# 資料復元研究支援データベースシステムの開発(3)

## ——3つのプロトタイプシステムとその比較——

上　島　紳　一

## 1. はじめに

近年のコンピュータの記憶媒体の大容量化とアクセス技術の高速化に伴い，研究支援を目的とするデータベースシステムの形態が変化しつつある。従来は，資料や関連文献の2次情報を格納するシステムが中心であったが，資料の画像や文献の全文，画像などの一次情報をそのまま格納したシステムが増加している。また，研究者が研究を進める上で，格納された資料や文献の1次情報をそのまま用いて個人の用途に応じたデータを付与したり，組み合わせたりすることでデータをカスタマイズできる機構の実現要求が高まっている。

我々のグループでは，これまで資料の研究作業における様々な切り口に焦点を当てシステム構築してきた。これらのシステムは，単にデータの格納を目的としているのではなく，研究者が行う作業のためのワークベンチとして位置付けている。つまり，システムに資料を格納し，それをもとに研究の本来の大きな目的の一つである資料復元作業を行う作業台とすることを目指している。例えば，木簡研究においては，複数の木簡を集めて冊書復元を行ったり，2本の木簡を仮想的に繋ぐことにより，見えない文字を類推したりする場合の作業台である。

工学的には，このようなシステムは柔構造データベース，ハイパーメディアシステムといった新しい分野のシステムとしてみなすことができる。また，資料データを構造が確定していないデータと見れば，資料の復元作業は半構造化データの構造化作業に相当する〔23, 21, 12〕。このような格納されたデータを加工することを目的とするシステムでは，利用者の多様な視点から予め構造の予測できないデータを発見的に取り扱う必要があり，システムとして柔軟な構造を持つことが必要である。このため，我々は，従来型の関係型のデータベースではなく，オブジェクト指向技術を用いた新しいデータモデルを構築してプロトタイプシステムの開発を行っている。

ここでは3つのプロトタイプシステムの設計方針，システム構成，特徴などについて述べ，プロトタイプシステムを利用して研究する上での可能性と問題点などについて述べる。

# 2．研究支援システム

## 2．1　システムの分類

資料研究を支援する目的で構築されるシステムは次の５つの種類に分類できる。

・デジタルアーカイブ

資料の記録を目的として資料の写真やビデオ映像，音などをデジタルデータ化してデータベースに格納したシステムを一般にデジタルアーカイブと呼ぶ。データ入力作業では，通常のカメラによる撮像に加えて赤外線カメラで可視範囲を越えるデータを入力する技術などが用いられる。また，取り込んだデータに処理を行って曖昧な部分や見えにくい部分などを鮮鋭化する画像処理や音声処理などの関連技術も同時に用いられる〔17〕。

データの管理機構としては，関係データベース（以下，RDB），オブジェクト指向データベース（以下，OODB），ファイルシステムなどが用いられる。データへのアクセス法としては，キーワードなどの資料の２次情報を用いて資料をアクセスする形態と，類似画検索手法などのパターン照合技術を用いて資料の１次情報を直接アクセスする形態がある。形態に応じて質問言語や応用プログラムが検索に用いられる。また，これらのシステムを要素データベースとして，複数のデータベースを統合する連邦型データベース，データベース内のデータ間に関連付けを持たせたハイパーメディアシステムなどが用いられることも多い。

・資料計測システム

デジタルアーカイブに格納された資料を積極的に利用する手法として資料計測システムがある。例えば，画像処理・認識の技術を用いて遺物，彫物などの部分片を３次元計測し，得られた３次元データから原データを仮想的に生成するシステムが構築されている。これらは断片化した遺物の破片の復元作業のシミュレーションに利用できる〔２，３〕。

また，大量の新規の資料をアーカイブにデータ入力する際に効率的に資料を整理・収納する目的で作成されているシステムがある。〔１〕は資料を複数の角度から撮像すると同時に重量を計測するシステムで，入力されたデータから資料の形状，色，重量などの情報を自動的に得る。これらの情報は，データへのアクセス情報として用いられる。

・全文データベースシステム

資料研究のための基礎文献の全文をテキストデータとして格納しておき，テキスト検索エンジンで必要な箇所を抽出して利用するシステムである。ワシントン大学と台湾中央研究院の共同で開発された中国24正史の全文データベースが代表的である〔４〕。日本でも京

都大学人文科学研究所において「史記」,「続日本記」などの文献の全文データベースが構築されている。またそれらをもとに一字索引も作成されている。

・ワークベンチ

ワークベンチとは，研究者が資料を用いて研究を進める際に，机上で行う作業をそのまま計算機上で行えるような環境を備えたシステムをいう。作業には，資料やその部分に対するコメント書き，資料の比較・合成などの操作，また資料に対する利用者の視点の設定と視点に基づく資料の操作などがある。このようなシステムでは作業の過程で生成されるデータを格納し，再利用できるデータベースが必要である〔7，9，25〕。また，研究者が研究上の仮説に基づき，多様な視点から資料を眺めて，データを付与しながら仮説を検証できる必要がある。研究過程で生成されるデータはアドホックなものが多く，ワークベンチではこれらの形式や書式の定まらないデータを管理する必要がある。既存のデータベースシステムはスキーマ（属性構造，関係構造）が堅く，これらのデータの格納や上記の操作に対応しにくい。このため，ワークベンチは，既存のシステムの枠組みを拡張した柔軟な構造を持つシステムを用いる必要がある〔7，11〕。類似なシステムとして発想支援システムがある〔5，25〕。デジタルアーカイブでは，大量データの効率的な蓄積と抽出を目的とすることが多く，データの再利用は考慮されていないことが多い。

