイルのビュー: ある形式のファイルを別の形式に変換してから処理したい状況は多く存在する。例えば、CSV 形式のファイルを XML 形式に変換して処理する場合などがあげられる。通常、そのような場合には変換ソフトウェアなどを利用して一時的なファイルを作成する。しかし、そのような一時的なファイルを生産することはファイル管理の観点から望ましくない。また、アプリケーション側で対応するためにはそのための機能を個別に実装する必要がある。提案ミドルウェアでは、この問題への解として、幾つかの形式に関して異種形式のファイルビューを用意する。これにより、一時的ファイルを作成することなく、アプリケーションが異種形式のファイルを容易に扱える。

(6) 各種制約の設定と制約維持の監視: ファイルやディレクトリに

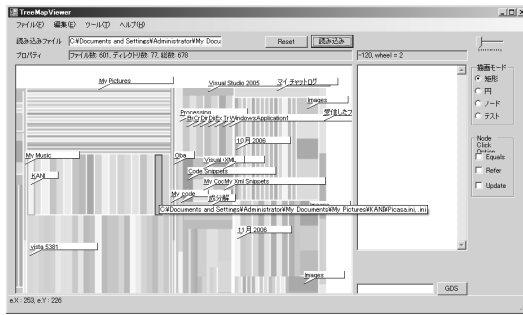


図 6 InfoSpace Maps
Fig. 6 InfoSpace Maps

関する制約を指定し、違反があったとき、もしくは違反の可能性がある時にアクションをとる。例えば、あるディレクトリに特定のファイル形式以外のファイルやサブディレクトリを追加することを禁止するという制約を設定しておく、違反があった場合にその場で警告をしたり、あらかじめ設定しておいたメールアドレスにメールを送ったりする。

(7) 新しいアプリケーションの開発。InfoSpace API が提供する機能を用いて、新しいアプリケーションを開発することが出来る。この場合、本ミドルウェアが提供する機能をフルに活用することが出来る。例えば、InfoSpace Maps (図 6) [5] は、コミュニティ情報空間の視覚化ツールである。これは、コミュニティ情報空間に含まれるファイル群とそれらの関連を視覚化する。利用者は、ファイル群をそれらの関連を通じてブラウジングすることが出来る。

このような、ファイル間の関連等を利用するアプリケーションの構築が、InfoSpace API を通じて提案ミドルウェアを利用することにより可能になる。本ミドルウェアが高度ファイル操作を実現するライブラリを提供するため、アプリケーションの開発者が個々に機能を実装する必要はなく、アプリケーションプログラムのコード量の削減に寄与できる。

5. InfoSpace API の設計

本章では、アプリケーションからの提案ミドルウェアの機能を利用するための API である、InfoSpace API の設計について説明する。この API は特定の言語とは独立であるが、議論を具体的にするため、ここでは Java 用のパッケージを参照モデルとし、特に重要な java.io 以下のファイル入出力にクラスに関する議論を行う。

InfoSpace API を java.io 互換で実現するためには、(1) 既存のクラスの拡張と、(2) 新規のクラスの追加、が共に必要である。しかし、互換性を重視するためには、(2) を最小限にしたうえで高度な機能を利用可能にすることが望ましい。本フレームワークでは新規で追加するクラスは InfoSpace クラスのみである。

(1) 拡張クラス：File クラス以外。拡張クラスのうち、File クラス以外のクラスについては、メソッドの変更はない。ただし、java.io のクラスが持つ機能に加えて、それらを実行することによって分かるファイルやプロセス間の関係をメタデータ DB に格納することが行われる。例えば、FileReader、FileWriter などのクラスにおいて、ファイルの生成や入出力が行われた場合、メタデータ DB にそのプログラムを実行しているプロセスノードを生成し、ファイルノードとの入出力の関係を追加する。また、入出力でつながったファイルの内容が同じ場合は、copy ラベルを持つエッジでファイルノードを接続する。

(2) 拡張クラス：File クラス。他の拡張クラスと同様、File クラスのメソッドは、処理に応じてメタデータ DB にファイル間の関連などを追加する。ただし、他の拡張クラスと異なり、新規で追加されるメソッドが幾つか存在する。追加されるメソッドの機能は、メタデータ DB に格納されたファイルノードの属性を調べるものと、次に述べる InfoSpace クラスの機能を使って高度ファイル操作を行なうものがある。

(3) 追加クラス：InfoSpace クラス。これは、高度なファイル操作を行う機能を提供する。InfoSpace クラスで実現される機能について図 7 で紹介しておく。メソッドは 3 つに分類できる。(a) メタデータ DB への問合せメソッド (getXXX メソッド) (b) メタデータ DB の更新メソッド (c) データ形式変換メソッド (view メソッド) (d) 制約メソッド (setConstraint メソッド) (e) (a) ~ (d) を内部的に利用して高度機能を実現するメソッド。

