

民博コスチュームデータベース（MCD）の過去・現在・未来 —集中型システム（nihuONE）の試用を通して—

中川 隆
国立民族学博物館

高橋晴子
大阪樟蔭女子大学

及川昭文
総合研究大学院大学

国立民族学博物館では、1984年から、身装一身体と装い一に関連する情報を集めた民博コスチュームデータベース（略称 MCD）を構築してきている。このMCDを、人間文化研究機構の研究資源共有化事業において構築された集中型システム（略称 nihuONE）へ移行する試みを行った。

ここでは、MCDの概要、MCDを検索するために用いてきたシステムの変遷、nihuONEの概要、MCDの移行作業、nihuONEの試用結果、及び、その移行過程で行った、検索システムの機能とデータ項目との関連についての検討、MCDの構成の見直し、データベース作成のあり方についての考察結果を報告する。

The Past, Present, and Future of the Minpaku Costume Database - Through a trial use of the nihuONE -

Takashi Nakagawa
National Museum of
Ethnology

Haruko Takahashi
Osaka-Shoin Women's
University

Akihumi Oikawa
The Graduate University
for Advanced Studies

In 1984 the National Museum of Ethnology started constructing the Minpaku Costume Database (MCD), which stores the information of clothing culture in the world. We made attempt to transfer this MCD to the nihuONE that is a centralized system of the National Institutes for the Humanities. This paper presents the outline of MCD, transition of the system used to search MCD, the outline of nihuONE, and trial results of nihuONE. It also makes reference to the operation results of this transition process, such as review of MCD format, as well as the relationship between the function of search system and the data item.

1. はじめに

国立民族学博物館（以下、民博）では大阪樟蔭女子大学衣料情報室（2001年4月からは、高橋研究室）と協力して、1984年から身装関連分野の情報をデータベース化するための研究を行っており、その成果は、民博コスチュームデータベース（MCD）としてWeb上で公開している。[1]

一方、人間文化研究機構では、所属する5機関が持つ多様な研究資源を共有化し、先進的な研究・教育を進めるため、研究資源共有化事業を開催しており、その一つとして集中型システム（nihuONE）が構築されつつある。[2]

今回、研究資源共有化事業の一環として、MCDのnihuONEへの移行を試行した。その主な目的は「nihuONEの評価」ということであったが、その過程で「MCDの再検討」や「データベース作成のあり方」などについても考察を行った。

表1 MCD を構成するデータベース
(2007年10月現在)

DB-ID	データベース名	データ件数	画像件数
MCD1	服装関連日本語雑誌記事:カレント	73,723件	
MCD2	服装関連外国語民族誌	1,651件	
MCD3	服装関連日本語図書	26,720件	
MCD4	服装関連日本語雑誌記事:戦前編	8,896件	
MCD5	衣服原本データベース	10,776件	39,679件
MCD6	アクセサリー・身装文化ディジタルーカイブ	5,293件	22,865件
MCD7	服装関連外国語雑誌記事	3,193件	

2. MCD の概要

MCD は、表1「MCD を構成するデータベース」に示すように、7 本のデータベースからなっており、総データ件数は 189,603 件に及んでいる。このうち、MCD1,MCD2,MCD3,MCD4 と MCD7 の 5 本は文献データベースであり、MCD5 と 6 は画像データベースである。

これらの検索項目は、例えば MCD1 については、表2「MCD1 の検索項目」の通りであり、MCD5 や 6 は、地域・民族名、現地名の他、形態、素材、構造技術などからなっている。

表2 MCD1 の検索項目一覧

MCD1の検索項目			
文献番号		時代コード	
著者名	発行年月日	時代通称名	時代通称名ヨミ
論文名	論文名ヨミ	キーワード	キーワードヨミ
掲載誌名	掲載誌名ヨミ	服装専門分類コード	服装専門分類用語
OWC	地域・民族名	身装概念コード	身装概念用語
特定地域名	特定地域名ヨミ	抄録	抄録ヨミ

これらのデータベースには、書誌的な事項に関する項目の他に、地域・民族名(OWC)、時代コード、フリーキーワード、服装専門分類コード、身装概念コード（統制検索語）など、文献の内容を分析したデータも収録している。