・資料・関連情報の公開システム

従来型の学術文献の出版のみならず，可搬性に優れる CD-ROM を用いてデータを配布したり，また，インターネット上の Web による研究成果を公開する方法が代表的である。資料データ，資料に関する論文，解説などの情報が公開の対象である。

我々のプロジェクトでは，上記のうち，特にワークベンチの実現に焦点を当てて柔軟な構造を持つモデルを提案し，プロトタイプシステムを構築している。以下では，資料をデータ化したものを資料データという。資料データはワークベンチ上での操作の対象となる。

## 2．2　ワークベンチの要件

資料データのワークベンチの満たすべきシステム要件は次の通りである。

R１　資料データの安全性

入力機器によりデジタル化されて格納されたデータは，多様な作業目的に利用されるため，資料データが改竄されるのを防ぐ機構が必要である。つまり資料画像データや資料のテキストデータは厳重に保護されなければならない。

R２　システムの対話性

資料に対する研究者の思考のための仮説や仮説から導かれる結果などをデータとしてシステム上に表現できる必要がある。つまり研究者がシステムの利用時に生成したアドホックなデータを追加できることが重要である。

R３　システムの柔軟性

R２のアドホックなデータを格納し，再利用できる必要がある。データの部分などを自由に取り扱ったり，複数の資料を並置したり，合成しながら作業過程のデータを保存することができる必要がある。これらはシステムの柔軟性を要求する。

R４　協調作業の支援

複数の研究者が共同で資料を整理したり，思考を進めたりするため，協調的に共同作業できる枠組みが必要である。ネットワークを利用して研究成果を研究者間で相互参照し，情報を共有することも含む。

## ２．３　資料データ

本稿では，資料として漢代の一等資料である木簡を対象としている。既に電子化された木簡は中国敦煌で出土した敦煌漢簡およそ1,000件と，居延で出土した居延漢簡およそ20,000件があり，プロトタイプシステムでは前者を用いている。

木簡は，古代のドキュメントである。個々の木簡のドキュメント型は未知であるが，複数の木簡を紐で結んで冊書として用いられたために，各木簡がドキュメントの型や冊書の意図を明らかにする上で重要なキーとなると考えられる。ワークベンチでは木簡のこの特徴をもとに研究者が作業を進めることを前提としている。

資料データは，Maspero 本 (Oxford) などの資料集の中の木簡の画像をスキャナーでデジタル化したデータ，ならびに大庭脩博士を中心とする研究グループの釈読をテキストデータ化したものである。釈読の種類は以下のようである。

・簡番

木簡に付与された番号。敦煌漢簡の場合，複数の研究者により同一簡に対して異なる番号が付けられている。ここでは，利用者による検索を容易にするために，次の４つの番号属性「疎勒河番号」，「Chavannes, Maspero, Xinai 番号」，「流沙墜簡番号」，「出土地番号」を用いている。

・釈読データ

木簡に書かれた文字を研究者が判読し，テキスト化したものが釈文である。釈文の各部分

文字列に対してキーワードとして解釈を与え，次の属性を付与したものを釈文と併せてここでは釈読データと呼んでいる。基本的な属性は予め確定しており，「釈文」，「人名」，「地名」，「年号」，「官名」，「干支」，「成語熟語」から構成される。また，同時に各木簡の形状，欠損状況，また他木簡との接続関係なども釈読データに含めている。

**例1**　図1の簡では，簡番がそれぞれ疎勒河番号：672，CMX　番号：C21，流沙墜簡番号：C271，出土地番号：T22-c-22 である。

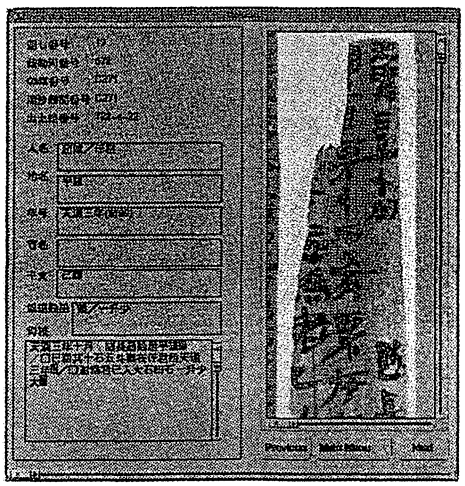

図1　資料データ

釈読データに関しては，釈文が「天漢三年十月隊長趙除居平望／□巳酉其十石五斗粟在任君所天漢三年／□遂為君巳入大石四石一升少大」である。「／」は各文字間が空いていることを示す。以下，前述のそれぞれの属性は順に人名が「趙除／任君」，地名が「平望」，年号が「天漢三年」，官名がなし，干支が「巳酉」，成語熟語が「粟／一升少」となっている。図中の簡番の中で通し番号とあるのはシステム上でデータの管理上に必要とする番号で簡番とは無関係である。

釈読データは資料の2次情報であるが，画像や釈読の中に新しい解釈を見い出したり，それらをもとに複数の木簡の並べ替えやコンピュータ上で合成を行う場合は，それらの2次情報が更に次の操作の対象となっている。つまり資料データは資料の復元作業（＝段階的な構造化）の基礎となるデータである。

## 3．資料データの部分の利用

本節では，利用者が資料データの部分を抽出し，それらを段階的に分類する機能を実現した

プロトタイプシステムについて述べる（図２）。このシステムは既存の OODB を拡張したデータモデルの上に構築されている［7，8］。ここではこのシステムを単にオブジェクトシステムと呼ぶ。

