

人文系利用者側の視点で構築するデジタルアーカイブ ー 画像資料とデータベースの有機的・可変的リンク ー

赤間 亮

立命館大学アート・リサーチセンター rat03102@lt.ritsumeikan.ac.jp

データ（文字・画像）・検索システム・インターフェイスまで、全体にわたって人文系研究者自身が構築した学術データベースを WEB 上で公開するプロジェクトを開始した。人文系で求められる情報公開のあり方を問う上では、格好のサンプルとなっていると思われる。また、人文系研究者が求めるデジタルアーカイブの方向を示す。

Digital archives constructed by aspect on the humanities system user side

AKAMA Ryo

Art Research Center, Ritsumeikan University

The projects that the science data base which the humanities system researcher had constructed is opened to the public on WEB was begun. This seems a shape sample in asking the ideal way of the information opening to the public requested by the humanities study. Moreover, the direction of the digital archives which the humanities system researcher requests is indicated.

1.はじめに

デジタル技術の進歩とともに、高度な情報検索、情報配信が可能となっており、その急激な変化には驚くばかりである。このことは、逆に人文系の研究者にとっては、それらの技術情報を取入れ、吸収していくことの困難さを、ますます感じさせられている状況であるといえよう。人文系の学術研究の世界においても、デジタルディバイドが発生しはじめている。

筆者は大学では、国文学を教えている。国文学の基本である”註釈”とは、用例検索の学問であるとも言えるが、電子テキストが出回り、これまでのように多大な労力を払わなくても註釈作業が可能な状況ができつつある。しかし、伝統的な教育方法では、辞書は勿論のこと、その作品に、時代的に近接する作品を対象に、人海戦術によって用例を探り、最も適当と思われる註釈を付すという作業を繰り返している。この作業の目的は、時間をかけて労力を払うことそのものであるようにも見え、こうした苦労が次回以降の同様の調査に役立つと教え込む。これが註釈学の神髄であるかのように教育している教員は、ネット上で用例が探せることに熟知しているとは考え難い。しかし、学生はインターネットで手軽に検索すれば、ある程度は、材料が集められることに気付いており、そのばかげた労力にうんざりしはじめている。古典文学への興味は、ますます失われていくという結果が生みだされている。

情報化社会の醸成にあたっては、人文系研究者の役割はコンテンツ作成にあるとも考えられているようである。政府予算を中心とした膨大な費用を使った強引なコンテンツ制作の事業に組入れられた研究者は、人文系研究にはあまりなじまない、厳格な締切を設定した「宿題型」プロジェクトの中で、働かされている。このところ各大学が積極的に進めている電子博物館・電子図書館プロジェクトの中で、人文系研究者が自ら積極的に進めているプロジェクトはあまり存在していないのではないと思われる。

私は、立命館大学アート・リサーチセンター（A R C）を活動拠点としているが、ここでは、人文系の視点から、デジタルアーカイブとデータベース公開を推進している。素材の選定や、デジタルコンテンツの製作、データベース構築、インターネット配信まで一連の流れは、筆者が統括しており、その意味においては、他にあまり例のない事例ではないと思う。

本稿では、人文系の視点から人文系研究者自らが作り上げた資料活用のためのデータベースシステムを紹介して、多くの機関・大学で進められているデータベース公開の参考に供したいと思う。

2. WEB 公開までの流れ

A R Cで開発しているデータベースシステムは、マルチメディア対応であるが、ビューアーはすべてブラウザやそのプラグインソフトに任せており、単純に文字データベースと画像・動画等のコンテンツの置き場所をリンクさせて読出してくるという構造である。したがって、最も重要な作業は、文字データベースの構築という作業である。

この文字データベースの構築は、今を遡ること約 15 年前から始ったものである。これが他のプロジェクトとの最大の相違点である。これまで、この素材を使いながら様々なデータベース構築作業を続けてきた。既に 3 年前には、画像データベース公開プロジェクトにも加えられ、ある程度の成果をおさめた経験もある。

(<http://pro2.project.mnc.waseda.ac.jp/nisiki/>) この時は、基本的な役割は、文字データの提供にあった。しかし、今回は、画像コンテンツの製作から、ネット公開用のシステム構築まで独自に進めたものである。

