

徳島大学 教育・研究者情報データベースの開発と運用

Development and operation of education and research database in Tokushima University

大家 隆弘* 上田 哲史† 矢野 米雄‡
Takahiro OIE Tetsushi UETA Yoneo YANO

キーワード: データベース, 教育情報, 研究者情報

Keywords: Database, Educational Information, Researcher's Information

1. はじめに

大学においては、教育・研究活動を広報する目的で、紀要、教育要覧、研究要覧などの定期刊行物（以下、刊行物）を出版している。刊行物に収録することにより読者に効果的にアピールできる項目としては、構成員の業績リストが挙げられる。一方で、大学の構成員に対する業績調査も頻繁に行われており、教職員はその都度調書の作成を行っている。

これの刊行物や調書を作成するためのデータは、例えば学歴、職歴のように職員個人にとっては不変の情報であったり、著者リストのように既存のデータを追加するだけで済む情報である。しかし、個人のデータの記録書式や、編集者のデータ収集方法を統一することは困難であり、よって、刊行物ごとにデータの再入力強いられ、結果、編集・校正に長い期間を要し、データ入力者のみならず編集者にとっても多大な負担となる。

そこで本学では、上記刊行物の網羅すべき項目を洗いだし、「情報の再利用」を設計思想として、ユーザの入力と刊行物の編集作業の援用を指向した「教育・研究者情報データベース」(EDB; Education and research database, 以下、EDB) を構築した。本稿ではこの EDB の開発と運用について述べる。

2. EDB の設計

EDB の設計において、以下の事柄に重点をおいた。

2.1 情報識別の完全性

データベースに登録されている情報（以下、登録情報）を分類する際に、情報を正しく分類することが必要である。

通常、教職員は業績情報を（ワードプロセッサや表計算ソフトなどを利用して）小規模なデータベースを作成して管理している。それらの情報は、その個人が所有するという点で特定個人の業績情報と判断できる。しかし、複数の利用者が登録を行うデータベースで、単純にそれらの情報を集めると情報の重複が発生するため、組織毎の分類では業績の重複が生じてしまう。

そのため、EDB では著作などの情報を特定の個人から

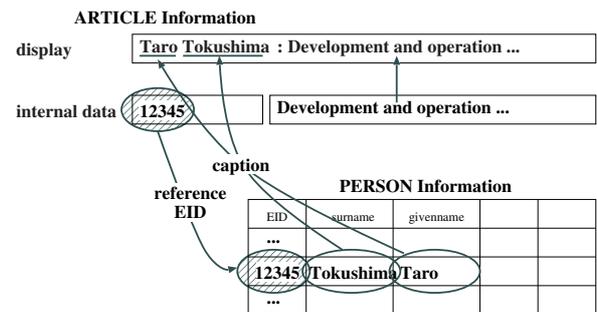


図 1. 参照型記述方式

切り離して管理し、「現実世界において 1 つの事柄を登録情報 1 つに対応させる」ことを大原則とした。この大原則のもとでは、複数著者の著作においても登録情報は 1 つであり、複数の著者がその情報を共有する。さて、この原則においては著者の 1 人が情報登録を行えばよいが、他の著者がその著書について既登録かどうかを知るための明確な方法が必要になる。

しかし、単純な姓名テキストによる分類では正確に個人を特定できない。例えば論文著者による分類を行う場合、人名には同姓同名、新旧字の問題があるためテキストによる分類には信頼性がない。そこで、他の情報を参照することによる情報記述方式（以下、参照型記述方式）を採用する。すなわち個人情報を含む全ての情報に識別子 (EID) を割り当て、著作の著者には個人情報の EID を記述する。識別子 EID の実体は整数値である。ただし、EID は利用者には判読が困難なため、閲覧インタフェースでは EID はそれが指し示す情報の見出しに置き換わることをとする。例えば著者欄に記述された EID は対応する個人情報の姓名テキストを表示する（図 1 参照）。

もちろん、著者欄にはテキストを記述できるが、EDB においては「参照型記述方式を用いた場合のみ、業績分類の対象となる」ということを大前提とした。テキストで姓名を記述した場合には、如何なる場合においても個人の業績と分類されることはない。