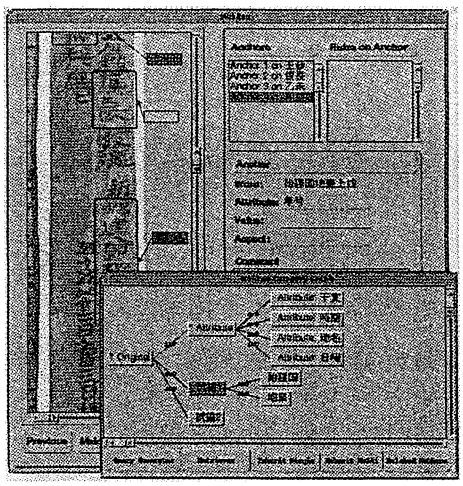

図２　木簡研究支援オブジェクトシステム

３．１　部分のデータ化

復元作業の初期の段階において，研究者は，まずシステムに格納されたデータの中から興味のある木簡の集合を収集し，１本ずつの木簡を精査する。この時，画像データや釈読データの意味のある部分を発見し，着目することが多い。例えば，木簡に含まれる文字の一部の文字や木簡画像の部分画像に着目する。これをオブジェクトの同定という。

研究者にとって抽出した資料データの部分に意見や意図を情報として付与する機能は重要である。更にそれらを分類したり関係付けたり，再分類して利用する。これらの操作は複数の木簡を集めて冊書復元を行ったり，木簡を仮想的に繋ぐことにより見えない文字を類推したりする場合に有効である。

これら一連の研究者の作業を計算機の観点から述べると，１）オブジェクトの同定，２）オブジェクトに対する実行時の属性／属性値の付与，３）オブジェクトやその部分の集約，分類，関係付け，などの機能に相当する。このような機能がに格納された資料データを用いて利用者の仮定に基づきながらシステムを操作する際に利用できねばならない。通常のデジタルアーカイブやデータベースシステムでは，オブジェクトの同定に対応する部分データをコンピュータ上で取り扱うことができない。そこで，まず，部分データを利用できるようアンカーの概念を用いる。

3．2　アンカー

アンカーはデータを指示する機構を持つオブジェクトで，もとのデータを複製せずに多重に利用することができる特徴を持つ〔6〕。図3にアンカーの概念を示す。$O_1$ が資料データ，$O_2$ がアンカーである。$O_1$ の2次情報である sentence の着目する部分を $O_2$ が指示している。

また，アンカーを資料データと同様に独立なオブジェクトとして取り扱うことで，資料データを整理するのと同様に，アンカーをも整理できる〔7，9〕。アンカーの属性構造は，属性名と属性値の組の集合で表しているため，利用者は各アンカーに自在に属性：属性値を付与することができる。

図3で，利用者は $O_2$ が指示する部分の属性を type: person's name, age: 23 のように与えることができる。このようにして資料データの部分を自在に抽出し，コメントなどを書き加えることができる。$O_2$ はオブジェクトとしてシステムの中に自立して生成でき，資料自身 $O_1$ と同等に取り扱える。

**例2**　図2後窓では，この中から特定の木簡を取り出して表示している。この木簡にはアンカー1から4が付けられており，後左窓において各アンカーが画像の部分を指定していることがわかる。後右窓には，アンカー4の情報が表示されており，利用者が付与した読みが「始建国地皇上戊」で，その属性が「年号」であることを示している。

図2の前窓では，木簡33番には「干支」として「戊午」が書かれた緑のアンカーが付けられ

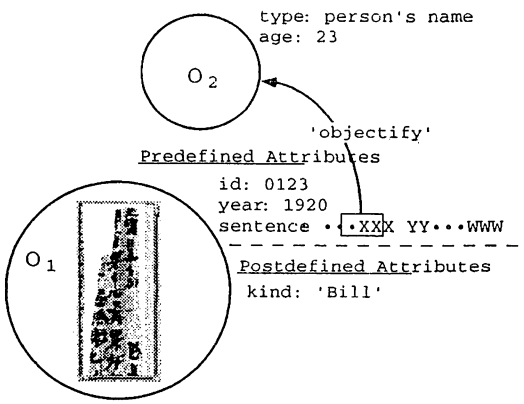


図3　資料データへのアンカー

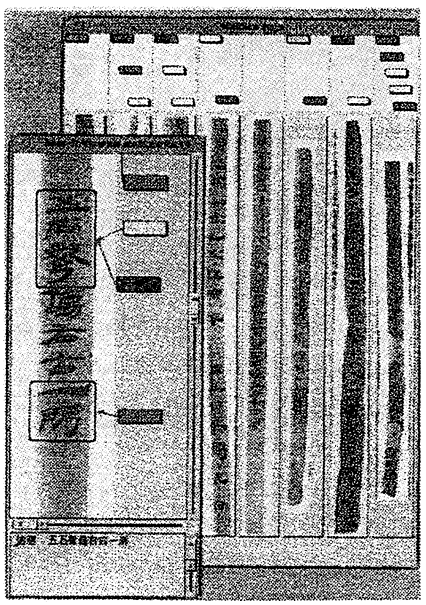

図4　アンカー一覧

ていることを示す。同簡の釈文は，「十七日　戊午　丁亥　丁巳　丙戌　丙辰　乙酉　乙卯　甲申　甲寅　甲申　癸丑　癸未」と与えられている。

**例3**　図4は，各木簡画像に対して利用者が画像の部分に着目して付与したアンカーを一覧している。各木簡でアンカーの個数は異なっており，色で区別されている。

3．3　アンカーの分類と監視機構

利用者は自分で生成したアンカーを自由に分類することができる。この分類項目のことをシステム上でカテゴリと呼んでいる。利用者はカテゴリを自由に新規生成でき，アンカーをその属性に応じて試行錯誤的に分類できる。