分析データのひとつである「OWC」とは、民族学で使われている地域・民族分類で、エール大学で開発された Human Relations Area Filesにおいて作成されたものである。MCD ではこれを簡略化して用いており、例えば、朝鮮半島は「AA」、日本は「AB」というように一部の地域を除いてアルファベット 2 文字で表している。

「服装専門分類コード」は、ファセット構造を持つ独自の分類で、横軸に生産から消費への流れを、縦軸に物としての個々の衣服あるいは関連品を挙げた形の表として定義され、英文字 2 術と数字 1 術の組合せて表されている。

「身装概念コード」は、近い内容を持った言葉をグルーピングし、それに服装専門分類コードの英文字 2 術と、数字 2 術から 5 術を組み合わせたものを付したものである。

「時代コード」は MCD 独自の表記法で、例えば、20世紀前半は「A20F」、後半は「A20S」というように表している。

現状では、雑誌記事のカレント(MCD1)と戦前編(MCD4)のように、かなり似通った項目を持つデータ群であっても、それぞれを独立させたデータベースとして扱っている。その理由は、それぞれのデータ群の性格を重視したことのほかに、検索システムの制約によることが大きかった。これについては次項で述べる。

3. 検索システムの変遷

MCD を利用するために用いてきた検索システムを古いものから順に挙げると、次のとおりである。

1) STAIRS (STorage and Information Retrieval System) 日本アイ・ピー・エム社の製品

1979 年に導入されたもので、汎用計算機上で稼動するオンライン会話型の情報検索システムである。

当時のシステムでは、一般的には、漢字データを扱うことができなかつたためカスタマイズを行い、漢字データを JIS コードで格納し、数台の特殊な端末に表示できるようにしていた。検索は、漢字データからは行えず、英語のように分かち書きされた 1 バイト文字列のテキストから単語単位で行うというのが基本であった。従って検索の対象となる項目にはすべて、半角カタカナで分かち書きしたデータを付加しておく必要があった。汎用計算機の端末で漢字が使用できるものが導入されたのは、1980 年代になってからである。

2) JAIRS (JApanese Information Retrieval System) 日本アイ・ピー・エム社の製品

1991 年に導入したもので、1) と同様に汎用計算機上で稼動する製品であるが、漢字データでの検索が可能になった。

1993年10月には、ゲートウェイを追加し、インターネット経由、又は電話回線に接続した300～9600bpsのモデムを経由して、パソコンのVT100端末エミュレータを使い、民博外からもMCDが利用できるようになった。

3) MMIR (Minpaku Multimedia Information Retrieval)

1997年に民博が策定した仕様に基づいて開発したシステムで、UNIX系のサーバ上で稼動するものである。全文検索エンジンとRDBMSを組み合わせ、C++でプログラミングしたCGIアプリケーションである。MMIR内の横断検索機能や、検索結果の一覧や詳細表示時に表示する項目を指定したり、表示順序・長さ等もユーザが指定できる多機能なシステムである。一方、民博の写真資料と標本資料のデータを強く意識して設計されていて、各レコードに付けられる画像は1枚だけ又は4の倍数枚でなければならないという制約がある。

1999年には、画像用のサーバを導入し、大量の画像を公開できるようになった。

4) MCD検索システム

1997年末から民博内で独自に開発を始め、次に述べるMARSに移行するまで使用したUNIX上で稼動するCGIアプリケーションである。3)のMMIRでは制約があった任意枚数の画像表示や、検索条件を入力する画面のカスタマイズなどの要望に応えて開発されたものであるが、一方、画像の管理方式をMCDに特化しすぎたためにメンテナンスが困難になってきた。

5) MARS (Multi Archives Retrieval System)

3)のMMIRの後継となるもので、2003年からレンタルを始めたシステム向けに、民博の定めた仕様に基づいて作られた汎用の検索システムである。JSPとjavaによって記述したもので、RDBMSやXML関連の標準的な技術を用いたシンプルなシステムとして構成されていた。しかし、その後、様々な拡張の要求が出され、データベースごとに検索条件の入力画面を用意できるよう

