

利用者登録型論文データベースシステムの構築

和田智仁¹

Implementation of a Theses Database System

Tomohito WADA

Abstract

This paper describes a method of implementing a database system for student's theses written in our institute.

One of the problem that arises at developing and running this kind of database system is the cost of collecting data. This system provides a mechanism for registering thesis data by users itself in order to share the cost of collecting data. All of thesis data registered by users are inspected by a professor who is in charge of the research to prevent registering inaccurate data.

The database system has build with 3-tiered client-server structure. The server consists of database management system(DBMS) and WWW server. WWW server accepts all requests from clients, and deals with DBMS to process the requests. Thus clients can access the database system using WWW browsers regardless of its platform.

KEY WORDS: データベース, 3 階層クライアント・サーバシステム, WWW

1 はじめに

スポーツ情報センターでは、学内に分散して存在する様々な研究情報を収納し、教官および学生の双方がそれらの研究情報を相互に利用できるようなマルチメディアデータベースシステムの構築を計画している。本研究では、目標とするマルチメディアデータベースシステム構築の足掛かりとして主にテキスト形式のデータを収納する論文情報データベースシステムの開発を行った。特に今回は、学生の作成する卒業研究に関する論文（卒業論文）および修士論文を対象としたデータベ

スシステムを構築することとした。これらの論文を対象としシステムを構築したのは、現在までにこれらの論文情報を一元的に収集し提供するシステムが本学には存在していないこと、論文の形式が定まっておりデータベース化が容易であること、さらに学生にとってこれらの情報は所属する研究室を選択する際あるいは配属後の先行研究調査の際に有用であることなどが理由である。

本データベースシステムを構築するにあたって予想された問題点の一つは、論文情報の収集コストである。本学は学士修士併せて毎年180名程度、また開学後これまでに延べ2,200名以上を排出し

¹ 鹿屋体育大学 National Institute of Fitness and Sports in Kanoya

ており、卒業論文および修士論文もこれと同じ数だけ存在する。毎年数多く作成される論文に関する情報を随時収集するとともに、これまでに作成された紙媒体の論文の情報を電子的フォーマットに変換しデータベースに格納するには非常に手間がかかるため、これをスポーツ情報センター独自で行うことは難しい。

そこで、本研究では利用者自身でデータベースへと論文情報の登録を行うことが可能である論文データベースシステムを構築することとした。利用者自身によって情報をデータベースへと登録できる場合には、データベースの管理者はデータの収集作業を行う必要がなく、システムそのものの管理のみを行えばよい。

本論文では、今回作成した利用者登録型の論文情報データベースシステムの設計から実装の方法について述べる。

2 システムの設計

構築する論文情報データベースには、すでに存在する大量の論文情報を格納するとともに、毎年作成される論文情報も随時収録していく必要があり、これらの情報収集コストがデータベースの構築・維持の際の問題となる。一方、データベースの利用を促進するためには、利用者が手軽にデータベースを利用できるような仕組みを実現する必要がある。これらの点を考慮し、構築する論文情報データベースの要求要件を以下のように設定した。

1. 論文情報の収録作業を利用者間で分担できるように、データベースの利用者自身によって論文情報をデータベースへと登録できること。
2. 正当でない論文情報がデータベースへ登録されることのないような仕組みを有すること。
3. 登録された論文情報を簡潔な方法で検索する手段を有し、検索結果を個別に提示できること。
4. 特定のプラットフォームに依存しない汎用的な方法によりデータベースを利用できること。

要件1は、データ入力作業を分担するための要件である。利用者が分担してデータの入力作業を

行えば、データベースの作成者や管理者がデータの収集作業そのものを行う必要がない。

今回は論文情報の登録作業を行う利用者として論文の執筆者である学生を想定しシステムを構築することとした。現在では、ほぼ全ての学生がパーソナルコンピュータを使用して論文を作成しており電子的なフォーマットで論文を所有していることから、論文の執筆者本人による論文情報の登録作業には手間を要しないだろうと考えたためである。また、研究室配属後の学生は、先行研究調査のために過去の論文を入手し検討する 경우가多く、このような場合にデータベースへの登録作業を併せて行ってもらえば、既存の論文に関する情報も分担して収集できるのではないかと考えた。