**例4**　図2では，前窓にはアンカーを付与された読み「始建国」，「地皇」がカテゴリー「試論1」に分類してみている様子を示している。上記以外にも属性として「干支」，「時期」，「地名」，「日時」などに対するカテゴリが，大分類項目を与えるカテゴリ「Attribute」に試行的に分類されている。

カテゴリに対する一連の操作は，システムによって監視し，制御することができる。操作には，カテゴリの生成や削除，カテゴリへの分類操作などがある。このような利用者の操作に対する監視・制御側はルールにより記述されている（図5）。これらのルールによりカテゴリを削除する際に利用者に注意を促すメッセージを発したり，誤って必要以上に多数のアンカーを1つのカテゴリに分類したりするなどの利用者の誤動作を回避することができる。

この監視機構は ECA (Event Condition Action) 機構により実装されている。ECA 機構

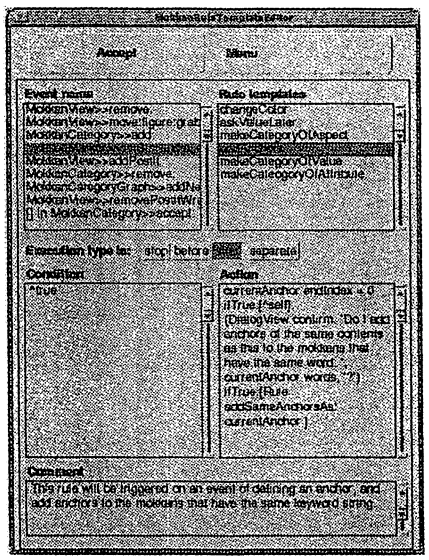


図5　アンカーの分類作業を監視するルール記述

用いるルールはカテゴリー毎に定義することができるため，利用者の分類作業をきめ細かに制御することができる。

3．4　システムの実装

本システムは，次世代データベースモデルとして開発されたインスタンスベースシステムのObase モデルの上で実現されている〔8〕。Obase モデルは集合オブジェクトが基本オブジェクトとして実装されている。システムの核には OODB を用いており，オブジェクトシステムは DBMS の外側に実装されている〔9，10，7〕。OODB では，技術的には，（1）主記憶上のオブジェクトをそのまま2次記憶に格納し永続化できる，（2）オブジェクト間の参照関係をそのまま保存できるため，複雑な構造を持つオブジェクトもそのまま格納できるという特徴を持つ。これにより利用者の用途に応じて木簡画像や釈文の部分を抽出したり，合成するといった操作，また検索結果を加工する操作が可能となる。また，グラフィックユーザーインタフェースの作成には Visual Works を用いている。

アンカーは格納層の上位レイヤーの機構として定義されており，マルチメディアデータを利用する上で基本機能である〔6〕。アンカーは部分領域の指定のみならず，任意に与えられたデータをオブジェクト化するためにも用いられる。

## 4．視点の利用

研究者が資料データを扱う場合，曖昧な動機，直感，仮定などに基づく試行錯誤的な取り扱いをすることが多い。また，同じデータを多様な角度からデータを扱い，複数の目的に対して取り扱うことが多い。本節では，通常はデータ化されない利用者の視点をデータ化し，それをキーとして木簡データに対して多様な視点から表現を与えながら段階的に集約するための枠組みを考える。また，この枠組みを実現したプロトタイプシステムについて述べる。このシステムでは，グラフを用いて研究者の思考や資料データの見方を表現しているので，システムをグラフデータベースと呼んでいる〔11，12〕。

4．1　利用者の視点とグラフ

研究者にとって研究上の仮説や仮定に応じて資料データを試行錯誤的に取り扱うことが，新しい着想を得る重要な行動である。研究者はこのようなデータに対する発散的な思考を行いながら，次第にデータを1つの視点からまとめる集約的な思考を繰り返すものと考えられる。そこでシステム上で利用者の視点をデータとして扱い，資料データを利用者の視点に応じて属性を与え，資料データを集約する作業を考える。

通常のデジタルアーカイブなどでは，システムに格納されたデータの表現は一意的で，利用

者が次々に思い付く自分の視点に応じて資料データをシステム上で仮想的に生成して表現することはできない。

**例5**　資料データの集合と利用者の視点を1つのグラフ上で表す〔12〕。図6に示すように，1つの資料データを複数の視点から表現する。つまり左側の人は，木簡を「手紙」とみなし，右側の人は「請求書」とみなしていることを示している。この場合，「手紙」と「請求書」が2つの異なる視点である。

上記のようにノードとエッジからなるグラフを考える。グラフ上で各資料データをノードで表し，資料データに対する利用者の視点を別のノードで表し，両者を有向エッジで連結する。有向エッジが該当する視点から資料データを見ることを表すものとする。エッジにはその視点から見た資料データの特徴を属性として与えるものとする。同じ資料データに対して異なる視点を設ける際は，新たな視点のノードを生成し，エッジで連結し，特徴を属性として与える。この作業を繰り返すと資料データの集合と利用者の視点の集合がエッジにラベルを持つグラフ構造をなす。属性は利用者が自由に記述できるものが望ましいので，属性名：属性値とも利用者定義とする。資料データに視点を次々に設け，視点から見た資料データの特徴を属性として追加することが，グラフを段階的に生成・更新することになる。