な機能（例えば、条件をプルダウンメニューで指定できること）や結果一覧をソートする機能などを追加していく、次第に複雑なものになってしまった。（図1,2）

また、Z39.50のサーバ機能及びゲートウェイ機能を持っており、データベースの横断検索も可能なシステムである。

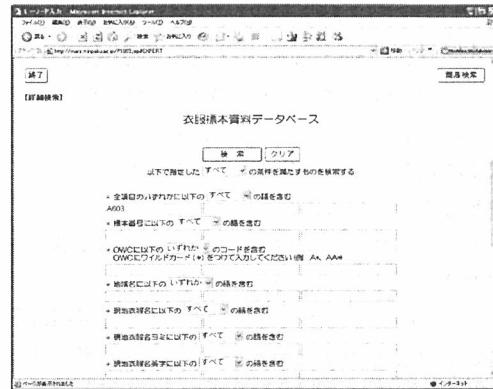


図1 MARSの検索条件入力画面の例



図2 MARSでの詳細表示画面の例

4)の検索システムを除いて、汎用的に使用できるシステムであり、MCD以外の民博の他のデータベースにも使用してきている。なお、これら以

外に、プロトタイプとして作成したものがいくつあるが、ここでは省略する。[3]

4. システムの仕様とデータ作成との関連

前述のように 1) の STAIRS では検索が行えるのは、1 バイト文字で分かち書きされた項目のみという制約があり、検索項目には、例えば、抄録にもすべて、半角カタカナで分かち書きしたデータを用意しておかなければならなかった。従って、データの作成には多くの手間と人手を要した。また、このころでは、データの入力にパンチカードが使われていた時代であり、専門のパンチャーに依頼してデータ入力をしていた。

また、1) や 2) の時代の汎用機の端末では、カタカナと英小文字とが同一画面上に表示できないという制約があったため、同じような項目から構成される日本語文献と外国語文献のデータベースであっても、それぞれ別のデータベースとして作成せざるを得なかった。

次に、画面のレイアウトや検索時の振る舞いについてふれると、まず、データベースごとに設定できる事項は、1) や 2) のシステムでは、項目名や表示順序をデータベース設計時に指定できる程度であった。また、検索の操作は、専用のコマンドまたはコマンドが割り当てられたファンクションキーによって行う方式であった。システムには USER EXIT 等と呼ばれる拡張のための出入り口がいくつか用意されており、多少の拡張は可能であったが、アセンブラー言語等の知識が必要で、カスタマイズは容易ではなく、必要最小限のものを導入時に業者が行うという程度であった。

カスタマイズの制約から、利用記録のログ情報も不十分であり、利用者の需要動向を知るために、データベースを分割しなければならないといったケースもあった。例えば、利用者がどの年代のデータをよく参照したかを知るために、同じ雑誌記事のデータであっても、カレント(MCD1)と戦前編(MCD4)のように分割しておく必要があった。

3) のシステム以降では Web ブラウザをクライアントとして使用するものとなり、画面上に機能

を表す名称を付けたボタン等を表示しておき、それをクリックする形で進めていくようになった。一方、3) や 4) を導入する頃には、適当な機能・性能を持った製品としての検索システムが見当たらず、仕様を策定し特注するという形になり、仕様の策定に手間取るようになっていた。また、この頃から、3) では C 言語、5) では JSP についてのある程度の知識があれば、導入後に民博内でもカスタマイズが行えるようになった。

5. nihuONE の概要

nihuONE は、Web ブラウザ上での簡単な操作により、データベースの定義から項目の設定、データのアップロード、Web 公開までが行えるシステムで、コンピュータの専門家でなくてもデータベースを作ることができるように配慮されたものである。



図 3 nihuONE の項目定義画面

データベースの作成を支援するための機能として、データのチェック機能やテーブル変換機能がある。テーブル変換機能とは、予めコードとその値の入ったテーブルを用意しておけば、アップロード時にコードに対応する値に展開し、検索や検索結果の表示の対象にできるというものである。また、コードにミスがないかチェックしてくれる機能 (RDBMS での参照整合性) も持っている。別途にチェック用のプログラムを作る必要がない。

また、nihuONE 内の自データベース又は他データベースのレコード ID を入れておけば、そのレコードへのリンクを生成してくれるという機能も用意されている。例えば、アクセサリ(MCD6)の各レコードに関連する衣服標本資料(MCD5)のレコードを表示させたい、といった場合に使用できる。(図4, 5)