データベースの文字列入力に関しては、表計算、エディタ、データベースソフト、それぞれ一般的なソフト（頻度の高いものは、エクセル、秀丸、桐である）を使い、ストレスなく構築していく。筆者の専門は、文学・演劇であるので、対象としている分野はそこに特化している。演劇情報は、時間芸術と呼ばれる分野であり、上演された舞台そのものは残らない。それ故に、派生した物品や情報が組合されて初めて再構築されるものであり、また、その一舞台に対する情報が複雑で、またマルチメディア向けでもあり、リレーショナルなデータベース構造が必要とされる。作業を進めながら、事典的なデータベースを成長させており、そのデータベースに対して＜表引き＞機能を縦横に利用することで、効率化が図られている。情報スキル教育の素材として

は最適であることを付加しておく。

コンテンツ画像や動画の作成に関しても、業者丸投げ方式はとらず、入力ノウハウの開発から行なっている。解像度や圧縮率などは、研究・利用目的と相互に関連しているもので、素材毎に専門家の判断する必要があるため、この感覚を独自に磨いていく必要があるのだ。そうすることで、人文系研究にとって最も重要な「大量のコンテンツの公開」という必須要件が安価に早く実現させられるのである。

画像形式はできるだけ単純化しており、静止画に関しては、様々な形式があるがWEB 公開用には JPEG 形式に統一していきたいものと思っている。また、どの所蔵機関も画像作成にあたっては、一旦フィルム化しておき、そこからデジタル画像へ落とし込むのを標準化しつつあるようであるが、これはあまりお勧めできない。デジタル画像そのものや、フィルムには寿命があるため、時間とともに過去の画像となるフィルムは保存しておいても意味がない。できるだけ原物を損ねることなく、直接取込むことが必要であるとすれば、今後は、デジタルカメラが主流とならざるを得ないであろう。現在の技術で記録された映像は、永久に完全で有用な画像と思ひこむのは、誤りではないだろうか。

最後のデータの WEB 公開にあたっては、ほとんど専門知識を必要としないファイルメーカー Pro の WEB 公開機能を利用している。すでに、Ver.4 の段階で登載していた機能であり、そのこと自体には、新味はない。ファイルメーカーの独自言語である CDML を使ってカスタマイズすれば、高機能を持ったアプリケーションにもあまりひけをとらない見栄えのあるデータベースが仕上がる。

最新の技術を使うのではなく、ある程度定着した技術の範囲であれば、情報科学を専門としない人文系研究者にとっても、こうしたシステムを、情報整理の援助システムとして利用するのにそれほどの負担を感じない。実際、本格的に稼働させるとすれば、システムの能力に限界があるため、相応のシステムを組む必要があるが、その場合でも、人文系研究者のリハーサルを経たサンプルデータベースが存在しているのとないのとは、大幅な経費の節減ができるだろう。

これらの作業の中で、難しい点として感じているのは、データベースを開発するのに特有の問題ではなく、人文系の研究者同士が共同研究時に感じるものと同じものである。これまでも、例えば調査カードを作る際、最も時間のかかる作業は、凡例の共通理解を得ることであったが、文字データ作成時におけるデータの入力者の凡例理解がそれである。専門性のある判断をしばしば要求されるデータベース入力こそが、人文系の研究にとって有効なデータベースであるが、すべての入力を専門性のある人材が入力できるわけではない。また、データベースの入力の役割分担と入力手順などのプロデュースをする人材が必須となる。しかし、少なくとも筆者の分野では、こうしたプロデュースの出来る人材は、ほぼ皆無に等しい。

3.作成中のデータベース (<http://www.arc.ritsumei.ac.jp/db/top.htm>)

データベースはA R C内に設置された UNIX サーバーから、同じくセンター内の複数のP C（WINDOWS-NT、WINDOWS98）に飛ばして運用している。

