

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——

上 島 紳 一

1. はじめに

考古学、史学、民族学や民俗学などにおいては、フィールド調査などにおいて収集した資料や遺物を整理・分類し、効率的に利用する目的で、多くのデータベースが構築されてきている。通常、これらの資料や遺物は、断片化された状態で発見されることが多い。資料等の断片の内容をもとに歴史的事実を明らかにすることは、考古学や史学等の研究者の中心的研究課題であり、特に断片を比較・合成することにより、その断片間の関係を発見したり、原形を再現すること（資料の復元作業という）の意義は大きい。

我々は、敦煌において出土した漢代の木簡に対する資料データベースシステムを試作した〔1〕。このシステムは、木簡研究の専門家が、破断した木簡の木片を整理するためにつけた属性と、その値および木片の画像から構成されている。この資料データベースシステムの有効性を検討した結果、専門家がこのシステムを木片の復元作業に利用する上で次の問題点があることが明らかになった。

1. 資料の情報処理を行う上で、計算機処理を行う前に専門家自身が行わなければならない前処理が多い。
2. 通常のデータベースの持つ単純な検索機能と、専門家の意図する検索方法とが乖離している。
3. 資料が断片化しているため、専門家の欲しいキーワードが資料から欠落している場合が多い。そこで専門家は、不足情報を補うために、なかば当推量的な仮定をたてて情報処理を行う。この仮定されたキーワードをデータベースに一時的に格納する事が困難なため、計算機を用いた有効な情報処理が行えない。
4. 新事実発見後のデータの変更・再編集（画像の合成を含む）が容易でない。
5. 木簡研究者にとって、木片間の関係を判断する決定的な方法は、木片の画像を比較することにある。比較する際の判断基準は、研究者に蓄積された経験に基づくこと

が多く、データベース化しにくい。

これらの問題を解決するためには、計算機側が、より高度な作業過程を負担するシステムを構築する必要がある。本稿で述べる資料復元研究支援データベースシステム（以下、本支援システムという）は、次の点を意図している。

1. 利用者の知識をデータベース化し、検索、更新などの処理機能を向上する。
2. 仮定に基づく一時的情報をデータベース化する。
3. 画像表示機能を充実することにより、専門家の目による判断を助ける。
4. データや画像の変更・再編集作業を簡易化する。

本支援システムの機構は、木簡研究のみならず、断片などから原形を復元する研究に共通するものであり、人文系、自然系を問わず、多くの分野に適用できる。2節では、資料復元研究支援データベースシステムの開発の方向と専門家の作業モデルについて述べる。3節では、システムの基本的動作形態をモデル化する。

2. システム開発

2.1 背景

人文科学研究の様々な分野において、収集した資料の情報処理を行う目的で研究支援データベースシステムが開発されている。博物館、美術館などの大量の収蔵物に対する大規模データベースシステム [2] [3] [4] [5] [6] [7] や、文化財科学、保存科学などで加工処理の対象とされるデータを対象としたシステム [8] などがよく知られている。これらのシステムは、格納されたデータの高速な検索機能や、文字データや画像データの効果的な表示機能 [9] などを持つ。更に高度なシステムとして、データの分類方法や格納方法に工夫して、類似画機能を持たせたシステム [10] [12]、対話的な利用者インタフェースを持たせたシステム [11] なども盛んに研究されている。他にも抽象度の高い人間の検索要求に基づく検索機能の実現の為に、認識処理、ルールを用いた検索システム [13] [14] [15] なども提案され、様々な分野への応用の可能性を持つ。

これらのデータベースシステムが、蓄積されたデータに対する効果的な問い合わせ処理機能の実現を主眼としているのに対し、本稿で述べる資料復元研究支援データベースシステムは、様々な局面で研究対象となる断片化した資料の復元作業を支援する目的で設計されている。次節に述べるように、資料の復元作業は複数の作業工程からなる為、システムは、各工程において蓄積されたデータに対して処理を行ったり、また新たなデータを生成したりできることが要求される。即ち、システムは、データの問い合わせ機能に加えて、データの加工処理機能、