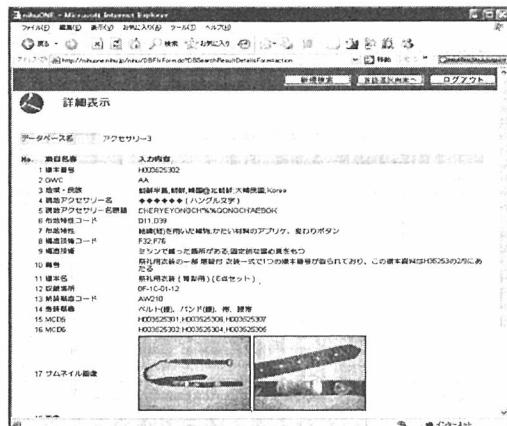


図4 nihuONE の詳細表示画面の例



図5 関連資料へのリンク先の表示例

その他、リプレース機能も用意されており、指定した項目の値を、予め定義した文字列の中に埋めこんで、別の項目の値とすることができる。この機能を使えば、画像ファイル名を入れた項目から、画像を表示するための html を生成して別の項目にセットしておくことも可能である。

検索機能としては、全項目からの全文検索や項目指定の検索、絞込み検索機能等が用意されている。

また、横断検索機能もあり、指定した複数データベースを対象にして、全項目から検索したり、同一項目名を持つものについては項目指定で検索したりすることが可能である。

結果の表示画面は、一覧と詳細の他、KWIC 形式での表示が用意されている。

さらに、従来の検索システムではあまり見られなかった統計機能も用意されている。2 項目まで掛け合わせた頻度統計を取ることができ、結果をダウンロードすることもできる。これにより、データベースを「検索する」から「活用する」システムへと発展させている。

6. nihuONE への移行

こういった特徴を持つ nihuONE に、MCD を移行した過程について簡単に述べておく。

MCD の場合、検索システムに載せる前の文字データは、汎用機の時代では、作成から編集、保管までを独自形式のシケンシャルファイルによって行っていたが、最近では、エクセルをマスターとしているため、nihuONE に載せる作業は簡単であった。

まず、Web ブラウザ上からデータベースの定義-->項目の定義-->一覧表示定義-->詳細表示定義へと操作を進めた後、エクセルから出力した CSV 形式のファイルをアップロードするだけによかった。

そのさい、OWC、服装専門分類、身装概念、構造技術、布地特性などのコードを入力している項目については、テーブル変換機能を使用して、コードに対応する値を表示させることにした。

項目の定義時には、漢字項目とそれに対応するヨミ項目がある場合には、とりあえず、両方ともデータベースに取り込むことにし、ヨミ項目の方は非表示に設定して、表示画面を簡潔なものにした。

なお、画像については、民博の MARS 用のもの、すなわち民博の Web サーバに置いているものをそのまま使用することにした。そのための html の作成は、先に述べたリプレース機能を使うという選択肢もあったが、MCD の場合、1 レコードに付く画像の枚数が不定であるため、エク

セルのVBAを用いて数ステップの簡単なプログラムを作り、エクセル上でhtmlに変換しておくことにした。

アップロードについては、今回は、エクセルから一度に出力するCSVファイルのレコード数を、1万件以下になるようにし、数回に分けて行った。アップロードの開始からデータのチェックと検索エンジンへのロードが完了するまでの処理時間は、1万件当たり16分程である。アップロードが完了すれば直ちに利用できる。

7. データベース分析

nihuONEの特長の一つである統計機能を使用してみた結果を紹介する。

1) 頻度統計の例1

例えば、「現在の研究者が、近代日本における身装分野のうち、さらには、どういった主題に関心があるのか」を知ろうとする場合、服装関連日本語雑誌記事:カレント(MCD1)を、地域・民族が日本(OWC=AB)という条件で検索すると57693件ヒットする。その内訳として時代コードの頻度を見ると、20世紀後半に関する研究が58%を占めているものの、近代にあたる19世紀後半および20世紀前半に関する研究は、頻度統計の6位、7位に位置していることがわかる。そして、この文献件数は、1562件にのぼる。さらに、この1562件について、統制語にあたる身装概念で頻度統計をとると、「紡績業、製糸業」、「服装史」、「西欧化」や、「風俗」の主題についての関心が1位~5位までを占めることが把握できる。