利用者の視点と集合化作業の過程をシステム上で表すため，形式的に一般化したグラフを階層構造グラフとして次のように定義する。

**定義1**　階層構造グラフとは次の条件を満たすノードとエッジからなる非巡回的な有向グラフである（図7）。葉ノードは資料データを表し，その他のノードは利用者の視点を表す。ノードとエッジはそれぞれの性質を表すラベルを持ち，ノードのラベルはそれ自身の性質を表し，エッジのラベルは2つのノード間の関係を表す。有向エッジは2つのノードの上下関係，即ち視点の上下関係と視点と資料データの関係を表し，方向は下位に位置づけられたノードから上

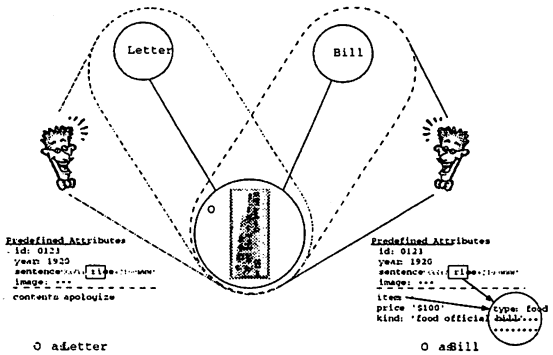


図6　木簡に対する2つの視点

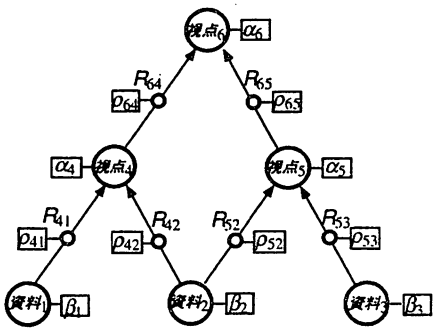


図7　階層構造グラフ

位ノードへ向かうものとする。□

視点と資料データの連結は，その視点からの資料データの集合化をも表す。ここで視点は，(i)利用者の興味に応じたデータの範囲を規定し，(ii)データへの自在な属性の付与と集合化作業における操作単位となるもの，と仮定している。

階層構造グラフでは，視点は多段に階層化されている。本来，視点の上位／下位の関係と視点の集約関係は異なるが，モデルの簡単化を図るため，エッジで両者を表すものと仮定している。これは資料データを半構造化データ〔21，22〕と見なし，半構造化データをグラフを用いて段階的な構造化作業と捉えることができる〔12，11〕。

## 4．2　グラフの利用法

階層構造グラフの構成方法から考えて，1つの視点に基づくデータの表現は両者を結ぶ部分グラフにより表されることになる。この部分グラフは利用者の視点に基づく資料データに関するすべての属性を含み，資料データの見せ方を仮想的に生成するあらゆる情報を含んでいる。このためこの部分グラフを視点に基づく仮想オブジェクトと呼んでいる。

定義2　グラフオブジェクト $g$ 上のオブジェクト $O$ が上位のカテゴリ $C$ に連結されている時，$C$ から $O$ に到るすべての経路からなる $g$ の部分有向グラフ $g'$ を視点 $C$ に依存した $O$ の仮想オブジェクトといい，$\tilde{O}[C]$ と表記する。この場合，$O$ はカテゴリでもよい。

この定義でわかるように仮想オブジェクトは視点に基づく資料の情報の集合である。利用者の視点に依存した資料データは，階層構造グラフ上の両者の間の経路に与えられた属性を集めることで得ることができる。集めた属性を，仮想オブジェクトの属性として定義することは自然である。仮想オブジェクトの属性は資料データの見え方を定義する。1つの資料データに対して視点を与えると，両者を結ぶパスの集合がその視点に基づく資料データの見方を与えることになる。また，視点に依存した資料データの属性が定まる。同じ資料データに対して，別の視点を与えると部分グラフの形状が変わり，異なる視点に依存した資料データの属性が定まる。

更に，階層構造グラフ上では，同一視点から統一的に仮想オブジェクトを取り扱うため，スコープの概念を導入している。これにより，一つの視点に基づく複数の資料データの見方を階層構造グラフ上で定義することができる〔14，13〕。スコープは視点階層における視点の切り替え操作を行う。

階層構造グラフを用いることで，(1)視点に基づく属性を実行時に生成，(2)視点の切り替えによる異なる資料データの属性構造（＝オブジェクトビューという）の実現，(3)利用者の多様な

視点と視点に依存する属性構造の収納，などが行える。

システムの中でグラフを単位として扱うことで，資料データに対して利用者の視点に依存した構造化作業の過程を表すことが可能である。特に，利用者毎のデータベース作成，内容に応じた多重分類，複数の視点からの意味付けなどが行える。

## 4．3　システムの実装

本システムでは，Web データや科学技術データなどに代表される半構造化データを段階的に構造化するための枠組みとしてグラフモデルを実現し，代表的な構造化作業としてデータの集合化と属性の付与の２つの構造化作業について実装している。このモデルは利用者の視点を陽に扱い，視点をデータとして取り扱う特徴があり，利用者の複数の視点から多様な構造を持つデータを柔軟に構造化することができる。

**例6**　図８で後中央の窓に階層構造グラフが見える。今，最も大きな視点Wが選ばれており，利用者はこの窓中に次々に自在に視点を生成できる。視点Wに基づく木簡の属性が前窓に表示されている。左側が属性で，右側が属性値である。利用者は視点に基づく属性ならびに属性値を定義し，見ることができる。