専門家が一時的に仮定する情報の処理機能，作業の各工程で生成されたデータを適宜，データベースに格納し，次の工程で再利用できるような機能などを持つことが必要である。このようなデータの加工と生成を中心とした機能を持つシステムを構築するには，RDB (Relational Database Management System) は，スキーマ構造が柔軟性に欠け，作業の途中で生成されるデータの取り扱いが容易でないことが指摘されており適当でない [20] [21]。

このような状況を鑑みて，われわれは資料復元研究支援データベースシステムの基本的枠組みとして OODBMS (Object-Oriented DataBase Management System) [16] [18] [19] [20] [21] を用いている。OODBMS は OOPL (Object-Oriented Programming Language) を基礎に，オブジェクトの識別性と永続性，データベース機能を持つソフトウェアシステムであり，システムの目的の添った形で再利用性，拡張性の高いシステムを設計できる特徴がある。

システムを設計する上で重要な課題の一つとして，従来専門家が手作業で行っていたデータ処理作業の計算機への移植がある。特に，OODBMS を用いたシステムでは，システムモデルの構築とスキーマの定義が重要である。これまで，ER (Entity-Relationship) モデルをもとにシステムのスキーマを構築する手法や [33] [34]，システムの各開発工程を複数に分け，各段階においてオブジェクトの静的側面，機能的側面，状態遷移を定義することにより大規模ソフトウェアの開発を可能にする手法などが提案されているが [35]，一般には，事例の特徴に応じてシステム設計者が両者を定義することが多い [23] [27] [28] [29]。本支援システムの開発にあたっては概ね [35] の手法を用いている [16] [17]。

2.2 専門家の作業モデル

ここでは，システム開発の第1段階として断片資料の復元作業における専門家のデータ処理作業についてモデル化する。

専門家が断片につけるキーワードは，断片から読みとることのできる内容に限られているため，予め用意した属性の一部分の情報しか持たないものが多い (属性値の欠落)。また，書かれている属性値の内容も意味が曖昧なことが多い (属性値の曖昧性)。

専門家は，属性値間および属性間の関係に関する知識を持っており，その知識を利用して各断片について属性値を検索し，断片間の関係を評価する。データ処理の結果得られる情報は，各断片のもつ情報量が少ないため不確実なことが多い。そこで，次の2種類の評価を行う。即ち，比較する断片間の関係 (例えば，原形が同一であるなど) が正しい可能性および正しくない可能性に対する評価を求める。この評価をもとに関係を推測する。

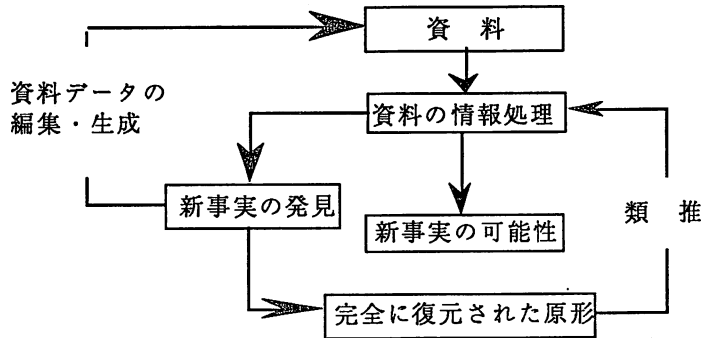


図1 専門家の作業モデル

研究者が断片資料間の関係を見つけ出すために行う作業は、図1のようにモデル化できる。まず、資料の中から鍵となる断片を決定する。次に、その断片に関係が有り得る断片を、資料に情報処理を加えることにより抽出する（データの絞り込み）。更に、別の角度からの情報処理を加える。この手順を研究者の知識をもとに試行錯誤しながら繰り返し、断片間の関係を発見する。その後、発見事実をもとの資料に書き加えたり、断片を合成するなどの再編集を行う。また、完成型データ（原形の復元）をもつくり出す。