ただし、不特定の利用者からも論文情報の登録を受け付ける場合、入力ミスや不正確な情報などを含む、不正なデータの登録を排除する仕組みが必要となる。そこで、要件2では不正な情報を排除するための仕組みの必要性を挙げた。

今回構築したシステムでは、新規に登録される論文情報は登録者以外によって内容の確認作業を行い、正当な情報であると確認された情報だけをデータベースへと登録することとした。また、登録される論文情報の確認作業は論文の執筆を指導した教官に行ってもらうこととした。指導教官は登録される論文の内容を把握しており、また登録される論文情報には指導教官名も含まれていることから、指導教官に確認を依頼することが最適と考えた。

また、大量のデータを有するデータベースを有効に利用するためには、大量のデータの中から欲しい情報のみを抽出して入手するための仕組みが必要である。

要件3では、簡潔なデータベース検索手段を提供する必要性を挙げた。

本システムでは、キーワードによる簡易検索と、発表年度や所属講座等を指定しての条件検索手段を提供することで、登録情報の検索、検索結果の閲覧を行えるような機構を実現することとした。

要件4は、構築するデータベースを広く利用してもらうために必須の要件である。

今回のシステムでは WWW サーバとデータベースとを連携させ、データベースの利用に関する全ての操作を Web ブラウザから行えることとした。

3 実装

3.1 システム構成

前節の要求要件を考慮し、本研究では 3 階層構造のクライアント・サーバシステムによりデータベースシステムを構築することとした (図 1)。本システムでは、データベースサーバとクライアントの間にアプリケーションサーバを配し 3 階層構造とした。このような 3 階層のクライアント・サーバ方式は従来の 2 階層方式に対し柔軟性が高く、最近では多くのシステムでも採用されている [1, 2, 3]。3 階層方式ではクライアントからデータベースが完全に分離できるため、データベースにおけるテーブル構造の変更、あるいはデータベースサーバの二重化といった構造変更などのデータベースの変更の影響をクライアント側から隠すことが可能である。本システムにおいては特にデータベース側での将来的な拡張が予想されることから、この方式を採用した。

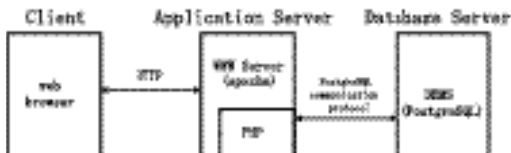


図1 1: 3階層システム

3 階層方式では、クライアント側で専用インタフェースを使用し、クライアントとアプリケーションサーバ間での通信を CORBA や Java RMI などのミドルウェアを用いる場合もある。これらを用いればクライアント上での複雑な計算やサーバとのより複雑な連携などが可能となる。本システムではクライアント上で特に計算を必要としないこと、クライアント・サーバ間の通信が複雑でないこと、クライアントインタフェースは簡潔でよいことなどを考慮して、通信プロトコルおよびクライアントインタフェースにそれぞれ HTTP と Web ブラウザを使うこととした。これにより、

多くのプラットフォームにおいて標準的に使用可能である Web ブラウザを使用して、利用者がデータベースにアクセスすることが可能となる。

アプリケーションサーバでは、HTTP を通じたクライアントからの要求を受け、要求内容を元にデータベースサーバと通信を行い、その結果をクライアントに返さなければならない。このような機能を実現する方法としては、WWW サーバと CGI (Common Gateway Interface) やサーバサイドスクリプト、または Java Servlet といった機構を組み合わせる方法が一般的である。本システムではサーバサイドスクリプトによってシステムを構築することとした。サーバサイドスクリプトは、HTML ファイルの中にプログラム自体を埋め込む形式のもので、CGI に比べ、プロセスを生成しないため処理が軽い、HTML ファイルとプログラムとの関係が理解しやすい、といった特徴がある。最近では Microsoft の ASP (Active Server Pages) や PHP などが広く使用されている。