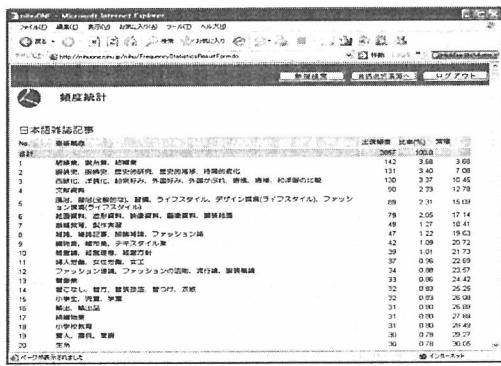


図6 nihuONEによる頻度統計の例

2) 頻度統計の例2

エクセル等の表計算ソフトと頻度統計の機能とを組み合わせれば、もう少し複雑な集計も可能である。簡単な例として、例えば上で使ったMCD1に20世紀後半という条件を付けてヒットしたものの頻度統計を取って、それをダウンロードしたものを、OpenOfficeのCalcにあるデータパイロット(エクセルのピボットテーブルでも同様)を使って、クロス集計してみると表3「クロス集計の例」のような結果が得られる。

表3 クロス集計の例

合計・出現頻度 身装概念コードの2桁目	身装概念コードの1桁目						合計結果
	A	B	C	D	E	F	
P	611	5221	10265	13161	4359	9594	43211
Q	3198		51	141	9145	1	12536
R					4	2	34
S					21	12	64
T						6	39
U							3617
V					8	615	1405
W				46	4	49	185
X						1328	1481
Y					306	8	1978
合計結果	10947	5573	10328	14698	16166	10212	67924

さらに、これを発行年の項目で絞り込んで、発行年代ごとに同様の処理を行うことによって(発行年代ごとの集計表は割愛するが)、20世紀後半は、全体的には、(Q-素材)の(E-製作・製造業)に関する研究、並びに、すべてのアイテムを含む(P-総記)の(D-デザイン)に関する研究が多い。しかし、各年代のクロス集計をみると、80年代と90年代には、(P-総記)の(C-ビジネス・物流)および(F-小売)に関する研究が増加し、とくに90年代には、(F-小売)に関する記事が、80年代の約3倍となっている。この結果から、研究の対象が、川上の問題から川中・川下の問題に移行していることが理解できる。衣服の生産地が国外に移り、国内ではマーチャンダイジング等の流通に力を注いでいる日本の繊維・ファッショング産業のあり方が、そのまま研究に反映されている結果となっているということがわかる。

3) KWIC機能の使用例

次にKWIC機能について触れておくと、身装関連分野では、第二次世界大戦後、「服装」「服飾」「衣服」「被服」の概念が混乱している状態が続

いている。検索項目を‘抄録’に指定して KWIC 検索を行い、後方の文字列でソートしてみると、一覧表示を見ればその混乱状態を具体的に把握することができる。こういった言葉の概念規定に本検索は有効である。

8. nihuONE の評価

nihuONE を試用してみた結果、民博の仕様に基づいて作ったわけではない nihuONE でも、MARS 等のシステムと実用上は同等とみなせる検索機能を実現できていることが分かった。ただ、MARS では、図1「MARS の検索条件入力画面の例」のように、検索条件の指定画面で、どの項目にどの用語が含まれているといった条件を複数指定できるため、1度の操作で行える検索が、nihuONE では絞込み検索を使って何ステップかに分けて行わなければならないといった手手続きが必要となる。一方、MARS での利用ログを見てみると、約2年間に MCD について約6万回の検索が行われ、その約 3 分の 2 が検索項目を指定しないで行われていたという結果が得られており、複雑な条件指定が使われる頻度はかなり少ない。利用ログをさらに詳しく分析すれば、今後の検索システムの要件の決定にとって有用な情報が得られるものと考える。



図7 nihuONE の詳細条件入力画面

一方、nihuONE では、プログラマや SE 等の技術者の支援がなくても、パソコンで作ったデータをもとに、簡単にデータベースを作り、Web に公開できるという大きな利点がある。MARS のようなシステムでデータベースを作成する場

合には、設定ファイルを定義し、システム用のデータを準備した後、データベースへロードし、高速検索用にインデックスを生成するためのアプリケーションを起動するといった複雑な設定や手続きが必要となる。例えば MARS では、設定ファイルは複数の XML ファイルからなり、その間に矛盾のないように慎重な指定が必要である。また、データの準備に関しては、DTD と XML 形式のデータを用意しなければならない。nihuONE では、前述のように Web ブラウザでの GUI による定義とアップロードのみで、Web 公開までできてしまう。このため、研究者自身によってデータベースの作成、データの追加・修正、設定の変更などを行うことができ、データベースを管理・運用していくことができる。さらに、前項で述べたような研究支援機能として KWIC や頻度統計機能まで使える等、大きなメリットがある。