3. データベースシステムモデル

3.1 システムの目的

本稿で述べるデータベースシステムは、考古学、史学等の研究者が行う、断片化された資料の原形を復元する作業を支援することを目的としている。図1に示した作業過程において、専門家が断片間の関係を見出す際の決定的な方法は、目による判断である。即ち、断片そのもの、または、断片の写真等を自身の目で確かめることにより、断片間の関係を判断する。この目による判断を行う前に、まず膨大な資料の中から比較の対象となりうる断片を見つけなければならない。この作業をするために、専門家は各断片にいくつかのキーワードをつけて資料を整理している。専門家の作業の中で最も困難を極めるのが、キーワードから比較対象を探し出す作業である。言い換えれば、比較の対象物を見つけることができれば、後は、比較的簡単に「専門家の目」によって判断を行うことができる。また、この「目」は、専門家によって長年培われたもので、その判断基準を計算機に実装することは容易ではない。そこで、本支援システムは、専門家の目による判断を補助するために、（1）比較対象となる資料を探し出し、（2）その比較対象となった断片の画像（=写真）の比較に必要な部分を表示する、ことを主眼に設計されている。

3.2 システムモデル

資料復元研究支援データベースシステムのモデル（以下では、システムモデルという）を図2に示す。図において□は、システム（＝計算機）側に準備される部分、○は、計算機が取り扱わずに利用者自身が作業を行う部分を示す。以下に、モデルの構成要素について述べる。