本システムでは、データベースサーバ上の DBMS (Database Management System) として PostgreSQL 6.5.3 を [4]、アプリケーションサーバ上の WWW サーバとして Apache 1.3.20 を [5]、サーバサイドスクリプトには PHP 3.0.12 を [6] 使用している。これらを採用した理由として、全て無償で利用可能であること、いずれのソフトウェアも広く使用されており互いに連携させ使用する場合の実績や関連情報も豊富にあること、特に PHP は PostgreSQL などの各種 DBMS を利用するための豊富なインタフェースを備えており本システムのような Web-DB 連携システムの構築に向いていること、などが挙げられる。今回は計算機資源の都合から WWW サーバと DBMS を同一マシン上で動作させることとしたが、構造としては 3 層であることにはかわりはない。サーバには、Fujitsu GP7000S model5 (SunOS5.7) を使用した。

3.2 データ登録

本節では、実装したシステムにおける論文情報データ登録時の手順およびその仕組みについて説明する。論文情報の登録は、図 2 のように “仮登

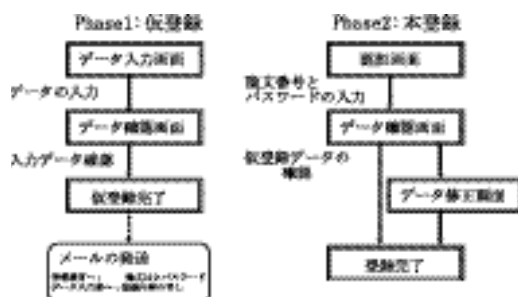


図2 仮登録と本登録

録”と“本登録”の2段階の手順によって行うこととした。以下、それぞれの手順について説明する。

1. 仮登録

仮登録は、種として紙の媒体で存在する論文の情報を電子的フォーマットにてデータベースへ登録する作業段階である。

図3は、論文情報の仮登録用フォームをWebブラウザ上に表示したものである。今回の実装ではデータベースに登録可能な情報として、著者名、学籍番号、所属講座、論文種別、発表年度、指導教官、論文題目、論文概要、論文キーワードおよび備考欄を用意した。本フォームを用いて適当なデータを入力あるいは選択すれば、論文情報の仮登録が行える。現在のところ、このページにはデータベースにアクセスできる全ての利用者がアクセス可能であり、全ての利用者は論文情報の仮登録を行うことができる。また、データの仮登録時には、登録者名と登録者メールアドレスも一時的にデータベースへ登録される。ただし、登録者メールアドレスは、本学ドメインのアドレスのみが許可されている。

本フォームを用いて論文情報がシステムに送信されると、システムは受信した各種情報の検査および修正を行う。具体的には、文字種 (1,2byte

文字)の統一、空白文字の除去、機種依存文字の検出と警告、学籍番号およびメールアドレスの検査、先に登録されている論文情報との重複検出などである。これらの修正が終わり、全ての検査に合格すると仮登録が完了する。

仮登録が完了後には、論文の指導教官宛てに登録内容確認依頼の電子メールが自動的に発送される。リスト1にメールの一例を示す。メールには、論文情報の仮登録が行われた旨の説明、本登録のためのURLおよび教官認証用の論文番号とパスワード、登録された論文情報の一部(タイトルと著者名など)が含まれている。また、論文情報の登録者メールアドレスにも、確認のため登録内容

図3 論文情報の仮登録フォーム

リスト1: 指導教官へのメール

*****様

論文情報データベース(<http://thesisdb.nifs-k.ac.jp/>)に、論文データの仮登録がありました。データを本登録するには、指導教官のデータ確認作業が必要となっています。以下のページにアクセスし、データの確認と本登録作業をお願いします。

『卒業研究・修士論文データベース登録確認ページ』
<http://thesisdb.nifs-k.ac.jp/commit.php3>

論文番号:	1234
認証パスワード:	*****
論文題目:	論文データベースの構築に関する研究
著者:	山田太郎 (012345)
発表年:	2001
データ仮登録者:	山田太郎 (s012345@sky.nifs-k.ac.jp)