筆者が開発したデータベースとして次のものが稼働している。

1) テキスト系データベース

○記録・年表検索システム

・江戸歌舞伎興行年表 ・近代商業演劇興行年表 ・現代商業演劇興行年表

日本の演劇の興行システムの特殊事情を理解しないと、年表化は不可能である。しかし、そこに出現する特殊事情は、専門家の間では常識であることが多く、また逆に例外も多いため、それらをすべて年表に反映させると情報は、一般からは理解を得られず、専門家からもくどく見える。また、同じ日本の演劇興行年表であるにも関わらず、大きく江戸期、明治～第二次大戦、戦後という日本演劇興行史の3時代区分を設定している。これは、社会制度の変化とともに、興行制度自体に変化があり、共有化できない項目も存在するからである。一般的に考えれば併合して一緒の操作で使える方が便利であると思われる向きもあるだろうが、内容面でもみて、ほとんどこの三区分別の中で運用をするので、分けておいて構わない。システムの言えば、P Cの連結による運用のため、分散型運用の方が安定を得られるという点もある。

これらのデータは、演劇資料を管理する上で、基本となる情報であり、次の課題は、このデータと資料別データベースとのリンクである。

○人名等検索システム

・歌舞伎役者データベース

人名に関わるデータベースであるが、日本芸能に関わる人名の特殊な点は、同じ名前が代数を重ねることである。一人の芸能人からみると、その名前は年とともに芸名が変わってゆき、生涯で複数の名前を名乗ることになる。一つの名前からみると、別人が同じ名前を名乗ってゆくことになる。名前を受継ぎ、名乗り始めることを襲名といい、普通の名前とは違うので、これを「名跡」と呼び、そしてこの芸能系の専門家になるためには、「何代目」を頭の中でどのような処理して覚えるかがポイントでもある。この点は、データベース化するにあたっての難題でもあるが、この難題に一つの解答を与えたつもりである。

・歌舞伎・浄瑠璃役名データベース（→コンテンツ部分は土田衛氏作成）

このデータベースは、歌舞伎や浄瑠璃に文学的な考察を加えるには、きわめて有用なデータベースであるが、問題は役名の多様さである。これは、単純によりがなを付与するだけでは解決できない。事実上、まったく同じ人物を表現する役名であっても複数の名称で呼ばれている場合が多いため、統一が難しい。この課題は解決できないままにしている。

○作業共通凡例検索システム

・漢字統一処理字典 ・作業進行管理システム

作業凡例の共有化を WEB 上ですること、作業の効率化を図っているものである。こうした情報公開は、人文系ではあまりされたことがない。漢字統一処理字典については、完全公開しており、知る人ぞ知るデータベースなのである。

○文献目録

- ・日本演劇文献目録検索システム

○蔵書検索システム

- ・映画パンフレット（ARC所蔵品）検索システム

○簡易版作品本文検索システム

- ・道頓堀花みち本文等検索システム

以上は、いずれも一般的に見られるデータベースであるが、この特徴は、数分でシステムの運用を始めたことである。個人構築している小さなデータベースが瞬時に公開できるのがこのシステムの有利な点である。

2) 書籍閲覧システム（電子画像）

○ページ物の閲覧資料検索閲覧システム

- ・テレビ・映画シナリオ検索システム 件数 40000 冊、画像数 40000 枚
- ・書誌情報イメージ閲覧システム 件数 1000 冊、画像数 6500 枚
- ・デジタル書籍閲覧システム 件数 2000 冊、画像数 20000 枚

「書籍閲覧システム」の場合、厚物・薄物・卷子物・複数巻物・辞書年表物によってページ表示方法を変える必要がある。書籍閲覧システムは、その点に考慮して作り上げたものであるが、まだ、改善の余地は残されている。

「シナリオ」や書誌情報イメージのシステムは、一般的な蔵書検索システムにイメージを組合せた蔵書検索システムであるが、画像数を逐次増やすことで、書籍閲覧システムに転じていくものである。完全に書籍閲覧システムとして完成していなくとも、部分的にでも書誌情報を補うページ画像を連動させることで、利用者の要望に応えた蔵書検索システムとして稼働することになる。

3) 画像検索システム

○一枚物の資料閲覧検索システム

- ・浮世絵検索システム 画像数 50000 枚、
※続物や組物などの表示（画像同士の関連づけ）が特徴
- ・舞台写真検索システム 画像数 130000 枚
- ・江戸時代版元検索システム 画像数 900 枚
- ・片山家能楽保存財団能楽写真検索システム 画像数 10000 枚