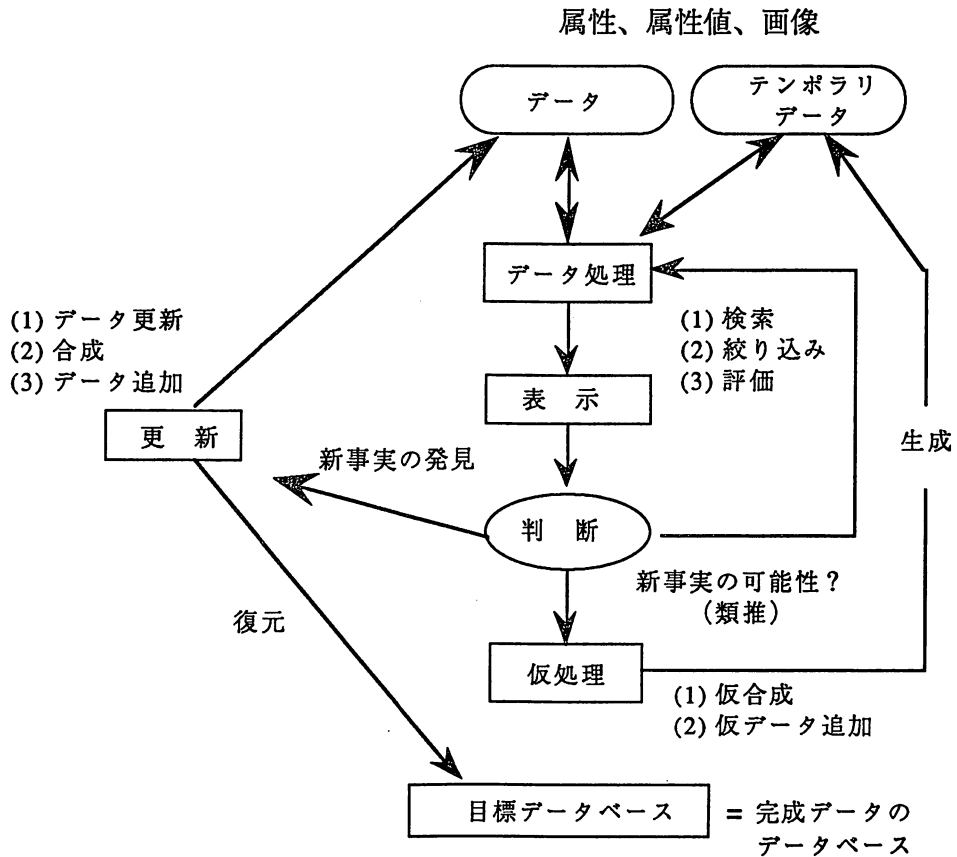


図2 資料復元研究支援システムのシステムモデル

(1) 資料

データベースに格納されるデータで、断片の写真などの画像データ、および専門家が各断片につけたキーワードである属性値データからなる。専門家は断片を分類・整理するために、複数のキーワードの要素（属性）を準備している。本システムで対象とする資料では、通常データとは異なり、各資料に各属性の情報が欠落している場合が多く、更に、属性値の内容が単純ではないという特徴がある。

資料は資料オブジェクトとしてデータベースに格納される。資料オブジェクトには、専門家がデータを操作する際の基本単位となる基本オブジェクトと、基本オブジェクトの一部分を抽出した部分オブジェクトがある(図3)。

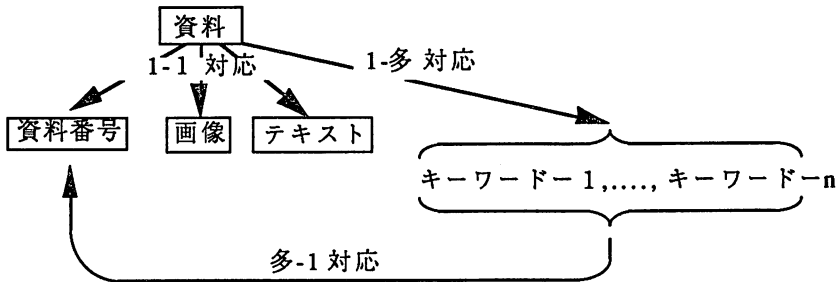


図3 資料オブジェクト

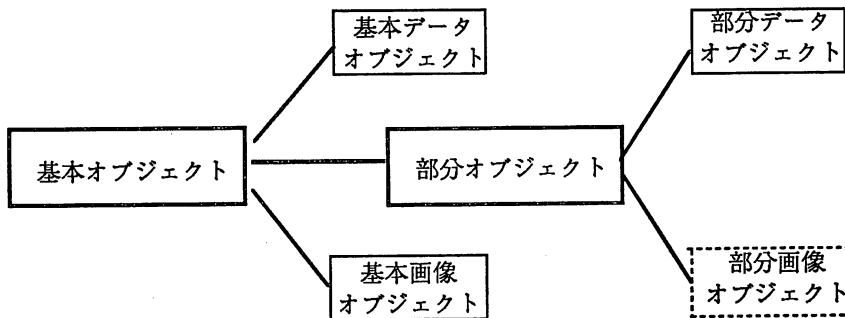


図4 基本オブジェクト部分オブジェクト

〔基本オブジェクト〕

基本オブジェクトは1つの資料の情報を持つオブジェクトである。基本オブジェクトは基本データオブジェクトと基本画像オブジェクトから構成される。基本データオブジェクトは、資料の番号、専門家が資料に書かれた文字などから抽出したキーワード(属性群)、および資料の作成者名、作成年、作成場所などからなる。基本画像オブジェクトは資料の画像である。

〔基本オブジェクトの特徴〕

資料は原形の一部であるため、単独では書かれている内容が確定しない。単に資料に書かれている文字列が取り出され、属性値としてオブジェクトに格納されることが多い。例えば人名という属性に格納されている属性値は資料の作者の名前とは限らず、資料に現われる人の名前である。この名前が、原形を復元する過程において文意により作者であることが判明した場合に、作者名に格納される。年号と作成年、地名と作成場所の属性についても同様である。