本グラフベースシステムにより，資料データを安全に保存しながら，視点を自在に生成し，資料データの見せ方を切り替えることができる。

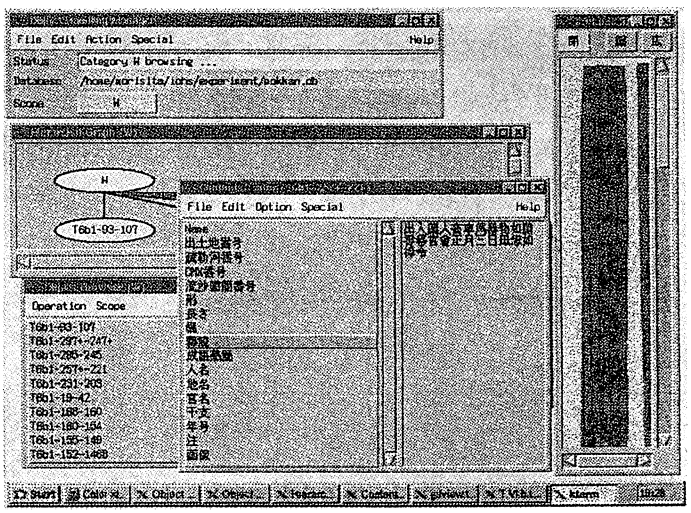


図８　視点を利用したグラフベースシステム

### 4．4　共同作業への発展

前節で述べたスコープを操作することで複数の利用者が１つの階層構造グラフを共有することができる。これにより他の研究者の作業の結果を利用したり，共同作業を行うことができる。

また，階層構造グラフを構造情報とコンテンツ上のの２つの分離し，両者でグラフを表すことにより，資料データの集合に複数のグラフ構造を生成することができる。言い換えれば，同一資料データを用いて，データを保護しながら複数の利用者が異なる目的で構造化することができる。

## 5．研究成果の公開

### 5．1　インターネットの利用

研究作業で得られたデータは，研究成果として公開し，共有することが望ましい。また一般に公開しなくとも研究者グループの中で共有する必要がある場合が多い。例えば同一課題を研究分担する場合や，複数の研究者が同一の基礎データを用いてなどがこれにあたる。

電子データを公開する場合，CD-ROM で配布する方法とインターネット上の Web を用いる方法が代表的である。CD-ROM の場合，データの携帯性がある反面，多数の複製が存在するため，データに変更があった場合や新しいデータが追加された場合などのデータ更新は容易でない。ここでは情報発信の手段の一つとして Web を用いてデータを公開する方法を実現している。この方法はインターネット上で情報を公開する最も簡便な方法として用いられている。

ここでは，敦煌漢簡に関する１次データの木簡画像と２次データの釈読データの両方を公開している。画像データは入力に用いた原典の著作権が一定の年限を経過しているため自由に公開できる。また，釈読データは作成者の了解を得て公開している。

### 5．2　ファイルの検索

本システムでは，前述の２つのプロトタイプシステムで用いたデータを１つのファイルに書き出して公開用のデータとしている。このファイルを検索プログラムが検索し，検索された文字列からの HTML 文書を動的にプログラムで生成している。実装には Perl を用い，サーバーのゲートウエイ機能を用いている。

利用者は Web ブラウザ上で任意の属性について検索文字列を入力し，適合する木簡を取り出すことができる。また照合パターンとしては完全一致と部分一致の両方により行える。

**例7**　図9は Web のブラウザーを用いて疎勒河番号が813の木簡を検索し，簡番，釈読データを表示した例である。また，同簡の画像も右窓に表示されている。図中で釈読データで該当する属性を持たない場合は，値の部分に黒線が引かれている（官名，干支)。