ARCのデータベースの特徴は、この画像検索システムにつきると思われる。単純に量的な面でも、日本で有数の画像数を誇る。スタンドアロンで動く汎用画像データベースシステムには、未だ使い勝手のよいものが少ないが、ネット経由でもスタンド

アロンでも両様で運用できるのも特徴である。スタンドアロンでも、まったく同様に動く。このシステムを運用するようになり、絵画や写真に関する共同研究が可能となった。この点が、一年前とは隔世の感を抱かせる、最も大きな改革である。

4. 有機的リンクと可変的リンク

さて、これらのデータベースの特徴をあげる。

1) 検索結果のほぼ全項目にそれぞれリンクを張っていること。

この機能は、最近、OPAC など、広汎な利用が見込めるデータベースにも採用されてきている機能であるが、学術的な限定目的のデータベースとしては、あまり例がない。ある程度の知識のあるのある利用者がこのデータベースに入った場合、一回の検索に留まらず、無意識に複数の検索を試みることになる。これにより、データベース利用の目的が「絞込み」のためにあるのではなく、多様な関連情報の掘り起こしにあることが分かってくる。検索という行為によって、興味の連鎖が作られ、データベースにアクセスした当初の目的を忘却して使い続け、最終的には関連性の薄い情報までも入手することが可能である。この様に、人文系の感覚からは、連鎖の糸を途切れさせない機能が求められていることを強調しておきたい。

2) 外部データベースへのリンクも可能なこと。

これも、特に珍しい機能ではない。この外部への連絡が、現在操作しているデータベースの中に留まらず、しかも、それを利用者に意識させることなく、外部データベースとも行き来することができる機能である。これは、リレーショナル的な検索に過ぎないが、データベース間の”ハシゴ”ができることも言い換えられ、前項でふれた連鎖の糸をより有機的に繋げていくことができるため、人文系の研究者が使うデータベースには必須の機能と言えよう。

3) 静的な文書上から検索語を指定してリンクさせられるので、結果が動的に呼び出せること。

長い標題になったが、具体的に言えば、デジタル文書で論文を書いたとする。この論文の注や、図版、図表にあたる部分は、文書の然るべき箇所にリンクをはり、特定の検索条件を指定しておく。ある結論付けを行なった手許の文書（論文）は、時間を経ても変更されずに公開されつづけるのが常である。一方でデータベース自体は、日々更新されて情報量が大きくなっていく。すると、リンクの張られたデータベースは、成長しているため、将来は、文書の結論を否定する結果を表示する可能性もある。もしも、そうなったら、この論文は、作成された時点では、確かに正しい結論を導き出していたのだが、時間を経て、自らを否定することになる。

人文系の研究では、一点の新出資料により、定説や常識を覆すことがままある。こうしたダイナミックな研究の動きが、データベースとの連携により、改めて指摘するまでもなく提示されてしまう。結果が可変的となる利用方法もデータベースに是非持たせておきたい役割である。

5. 人文系コンテンツのデジタルアーカイブが行なわれる条件

最後に、実践的にデジタルアーカイブを行なったきた経験から、こうした仕事を進める上で、筆者が獲得し、一般的に通用すると思われる条件を列举しておきたい。

人文系研究者にとってデジタルアーカイブに期待するものは、大量の資料・大量の情報を整理する手間の軽減という一点にかかっている。考えたり、発想したりするのは、我々であって、コンピュータではないので、偏に整理の支援システム、しかも短時間での構築のできる道具を追い求めることになる。したがって、作り込みに時間のかからない、手軽に使えるアプリケーションの開発がまずは先決であった。

また、既に紙媒体により普及している資料をデジタル化して何になるのか。各機関で整理公開している資料は、図書館を通じて複写を頼めば、使い易い紙媒体で送られてくる。それよりも費用のかかるデジタル化など、する気は起きない。これも、コピー代よりも安価にデジタル化ができるのなら、大いに進むだろう。安価に最低限の画質でデジタル化するノウハウの構築が必要であった。