〔部分オブジェクト〕

基本画像に含まれている一部分（例えば文字やパターンなど）に注目して、表示や処理を行ないたい場合がある。基本画像の一部を取り出したものを部分オブジェクトと呼ぶ。部分オブジェクトは、部分画像オブジェクトと部分データオブジェクトからなる。部分画像オブジェクトは原画像の一部をデータとしたものである。部分データオブジェクトは取り出した画像に関するデータであり、文字の読みや部首、部分画像の特徴などの属性からなる。

部分オブジェクトには単語オブジェクト、一文字オブジェクト、部首オブジェクト、ハネオブジェクト、部分的な特徴パターンオブジェクトなどがある。これらの部分オブジェクトをすべて準備しておく数が膨大となるため、ユーザーが頻繁に使うことが予想されるオブジェクトのみをシステム定義オブジェクトとして予め準備しておく。本システムでは一文字オブジェクトがこれにあたる。他のものはクラスのみを準備しておき、ユーザーが必要に応じてオブジェクトを生成する。

〔部分オブジェクトの特徴〕

部分画像オブジェクトは画像の実体を持たず、原画像における画像の占める領域の情報のみを持つ。部分オブジェクトは、表示や処理などで必要な場合に基本オブジェクトの画像から必要な部分を取り出され、部分画像オブジェクトとして実体化する。

(2) 目標データベース

復元作業により再現された、原形に関するデータを格納するデータベースシステムで、完成データのデータベースとなる。

(3) データ処理

計算機によって、資料間の関係や専門家の判断や推測の正当性を評価する部分である。資料の復元作業において、専門家にとって最も重要な課題は、ある2つの資料が同一原形を構成していたかどうかを調べることである。これを行なうための主なデータ処理方法としては、比較の対象とすべき候補を絞り込む処理と、絞り込まれた候補間の評価する処理、専門家の判断や仮定（2つの資料が同一の原形であるなど）の正当性を、属性値内容を調べることにより評価する処理などがあげられる。

資料の中から同一原形を持つ木片を探し出す方法としては、次の(1)、(2)がある。

- (1) 1つの資料に注目し、全資料を対象にそれと関係する資料を探し出す。
- (2) ある原形を想定し、適当なキーワードを選び出して、それらと意味的に同じグループに属する資料を探す。

また、専門家が2つの資料の原形が同一であると推測した時、資料間の関係性を評価して、そ

の推測の正当性を検証するデータ間の関係を調べるためには、2.2で述べた専門家のデータ処理方法を計算機に実装することが必要である。そのために専門家の持つ属性値間の関係の知識やデータ処理の手順をデータベース化した知識ベースを作成する。専門家の知識は個人によって異なることが多いため、専門家個々に定義しなければならない。専門家の共通の認識はシステム定義として予め準備する必要がある。類似画検索など、画像データに直接処理を行い資料間の関係を調べる機能もこの処理部分にあたる。

(4) 画面表示

データを処理した結果、比較対象となった候補のリストおよびその評価内容を画面に表示する。専門家の「目」による判断を補助するために、候補の中から専門家の指定した資料の画像を複数個同時に画面に表示する。表示される画像には、原画像、必要とする部分だけを抽出した部分画像、および、複数個の原画像をつないで生成される合成画像がある。候補となったオブジェクト数が多い場合には、基本画像すべてを同時に画面に表示させることは困難である。同時に多数の候補を表示させるために一文字オブジェクトを利用する。この場合は、一文字オブジェクトに各基本画像オブジェクトを代表させる必要があるため、共通の特徴を持つ文字を選び出さなければならない。(例えば同一文字を選んだり、同じ部首を持つ文字を選び出すなどの方法で)一文字オブジェクトを見て次の絞り込みを行い、その基本画像を表示する。

候補となったオブジェクトに対する正、否の評価結果に基づいて、表示させるオブジェクトを限定する機能や、表示位置を制御する機能が必要である。また、利用した知識ベースや結果に至った処理過程などを表示・保存しておくこともユーザーにとって有用である。