InfoSpace クラスのメソッドは、拡張クラスのメソッドからも内部的に利用される。例えば、FileReader や FileStreamReader の拡張クラスでは、オプションで指定することにより、CSV ファイルを XML ファイルとして開く、といった機能を提供するが、それは内部的には (c) のメソッドを呼び出すことにより定義されている。

プログラム例。提案 API を用いたプログラムの例を図 8 に示す。このプログラムは、与えられたファイルの最新バージョンを開くためのプログラムである。例えば、複数のユーザによってある文書を共同で校正している場合、他の人が別の場所にコピーして編集したものが最終版であるという状況がありうる。このような時、最新版を発見し、入手するプログラムを作成することは困難であった。本ミドルウェアは、このようなよく利用されると考えられる機能を直接提供する API を用意している。したがって、プログラムはメタデータ DB などに直接アクセスせずに、これらの機能を組み込んだアプリケーションを簡潔に実現できる。

```
import infospace.*
public class sample {
    static void main(String[] args){
        File[] farr;
        File file = new File(args[0]);
        try{
            InputStreamReader input =
                new InputStreamReader(System.in);
            farr = InfoSpace.getUpdated(file);
            for(int i=0; i<farr.length; i++){
                System.out.println(i+" "+farr[i].filename());
            }
            System.out.println("Select a file to open:");
            String str = input.readLine();
            int n = Integer.valueOf(str);
            FileStreamReader output =
                new FileStreamReader(farr[n]);
            ...
        }catch(Exception e){
            ...
        }
    }
}
```

図 8 infospace パッケージを利用したプログラム例
Fig. 8 Example program using the infospace package

6. ミドルウェアの実装

ミドルウェアは、次の 3 種類の部分に分けて実装される (図 9)。

(1) 機能実現部は、本ミドルウェアの中心となる部分であり、5 章で説明したような機能を実際に提供する。(2) アプリケーション連携部は主に API を実装したライブラリとして実装され、アプリケーションプログラムの動作からメタデータを入手したり、アプリケーションと機能実現部のインタフェースとして動作する。(3) ユーザインタラクション部は主に OS のコマンドラインや GUI への追加モジュール等で実装され、OS レベルでのユーザによるファイル操作からメタデータを入手したり、機能実現部が提供する機能のユーザへの提供を行う。

6.1 機能実現部の実装

メタデータ DB への格納。機能実現部は、アプリケーション連携部およびユーザインタラクション部からの要求に従い、メタデータ DB にデータを格納する。メタデータ DB への格納データは DSL (Document Space Language) [6] で記述する。DSL とは、RDF 上にファイル群管理に必要なクラスや関連を表現するポキャブラリを独自に定義した言語である。DSL では、リソースとして、ファイル、プロセス、ユーザなどを持つ (図 10)。

type	メソッド名	説明
(a)	File[] getChildren(File f, int DSL_PROPERTY)	ファイル f の関連 DSL_PROPERTY である子ノードの配列を得る
	File[] getDescendants(File f, int DSL_PROPERTY)	ファイル f の関連 DSL_PROPERTY である子孫ノードの配列を得る
	File[] getParents(File f, int DSL_PROPERTY)	ファイル f に関連 DSL_PROPERTY である親ノードの配列を得る
	File[] getAncestors(File f, int DSL_PROPERTY)	ファイル f に関連 DSL_PROPERTY である祖先ノードの配列を得る
	int getType(File f)	ファイル f のノードタイプを返す
(b)	boolean makeNode(File f)	ファイル f のノードを生成する
	boolean setRelation(File f1, File f2, int DSL_PROPERTY)	ファイル f1 からファイル f2 へ、関連 DSL_PROPERTY をつける
(c)	File view(File f, int VIEW_TYPE)	ファイル f を VIEW_TYPE の形式に変換した一時ファイルをビューとして返す
(d)	boolean setConstraint(Cond cond, Act act)	制約 cond に違反した場合、act を実行することを要求する
(e)	boolean msCopy(File f1, File f2)	ファイル f1 をマスターとして、スレーブコピーを f2 に行う
	File[] getUpdated(File f)	ファイル f もしくはそのコピーの派生の末端である DSL ノードの配列を返す
	boolean backup(File f, URI uri)	uri が示す領域にコピーを作成し、バックアップの関連をつける