**例8**　図10は，敦煌漢簡971本に対する釈読データに「食」の文字を含む木簡の疎勒河番号，

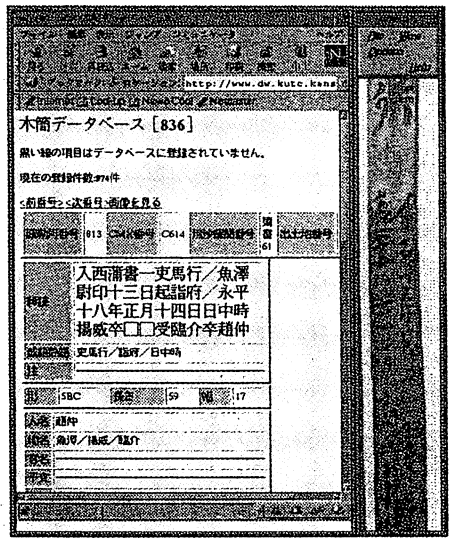


図9　Web による情報公開

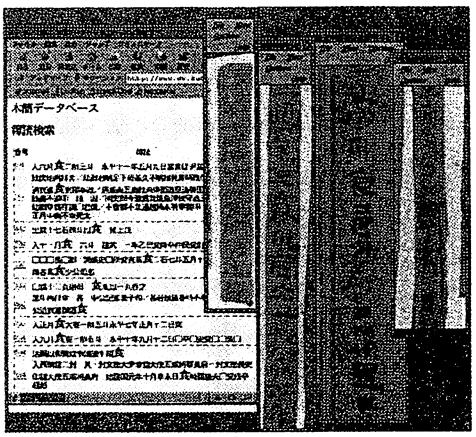


図10　釈文に「食」を含む木簡の検索

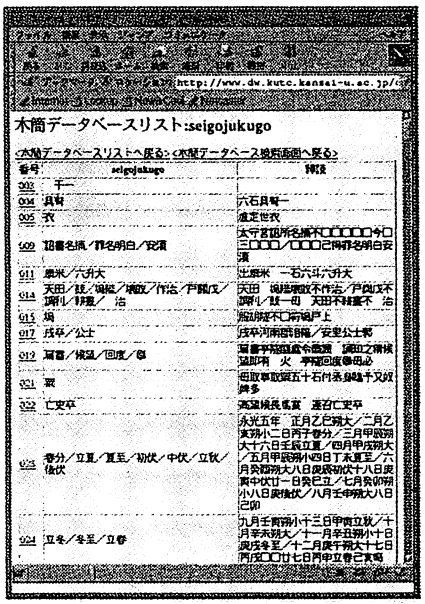


図11　敦煌漢簡における成語・熟語の一覧

釈文，画像を表示している。図中では4書類の「食」の肉筆が見えている。また，成語・熟語の検索結果の一覧が図11に示されている。

Web を用いたデータの公開手法はインターネット上で容易にデータにアクセスできるため，データに頒布性が高い。また情報を一カ所に配置しているため，データの更新が容易である。また，本手法のようにファイルシステムへのアクセスのみならず，データベースのフロントエンドとして用いることができる特徴を持つ。

しかし，Web が用いている http プロトコルは，文字をストリームとして転送する機能のみを備えており，利用者側から文字データ以外の画像データや線画データを検索キーとして転送することができない。そのため本枠組みではスケッチ画や文字の特徴による木簡画像データの検索は行えない。

## 6．3つのシステムの比較

表1に，3，4，5節で述べたシステムを比較する。ここでは2つの切り口から比較する。まず，資料復元研究支援システムとして，特に木簡データの整理を目的とした場合のシステムの有効性の観点から上半分に，次に技術的側面から下半分に示している。

オブジェクトシステム，グラフベースシステムはいずれも資料復元研究におけるワークベンチの形態を持つ。従って作業過程における必要なデータや思い付き，仮定，簡単なコメントなどをデータとして格納できるようになっている。後者は資料データの構造化を行うワークベン

表1　3つのプロトタイプシステムの比較

| | オブジェクトベース<br>システム | グラフベース<br>システム | ファイルの Web<br>インタフェース |
|---|---|---|---|
| 特徴 | 部分データの利用<br>逐次分類 | 視点のデータ化，視点に依存<br>したデータの扱い | インターネット上<br>での情報公開 |
| 形態 | ワークベンチ | ワークベンチ | サーバー |
| 資料復元作業<br>における役割 | 部分データの利用，<br>欠落簡・冊書の組合せや合成 | 仮定に基づく試行錯誤的な<br>構造化（属性付与と集合化） | 情報公開<br>索引付け |
| 技術的新規性 | 新規 | 新規 | 既存 |
| モデル<br>主要素技術<br><br>技術的焦点 | Obase モデル<br>アンカー操作<br>ルール機構<br>オブジェクト化<br>アンカーの逐次分類 | 階層構造グラフモデル<br>グラフ操作<br>2層構造<br>オブジェクトの多重表現<br>半構造化データ | ファイル<br>ゲートウエイ機構<br><br>文字検索<br>Web 文書の自動生成 |
| 核システム<br>実装言語<br>稼働システム | OODB<br>smalltalk<br>WS | インスタンスベースシステム<br>Lisp + ELK<br>WS / PC | ファイルシステム<br>perl<br>PC |
| DB 種別 | 柔構造 DB<br>科学 DB | 柔構造 DB<br>科学 DB | 固定構造 DB |
| 文献（人文科学）<br>文献（工学） | [18, 19, 16]<br>[7, 16, 17] | [20]<br>[11, 12, 13, 14, 15] | |

チであり，構造化の過程自身がグラフとしてデータベースに格納されている。両システムでは，属性と属性値とも利用者が定義できる形式を持つため，複雑なデータ構造を持つデータをこの形式で表現することが可能である。また，データ構造として構造情報とコンテンツ情報から構成される2層構造としているため，複数の利用者が互いに異なる目的のためにシステムを利用したり，また共通の目的で作業結果を互いに利用することができ，拡張性がある。

前者は smalltalk，後者は scheme を用いて実装されているため処理速度が遅い。これらはプロトタイプシステムの開発に適しているため用いており，他の高速処理の可能な言語で代替することで高速化できる可能性が高い。いずれも簡単な処理系でデータの格納系を実装できることがモデルの有効性を示している。

また，計算機でグラフを取り扱う場合は，処理が複雑になり必ずしも有限時間で終了しない問題となることが多い。しかしグラフベースシステムで利用されるグラフは，利用者の視点間の関係を表現しているため，深さが高々5段，幅が資料データの数の大きさを考えれば十分と考えられる。この場合，グラフ処理に有効なアルゴリズムもで提案している〔12〕。

## 7．おわりに

本稿では，オブジェクト指向技術を用いた復元研究を支援する3つのシステムについて述べた。システムは，研究を迅速に進めるためのシステムの効率化，また研究成果を効果的に相互利用する方法として複数の利用者によるワークベンチの利用などが今後の問題である。

最後に，オブジェクトベースシステムの開発に多大なご指導とご尽力を頂いた Obase consortium（代表：神戸大学田中克己教授）のメンバー各位にここに記して衷心よりの謝意を表す。神戸大学大月一弘助教授，同大学森下淳也助教授，姫路独協大学杉山武司助教授にはグラフモデルの基礎的枠組みに関して本質的な示唆を頂いた。また，関西大学総合情報学部学生中山怜志君には Web システムの開発にご協力頂いた。ここに記して衷心よりの謝意を表す。

**参考文献**

〔1〕 杉田，“民族学研究のための画像データベース”，システム制御情報，Vol. 33 No. 6, pp. 273-281 (1989)