(5) 利用者の判断

専門家は画面に表示された情報をもとにして資料間の関係を判断する。新しい事実が発見された場合は、画像の合成、属性値の追加、変更などの手段によって資料の内容を更新する。更に、原形が復元された場合は、目標データベースにデータを追加する。

利用者の判断した新事実の確認のために、別の観点から再度データ処理を行い、判断の正当性を調べる場合もある。また、候補が多すぎて関係が判断できない場合には、更に候補を絞り込む必要がある。この場合には、再度、別の観点からデータ処理を行うか、または一度仮処理を行ってからデータ処理を行う。

(6) 仮処理

計算機によるデータ処理を行う前に、専門家の推測による仮定を計算機に一時的に登録する処理を行う。属性値の仮データ追加と、資料の仮合成の2種類の仮処理機能を持つ。前者は、注目している資料に属性値の情報が不足している場合に、専門家の勘などを頼りに仮定をた

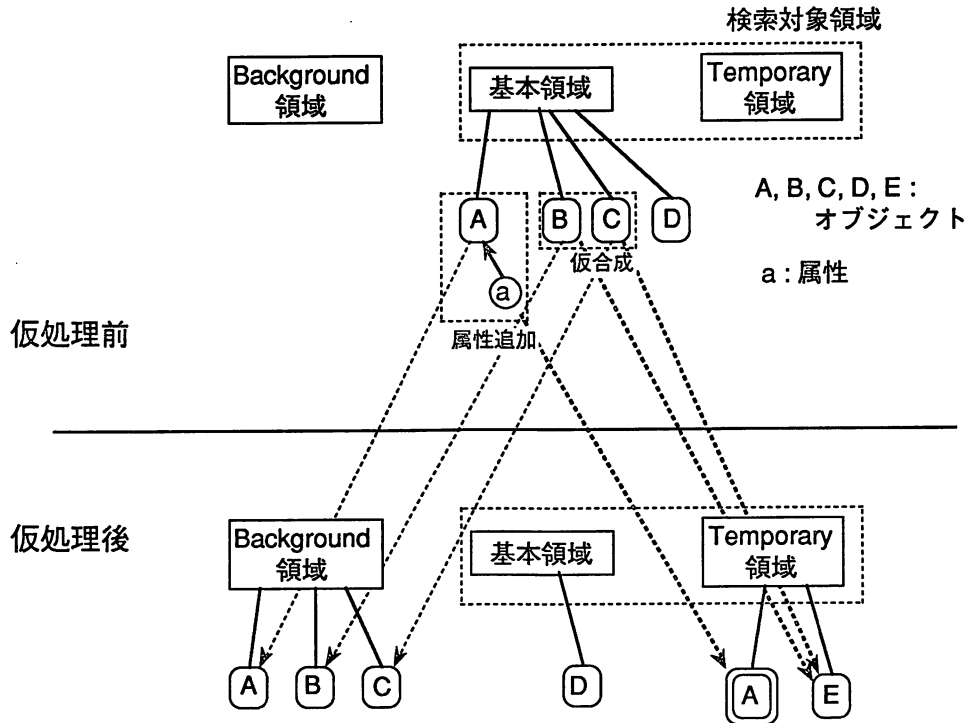


図5 仮処理におけるオブジェクトの移動

て、属性値を追加する方法である。後者は、複数個の資料を同一の原形の断片のものと仮定してそれらの資料を合成し、属性値情報を増やす方法である。仮処理によって生成された新しい仮資料は、テンポラリオブジェクトとしてデータベースに格納される。仮処理には、仮オブジェクトの生成と、仮合成オブジェクトの生成がある。

〔仮オブジェクトの生成〕

当推量的なデータ処理を行なう場合、ある基本オブジェクトに属性値の存在を仮定し、それを追加して新しいオブジェクトを生成する。新オブジェクトは仮オブジェクトと呼ばれ、テンポラリ領域に置かれる。ユーザーから見た場合、この仮オブジェクトはもとのオブジェクトと同じ名前で扱われ、データ処理の対象となる。