さらに、大量に資料があり複製も作られたことがなく、多くの人が接触できない分野の資料からデジタル化すべきである。各機関の資料庫に未整理のまま眠っているものをデジタル化すべきである。目録などは完備している必要はない。ナマのまま公開することで利用者のネットワークができ、利用者の側で目録化が進む。このインタラクティブな仕組の構築も必須であろう。

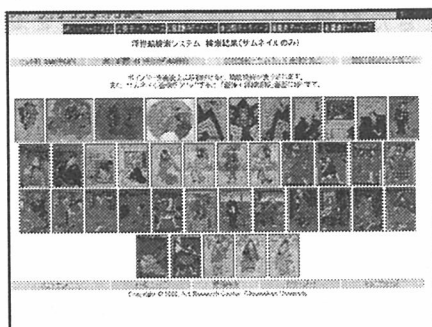
このところ、国立大学や大規模博物館を中心として書籍の紙面そのものをデジタル画像化した電子図書館の運営が盛んになってきた。しかし、構造は、国の情報化政策によって、莫大な資金が注ぎ込まれ、遮二無二文化系のコンテンツを供出させられているというのが実態であった。人文系の研究者からの積極的な参加者はほとんどなかったと思われる。これは、単純に研究者が必要ないと思ったからである。人文系の研究者が使いたくなるシステムは、人文系の研究者自らがプロデュースせざるを得ないのである。

浮世絵検索システムのインターフェイス

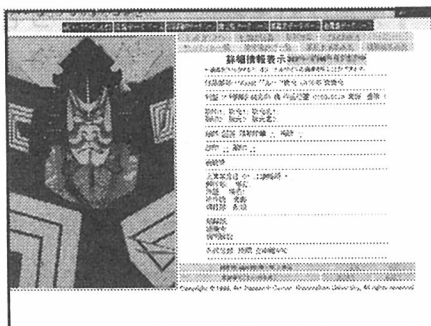
0. 検索初期画面

The screenshot shows the initial search screen of the 'Ukiyoe Search System' (浮世絵検索システム). The interface is in Japanese. At the top, there is a title bar and a header section with the system name. Below this, there is a main search area with various input fields and buttons. The fields include '検索条件' (Search Conditions), '検索' (Search), '検索結果' (Search Results), and '検索履歴' (Search History). There are also buttons for '検索' (Search) and '検索結果' (Search Results). The interface is designed for a web browser, with a clear layout for entering search criteria and viewing results.

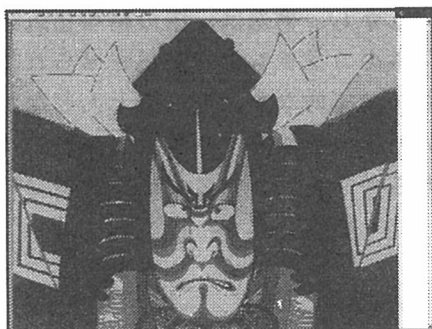
1,検索結果（サムネイル表示）



2,検索結果（詳細解説表示）



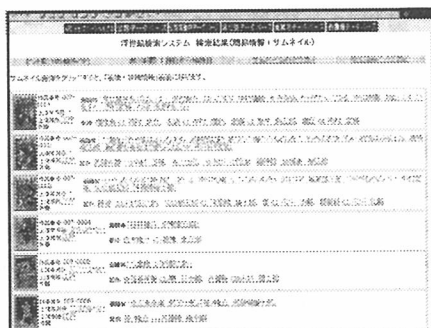
3,オリジナルサイズ表示



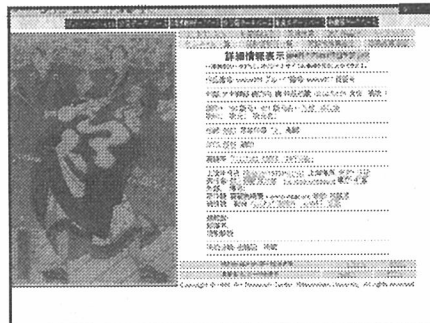
4,関連図表示（組合せ表示）



2.1,検索結果からリンク検索



2.2 詳細解説表示



2.3 解説編集画面