に関するメールが自動的に発送される。

■ 本登録

本登録は、仮登録された論文情報の内容を確認し、確認後の情報を正式にデータベースへと登録する作業段階である。

本システムでは、主として論文の指導教官が仮登録終了後に送付される電子メールの通知を受け、本登録用の Web ページにアクセスし、論文情報の内容を確認しデータベースへと本登録することを想定している。図 4 に、本登録のための利用者認証画面を示す。仮登録後に指導教官宛てに送付される電子メールにある論文番号とパスワードを用いて認証は行われる。



図 4 認証画面

認証が成功すると、図 5 に示すような論文情報の内容確認画面が表示される。内容確認画面では仮登録された論文情報が表示され、指導教官はその論文情報をデータベースへと本登録できる²。また、必要に応じてデータの再編集、あるいは消去も選択できる。

指導教官によって本登録された論文情報は、即時にデータベースへと登録されるとともに、データ検索の対象となる。

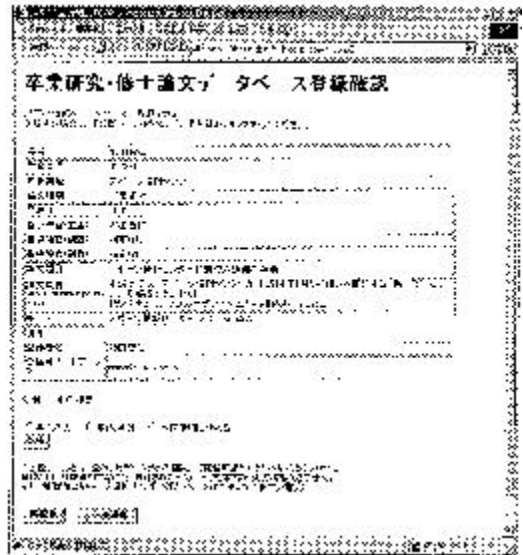


図 5 本登録フォーム

■ データ検索

データの検索には、簡単に検索を行うための簡易検索と、詳細な条件設定から論文を検索できる詳細検索の 2 種類を用意した。

簡易検索用画面を図 6 に示す。簡易検索では、検索キーワード³を指定して論文を検索できる。指定された検索キーワードは、論文情報中の題目、要旨、キーワードと比較される。



図 6 簡易検索

² 今回の実装では、論文の本登録時にその論文に関する情報をどの範囲まで公開可能とするかの「公開レベル」を指定することとしたが、現時点で公開範囲の制限については実装されていない。

³ 「検索キーワード」は、データベースを語句によって検索するための文字列のことで、論文自体の内容を示すための「キーワード」とは異なることに注意。

詳細検索用画面を図7に示す。詳細検索では、論文種別、発表年、所属講座、指導教官といった項目をリストから選択したり、著者名、指導教官名、論文題目、要旨、キーワードなどの語句を指定することで複雑な検索条件から検索を行うことが可能である。

図7 詳細検索

検索結果は図8に示すように、論文の題目、著者名、論文種別、所属講座、発表年、要旨の一部の一覧として表示される。論文の題目は、より詳細な論文情報へのリンクとなっており、このリンクを選択することにより、個別の論文情報を表示できる。検索結果の詳細を表示した画面を、図9に示す。

図8 検索結果

図9 検索結果詳細表示

4 考 察

本システムは、2000年12月から現在まで一部の教官の協力をいただいてテスト運用を行っている。現在までには特に大きな問題は見付かっていない。ただし、幾つかの改良すべき点も見付かっている。