図 7 InfoSpace クラスのメソッド

Fig. 7 Methods of InfoSpace class

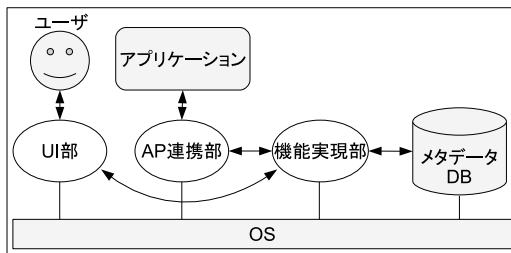


図 9 ミドルウェアの実装

Fig. 9 Implementation of the middleware

高負荷処理の遅延実行．提案ミドルウェアではファイル操作のパフォーマンスを落とさずに高度機能を提供する必要がある．そこで、処理の一部はファイル操作後に遅延して行うよう実装し、応答時間の短縮を図る．例えば、同じプロセスから入出力された 2 つのファイルの内容が同一である場合、メタデータ DB に格納されているこれらのファイルを表すノード間に copy という関連を追加する必要があるが、このファイルの同一性のチェックは時間がかかる．そこで、同一性のチェックはすぐには行わずに制御を返し、その後バックグラウンドで実行するなどの工夫を行う．

クラス	説明
dsl:dir	ディレクトリを表すクラス
dsl:file	ファイルを表すクラス
dsl:symbolicLink	シンボリックリンクを表すクラス．
dsl:phantom	存在しないが参照されるディレクトリもしくはファイルを表すクラス
dsl:process	プロセスを表すクラス
dsl:user	ユーザを表すクラス

図 10 DSL で用意するクラス

Fig. 10 Classes defined in DSL

6.2 アプリケーション連携部

アプリケーション連携部は、InfoSpace API を実装したモジュールとして実装される．本モジュールの主要な機能の一つは、API に基づくファイル操作にしたがって、適切なメタデータを入手し、機能実現部に送信することである．特に、アプリケーションの実行時に読み書きしたファイルの関連を作るために、アプリケーションの実行開始時にジョブノードを作成し、ファイルの読み書きに応じて read および write 関連の作成を行う必要がある．

6.3 ユーザインタラクション部

ユーザインタラクション部の実装では、GUI および CUI によるファイル操作の情報を入手しなければならない．その実現方法は幾つか考えられるが、現在は下記の方法で実装を進めている．**Windows GUI**．Windows GUI を対象としたユーザインタラクション部の実現のために、C#のライブラリを用いて実装を行っている．コマンドラインインターフェース．Unix の Shell や Windows のコマンドラインなどの CUI インタフェースのシステムでは、基本コマンドの置き換えや追加によってユーザインタラクション部を実

現する．例えば、UNIX の cp や mv といった基本コマンドはメタデータ DB と連携するコマンドの実装と置き換える．また、高度ファイル操作・管理の機能を実現する新しいコマンドを追加する．

7. まとめ

本論文では、複数の計算機から構成されるコミュニティ情報空間に含まれるファイル群の高度操作・管理を実現するためのミドルウェアの提案を行った．本ミドルウェアでは、既存のファイルシステムと比較して大量のメタデータを保持し、それらを有効に活用する．今後の課題としては、提案ミドルウェアの実装の完成度の向上、パフォーマンスの検証、実利用におけるユーザビリティの検証、等があげられる．

[謝辞]

ゼミ等においてコメントいただきました、筑波大学大学院図書館情報メディア研究科 杉本重雄教授、阪口哲男准教授、永森光晴講師に感謝いたします．

[文献]

- [1] Daniel J.Gonçalves, Joaquim A. Jorge: An Empirical Study of Personal Document Spaces, Springer LNCS 2844, Interactive Systems: Design, Specification and Verification, DSV-IS 2003 Proceedings, pp 46-60 Funchal, Portugal, 4-6 June 2003.
- [2] Google. Google Desktop Search. <http://desktop.google.com/>.
- [3] W3C. Resource Description Framework (RDF). <http://www.w3.org/RDF/>.
- [4] 石川憲一, 森嶋厚行, 田島敬史: メタデータ DB のための到達ノード問合せと更新処理の効率化, 電子情報通信学会第 18 回データ工学ワークショップ (DEWS2007), 2007 年 3 月.
- [5] 望月祥司, 児玉麻莉子, 石川憲一, 森嶋厚行, 田島敬史: 大規模ディレクトリ空間視覚化のための論理構造抽出, 電子情報通信学会第 18 回データ工学ワークショップ (DEWS2007), 2007 年 3 月.
- [6] 石川憲一, 森嶋厚行, 田島敬史: 大規模ドキュメント空間管理のための意味ファイルシステムの構築, 電子情報通信学会技術研究報告, Vol.106, No.150, DE2006-115, pp. 139-144, 2006 年 7 月.

三森 祐一郎 Yuichiro MITSUMORI

筑波大学大学院 図書館情報メディア研究科博士前期課程在学中．情報処理学会学生会員，日本データベース学会学生会員．

森嶋 厚行 Atsuyuki MORISHIMA

筑波大学大学院 図書館情報メディア研究科/知的コミュニティ基盤研究センター准教授．筑波大学大学院 工学研究科修了．博士 (工学)．ACM, IEEE-CS, 情報処理学会, 電子情報通信学会, 日本データベース学会各正会員．