なお、検索速度や頻度統計等のレスポンスについては、今のところ、いずれもほぼ瞬時に完了しているため問題とはなっていない。

9. 今後の展開

3章「検索システムの変遷」で述べた 1) と 2) のシステムでは、カスタマイズはほとんど困難であった。その後のシステムでは、プログラムのソースコード類が提供され、カスタマイズの敷居も低くなり、変更の幅も広くなった。半面、ユーザからの要求レベルも高くなり、それにあわせて拡張に拡張を重ね、複雑怪奇なシステムへと変わっていき、バグの発生が多くなったり保守が困難になったり、という面も出てきた。

そんな中、nihuONE のように今のところ民博側ではカスタマイズが行えないシステムを試用して、検索システムの要件について改めて考えみると、システムのカスタマイズに手間をかける前に、YAGNI (You Aren't Gonna to Need It.) と割り切って、データの見直しやデータ作成に手間をかけてみることも大切ではないかと思われる。

検索システムの構築や維持に割振れる資源の規模によっても違いはあるが、また、カスタマ

イズにも設定ファイルで指定できる範囲（設定ファイルにも単純な設定しかできないものから、テンプレートや XSLT のようにプログラミングの要素を持ったものまでいろいろであろうが）のものから、ソースコードを修正するものまでいろいろであるが、いずれにしろ、レイアウトや画面遷移に凝ろうとするとかなりのコスト（時間と経費）がかかることになる。

また、MCD については、データ項目の見直しや、データ内容についても見直しを検討する時期に来ていると考えられる。

まず、項目についてはデータ作成の省力化の点からも見直す必要がある。検索システムの更新時には、移行作業に追われ、とりあえず移行することにのみ関心が向く傾向にあり、項目の見直しは十分に行えているとはいえない。例えば、旧システムでは必要だった分かち書きや読み仮名データが、現在のシステムにも引き継がれている、といったことである。この点はデータ作成の手間と利便性との兼ね合いから再検討する必要がある。

また、nihuONE の横断検索機能を有効に使うには、検索項目名をそろえておくことも必要となる。

さらに、MCD を単なる検索の対象から分析の対象へと発展させるためには、コードで入力している項目の内容も見直す必要がある。例えば、「発行年」の項目は、年と月が入力されている形が基本であるが、他に月が複数入っていたり、年と月の組が複数入っていたりするものもある。このため、頻度統計の機能を使って、発行年と身装概念コードのように 2 項目を掛け合わせた集計を取ろうとすると、組合せの数が多くなりすぎて処理できない、とのメッセージが出る。これは「年」だけを独立させた項目を設ければ防げる。即ち、検索のためのコードから、分析のためのコードに変えていく必要がある。

データベース全体の構成面では、データ群の分割は、検索時のレスポンスや、操作の手間といった面から見直したい。高速な検索システムでもいくつものデータベースを対象に処理すれば、遅く

なるのは明らかであるから、まとめられるものはまとめた方が良い。利用者の需要動向の調査を取りたいために分割しているデータベースについても、検索条件の入力画面で、どの範囲から検索するのかが指定でき、利用記録にその情報が残るような機能があれば、分割する必要はなくなる。

民博では、2007 年度末に計算機システムの更新が行われ検索システムも新しくなる。MCD にとっては、当面は、新システムへの移行や現在開発中の身装年表データベースとの連携などが課題となるが、nihuONE についても並行して稼動させていき、利用記録等を参考に今後の方針を検討していくことになろう。

参考文献

- [1] <http://www.minpaku.ac.jp/>
なお、現在、MCD の研究およびデータ作成は、MCD プロジェクトが行っている。
- [2] <http://nihuone.nihu.jp/>
- [3] ハ村広三郎、高橋晴子他 “民博コスチュームデータベース” 情報処理学会: 人文科学とコンピュータ研究会、1993.11.26
- [4] 高橋晴子 “民博コスチュームデータベース [MCD] の現状および今後の展望”、情報管理、35-8、pp.665-674(1992)
- [5] 木暮 仁 “システム開発におけるユーザーニーズは絶対か?”等、システム部門 Q&A <http://www.atmarkit.co.jp> (2003-)