〔仮合成オブジェクトの生成〕

複数の資料が同じ原形を構成することが推測される場合、合成と同様の操作により仮合成オブジェクトが生成される。仮合成オブジェクトはテンポラリ領域に置かれ、データ処理の対象となる。

仮オブジェクトや仮合成オブジェクトが存在する間は、もとのオブジェクトはバックグラウ

ンド領域に移動する。バックグラウンド領域にあるオブジェクトは仮オブジェクトや仮合成オブジェクトが存在する間、データ処理の対象からはずされる。その後、これらのオブジェクトが消滅した場合は、もとのオブジェクトがバックグラウンド領域から復帰し、再度データ処理対象となる。また仮定が正しいことが発見された場合は、これらのオブジェクトが資料データベースに移動し、バックグラウンド領域から削除される。仮定が誤っている場合は、これらのオブジェクトはもとにもどる。

(7) 更新

複数の資料が同じ原形を構成することが発見された場合、それらの資料オブジェクトを合成し、新しい基本オブジェクトを生成する。新オブジェクトの属性値はもとのオブジェクトのすべての属性値を引き継ぐ。更に合成により新たな発見、例えば人名が作者名であることが発見されたり、重ね合わせた画像から新たな属性値が発見されたりする。これらの新しい属性値を新オブジェクトの属性値に追加する。新オブジェクトの生成により、もとのオブジェクトを基本オブジェクトから削除する。

基本オブジェクトと部分オブジェクトは、互いの関係を保持している。従って基本オブジェクトを合成する場合にはその関係を書き直す必要がある。また、合成により不明であった文字が新しく読めた場合、その文字に対応する一文字オブジェクトが追加される。

同一原形を持つものと同定された資料間に欠落部分がある場合には、欠落部分をオブジェクトとして定義する(ダミーオブジェクト)。すき間の文字が推測できる場合はそれを属性値とする。ダミーオブジェクトはデータ処理の対象となる。

合成操作が繰り返された結果、最終的に原形が再構成されたオブジェクトは目標データベースに格納され、資料データベースから削除される。

参考文献

- [1] 大庭, 上島:「コンピュータによる漢代木簡の索引作成の基礎的研究」, 平成元年度文部省科学研究費補助金(一般研究B)(課題番号63450050), 研究成果報告書, pp.1-103 (1990)
- [2] 杉田:「人文科学におけるマルチメディアデータベース」, 情報処理 Vol.28 No.6, pp.765-772 (1987)
- [3] 星野他:「東洋学研究支援データベースの研究」, 昭和61年度文部省科学研究費補助金(総合研究A), 研究成果報告書 (1987)
- [4] 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)
- [5] 高橋, 中川:「歴史資料データベース MUROMATI」, 京都大学大型計算機センター広報, Vol.19 No.6 (1986)
- [6] 星野, 勝村:「東洋学文献類目データベースの研究と開発」, 情処学論, Vol.25 No.2 (1984)