ひとつは、仮登録時の利用者認証の問題である。今回のシステムでは、論文の仮登録時には特に利用者認証を行わず、任意の利用者が情報を仮登録できるようにしている。今回のシステムでは、当初は学内向けにのみ開放する予定であること、システムの実現が容易であること、といった理由からこのような簡易な方法を用いた。しかし、今後システムを学内外へ開放することを考慮すると、仮登録時にも利用者の認証を行い、正当な利用者(例えば本学の学生)のみが仮登録を行えるなどとし、関係者以外のいたずらなどによる不正な情報の登録を排除する必要があるだろう。利用者の認証方法としては、データベース専用の認証システムを設ける、本学の利用者認証システムと連携して行う、論文情報の本登録と同様に電子メールを用いて本学のメールアカウント宛てに登録用の認証コードを送付する、などが考えられるが、使いやすさの点から本登録と同様に電子メールを用いた方式を採用したいと考えている。

もう一つは、論文情報の修正機構の不備である。また、システムのテスト運用開始後には「論文情報のうち、まず題目と著者名だけを登録しておいて、順次要旨などの情報を追加していきたい」という意見もあった。このような利用法を可能にするためにも論文情報の修正機構は必要である。論文情報の修正については、登録済のデータを元に加筆・修正を加えたデータを再度仮登録できるようにすれば、現在の手順と仕組みをほとんどそのまま使用して対応できる。

また、今回実装したデータベースでは、著者名や論文要旨などのみが収録の対象で研究論文そのものを収録の対象としていない。今後は、論文そのものもデータベースへと収録できるようにデータベースを拡張したいと考えている。ただし、既存の紙媒体の論文をデータベースへ電子的フォーマットで収録するのには非常に手間がかかる。特に、本文そのものも全文検索の対象とするには、論文の内容も個別に収録しなければならないが、現実的にそれを行うのは非常に手間がかかり困難であろう。先に述べたように、学生は論文を電子的な手段によって作成する場合がほとんどであり、新規に作成される論文を学生から電子的フォーマットのまま直接収集しデータベース化すれば電子化のコストを最小限に留めることが可能である。本文を含む論文のデータベース化を進めるためには、電子的フォーマットのままで論文を回収していくための全学的な制度作りが必要であろう。

5 おわりに

本論文では、登録する情報の収集コストを分散するために利用者自身で情報を登録可能な論文情報データベースシステムの構築について述べた。

本システムは、多くの利用者から利用してもらうために WWW とデータベースとを連携させた。これによって、プラットフォームに依存しない幅広い利用者からのアクセスが可能となった。

本システムにより、本学に存在する数多くの卒業論文および修士論文、さらに今後作成される論文の情報をデータベース化することが可能となった。今後は、本システムを全学的に利用してもら

えるように活動を行うとともに、収録データ数を増加させ、より価値のあるデータベースへと発展させていきたい。

また、本学における研究の特徴としては、研究に使用する資料やその成果にビデオ映像が用いられていることが挙げられる。今後は、これらの関連する情報も併せてデータベースへと収録できるような、マルチメディアデータベースシステムの作成に取り組みたい。

謝 辞

本論文は、平成11年度教育改善推進費(学長裁量経費)による「スポーツ情報発信用データベースシステムの構築」プロジェクトの研究成果の一部である。プロジェクトの進行にあたり、スポーツ情報センターおよび運営委員の各教官にご協力いただきました。特に、論文情報の提供には志村教授、萩助教授、杉山助教授、松尾助教授にご協力いただきました。ここに感謝の意を表します。

参考文献

- [1] 安藤英俊．WWW を利用した 3 層構造型履修申告システムの開発，学術情報処理研究，NO.3，pp 21-31，1999．
- [2] 島田圭，中所武司．フレームワークによる 3 層 Web アプリケーション構築法，情報処理学会研究報告2000-SE-129，pp 1-8，2000．
- [3] 鶴野和応，永田幸次．3 階層 C/S システムで構築した PerfectTV! 顧客管理システム(SMS)，情報処理，Vol.39 NO.2，pp 152-157，1998．
- [4] PostgreSQL. <http://www.postgresql.org/>
- [5] The Apache Software Foundation. <http://www.apache.org/>
- [6] PHP: Hypertext Preprocessor. <http://www.php.net/>