〔2〕 金谷一朗，陳謙，千原國宏，“VR技術を応用した遺物復元システム”，情報考古学誌，Vol. 3 No. 1, pp. 35-46 (1997)

〔3〕 横矢直和，増田健，“多視点距離データを用いた3次元形状モデリング”，公開シンポジウム人文科学とデータベース「データ」を読む・観る・解く」，pp. 55-60 (1995)

〔4〕 台湾中央研究院，http://www. sinica. edu. tu/ftms-bin/ftmsw 3

〔5〕 國藤進，“発想支援システムの研究開発動向とその課題”，人工知能学会誌 Vol. 8 No. 5, pp. 552-

559 (1993)

〔6〕 Halasz, F, Schwartz, "The Dexter Hypertext Reference Model", Communications of the ACM, Vol. 37 No. 2, pp. 30-39 (1994)

〔7〕 Ueshima, S., Ohtsuki, K., Morishita, J., Qian, Q., Oiso, H. and Tanaka, K., "Incremental Data Organization for Ancient Document Databases,", Proc. of the 4th International Conference on Database Systems for Advanced Applications (DASFAA '95), pp. 457-466, Singapore (1995).

〔8〕 Tanaka, K., Nishio, S., Yoshikawa, M., Shimojo, S., Morishita, J. and Jozen, T., "Obase Object Database Model: Towards a More Flexible Object-Oriented Database System," Proc. of the International Symposium on Next Generation Database Systems and Their Applications (NDA '93), pp. 159-166, Sept. 1993.

〔9〕 上島紳一, 大月一弘, 森下淳也, 田中克己, "歴史的資料を対象としたサイエンティフィックデータベースのシステム設計", 電子情報通信学会研究会技術研究報告 DE93-47 (1993)

〔10〕 Shinichi Ueshima, Kazuhiro Ohtsuki, Jyunya Morishita, Qing Qian, Hiroaki Oiso, Katsumi Tanaka, "Design and Implementation of an Object-Oriented Scientific Database System for Historical Materials", Proceedings of Obase symposium III, pp. 204-215 (Obase Consortium) (1994)

〔11〕 Ueshima, S., Morishita, J., Ohtsuki, K., Sugiyama, T., "Hierarchical Graph Model: Collecting Objects and Generating Multiple Views for Semi-structured Data", in "Advanced Databases", World Scientific ltd. (1998).

〔12〕 上島紳一, 森下淳也, 大月一弘, 杉山武司, "階層構造グラフを用いた半構造化データの段階的な構造化手法", 情報処理学会論文誌 Vol. 39, No. 4 pp. 857-867 (1998)

〔13〕 森下淳也, 上島紳一, 大月一弘, 杉山武司, "仮想オブジェクトを用いた半構造化データの段階的な構造化手法の提案", 平成8年度文部省科学研究費重点領域研究「高度データベース」松江ワークショップ講演論文集, pp. 144-146 (1996)

〔14〕 上島紳一, 森下淳也, 大月一弘, 杉山武司, "階層構造グラフを用いた半構造化データの段階的な構造化手法の提案", 情報処理学会技術報告 DBS-111, pp. 9-16 (1997)

〔15〕 森下淳也, 上島紳一, 大月一弘, 杉山武司, "階層構造グラフにおける属性の取り扱い方に関する検討", 電子情報通信学会データ工学研究会 DE96-79, pp. 31-36 (1997)

〔16〕 上島紳一, "資料復元研究支援データベースシステムの開発(1)—開発の動機とシステムモデル", 関西大学東西学術研究所紀要26輯, pp. 1-12 (1993)

〔17〕 上島紳一, "資料復元研究支援データベースシステムの開発(2)—木簡画像データの扱い", 関西大学東西学術研究所紀要27輯, pp. 47-60 (1994)

〔18〕 森下淳也, 上島紳一, 大月一弘, "視点に依存した属性付け機構を持つ木簡研究支援データベースシステム—構造進化型データベースの概念—", 公開シンポジウム人文科学とデータベース「データ」を読む・観る・解く」, pp. 19-28 (1995)

〔19〕 上島紳一, 森下淳也, 大月一弘, "木簡研究と情報処理", 日本情報考古学会情報考古学会第1回大会発表要旨集, pp. 74-79 (1996)

〔20〕 森下淳也, 大月一弘, 杉山武司, 上島紳一, 大庭脩, "木簡研究支援データベースシステム—知見と仮説に基づく再構造化", 公開シンポジウム人文科学とデータベース「データ」を読む・観る・解く」, pp. 105-112 (1997)

〔21〕 Buneman, P., "Semi-structured data." See the URL, "http://www.cis.upenn.edu/db/tutorials/semistructured-paper.ps."

〔22〕 Buneman, P., Davidson, S., Hillebrand, G., Suciu, D., "A Query Language and Optimization Techniques for Unstructured Data," Proc. of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Canada, pp. 505-516, June 1996.

〔23〕 Zdonik, S., "Incremental Database Systems: Databases from the Ground Up," Proc. of the 1993 ACM SIGMOD International Conference on Management of Data, Washington DC, USA, pp. 408-412, May 1993.

〔24〕 原田正則，宝珍輝尚，中田充，都司達夫，"データ型に基づくマルチメディアデータ参照機構の実現とその有効性"，情報処理学会論文誌，Vol. 38, No. 8, pp. 1603-1612 (1997)

〔25〕 中田充，宝珍輝尚，都司達夫，"サイエンティフィックデータベースのための一データモデル"，電子情報通信学会第6回データ工学ワークショップ (DEWS '95), pp. 71-78 (1995)