- [7] 打浪：「フィールド調査データ処理におけるマルチメディアデータベース」, 情報処理, Vol. 28 No. 6, pp.773-783 (1987)
- [8] 奈良国立文化財研究所：「奈良国立文化財研究所における情報処理の現状」 (1991)
- [9] 佐藤他：「民族学研究支援のための標本画像検索システム」, 情報処理, Vol. 29 No. 12, pp.1108-1118 (1988)
- [10] 加藤, 下垣, 藤村：「画像対話型商標・意匠データベース TRADEMARK」, 信学論 D-II Vol. J72-D-II No. 4, pp.535-544 (1989)
- [11] 栗田, 下垣, 加藤：「主観的類似度に適応した画像検索」, 情処論, Vol.31 No.2, pp.227-237 (1990)
- [12] 黒川, 洪：「形状情報を用いた画像の類似検索システム」, 情処学論, Vol. 32 No. 2, pp.721-729 (1991)
- [13] 山根, 坂内：「キーワード自動抽出を考慮した画像データベース」, 情処研報, Vol. 91, No. 56, 91-CV-73, pp.1-8 (1991)
- [14] 佐藤, 坂内：「画像理解のための認識ルール獲得支援システムの一提案」, 情処研報, Vol. 91, No. 56, 91-CV-73, pp.113-120 (1991)
- [15] 坂内, 佐藤：「画像データベースにおけるモデル形成」, 信学論 D-I Vol. J74-D-I No. 8, pp.455-466 (1991)
- [16] 上島, 園田, 大月：「資料復元研究支援データベースシステムのモデル化」, 信学技報, DE 91-53 pp.9-14 (1992)
- [17] 上島, 園田, 大月：「木簡オブジェクト」, Proc.Obase symposium, pp.105-112, Obase consortium (1992)
- [18] W. Kim, F. H. Lochovsky (eds.): Object-Oriented Concepts, Databases, and Applications, ACM Press (1989)
- [19] M. Atkinson, F. Bancilhon, D. DeWitt, D. Dittrich, D. Maier, S. Zdonik: "The Object-Oriented Database System Manifesto", Proc. DOOD, pp.40-57, Kyoto, Dec (1989)
- [20] 宮崎, 川越編：「特集：オブジェクト指向データベース」, 情報処理, Vol. 32, No. 5 (1991)
- [21] 牧之内：「オブジェクト指向データベース」, 情報処理, Vol. 32, No. 9, pp.1032-1040 (1991)
- [22] 横田, 西尾：「演繹・オブジェクト指向データベース」, 情報処理, Vol. 31 No. 2, pp.234-243 (1990)
- [23] 加藤, 水島：「画像情報システムのためのマルチメディアデータモデル CHEMIN」, FGIS 89-3, pp.1-8 (1989)
- [24] 大川, 小泉, 馬場口, 手塚：「不完全情報を含むフレームの階層化」, 信学論 D-II Vol. J74-D-II No. 6, pp.786-795 (1991)
- [25] 美濃, 岡崎, 坂井：「対象物の属性特徴による画像検索法—風景画像中の山を例として—」, 情処学論, Vol. 32 No. 4, pp.513-522 (1991)
- [26] 高橋, 島, 岸野：「位置情報を手がかりとする画像検索法位置による検索」, 情処学論, Vol. 31 No. 11, pp.1636-1643 (1990)
- [27] G. Booch: "The Design of C++ Booch Components", Proc. OOPSLA, pp.1-11 (1990)
- [28] 中島, 田原, 加藤：「設計支援エキスパートシステム構築シェル：MAGIC」, 情処学論, Vol. 31 No. 11, pp.1572-1580 (1990)
- [29] 中島, 大藤, 藤田：「エキスパートシステムとオブジェクト指向 OODBMS との統合」, 信学論 D-I Vol. J74-D-I, No. 8, pp.476-484 (1991)
- [30] M. R. Blaha, W. J. Premerlani, J. E. Rumbaugh: "Relational Database Design using Object-

- Oriented Methodology”, *Communications of ACM*, Vol.31 No. 4, pp.414-427 (1988)
- [31] K. Tanaka, M. Yoshikawa, K. Ishihara: “Schema Design, Views and Incomplete Information in Object-Oriented Databases”, *Information Processing* Vol.30 No.12, pp.239-251 (1989)
- [32] T. Matsuyama, V.S. Hwang: *SIGMA: A Knowledge-Based Aerial Image Understanding System*, Plenum Press (1990)
- [33] P. Coad, E. Yourdon: “Object-Oriented Analysis”, Englewood Cliffs, NJ, Prentice-Hall (1991)
- [34] P. Coad, E. Yourdon: “Object-Oriented Design”, Englewood Cliffs, NJ, Prentice-Hall (1991)
- [35] M. R. Blaha, W. J. Premerlani, J. E. Rumbaugh: “Object-Oriented Modeling and Design”, Prentice-Hall (1991)