以上のように、業績分類を簡潔な規則の上で行うことで、利用者は登録情報中の自分の業績を正確かつ素早く抽出することが可能になる。また、この他に情報登録時

*徳島大学 工学部 電気電子工学科

†徳島大学 高度情報化基盤センター

‡徳島大学 工学部 知能情報工学科

に関係者（共著者など）に電子メールで通知する等の方法で、重複情報の排除に勤める。

2.2 データ構造

本データベースが目的とする情報公開は日本語文化圏に限定されない。よって、各登録項目を英語¹、日本語²、日本語読み、参照型記述方式の EID のセットとした。EID が無効なとき、英語、日本語、日本語読みが有効となる。

各情報に属する項目の定義は、一般の RDBMS (Relational Database Management System) と同様の手法で、情報の正規化に注意しながら設計を行った。

2.3 プラットフォーム非依存

様々な利用者が属する環境では、特定の OS やアプリケーションに依存するシステムを構築すべきではない。

特定ベンダへの依存性をなくしかつデータ様式の加工や変更に対応するため、登録情報の表現には XML (Extensible Markup Language) [1] を採用した。XML は文字コードとして Unicode [2] を採用しており、多種の言語を表現でき有用である³。

ただし、大多数の利用者には XML での情報記述は困難を伴うため、情報登録は WWW の CGI で入力インタフェースを用意し、CGI において XML 表現に変換し登録する。また、情報抽出には XML 表現の他に、各アプリケーションに適した表現形式 (CSV, HTML, LaTeX[6] など) に加工・出力するインタフェースを準備した。

2.4 開放されたインタフェース

登録情報の利用を制限すると利用範囲が限定され、データベース利用の活性化に矛盾する。登録情報の中で公開可能な情報を抽出、公開用のデータベースを作成し、利用者が任意に登録情報を抽出可能な状態を実現する。また、利用者の習熟度にあわせ、データベースに直接アクセス可能な低レベルインタフェース、プログラミングインタフェース、また簡潔な条件式で情報を様々な形式で抽出可能なインタフェースを用意する。

2.5 セキュリティ

WWW によるデータベースへのアクセスは、HTTPS (Secure HTTP) により保護し、その他のアクセスのうち非公開情報にアクセスするものは SSL (Secure Socket Layer) による通信路暗号化を行う。非公開情報へのアクセスは、パスワードによるユーザ認証を経ることを条件とする。また、登録情報が「いつ」、「誰によって」、「どのように（登録）変更された」かを履歴情報として保存する。

一方、登録情報毎に読み、変更、無効化の権限を各情報に付加する。基本的に登録情報は登録者に変更、無効化の権限があるが、著作などの共著者にも権限を与える

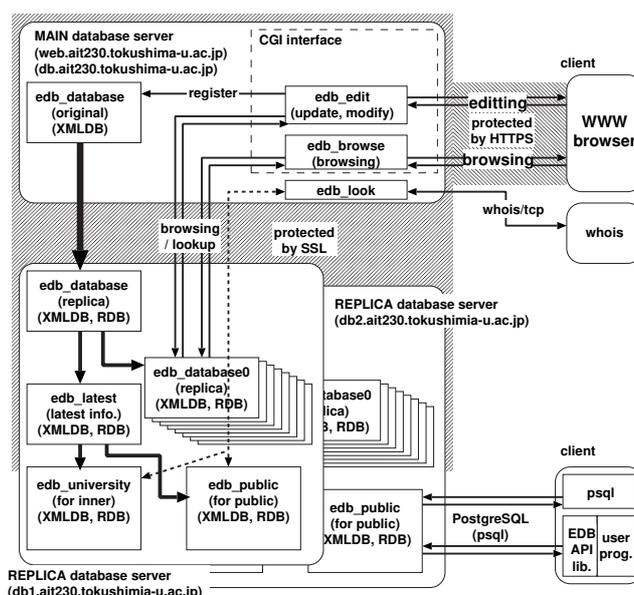


図 2. システム構成

ために、情報の定義自体に権限の譲与構造を設ける。また、データベース内での権限の譲与構造（たとえば、学科内の情報は学科長が修正、削除の権限を有するなど）を定義し、実社会の管理構造を反映する。

3. EDB の構築

3.1 システム構成

本システムの概要を図 2 に示す。本稿の執筆時点において、EDB は 3 台のサーバで構成されている。OS として FreeBSD [3] を採用し、データベースコアに PostgreSQL 7.2[4]、WWW サーバに Apache 1.3[5] を用いた。

1 台がメインとなり原典情報をデータベース XML 形式で保持する (XMLDB)。あとの 2 台は複製 (replica) を保持するほか、XML 形式の登録情報を RDB 形式に変換したデータ (RDB) を保持し、検索を請け負う。edb_database0 に連なるレプリカが数多くあるのは、アクセス混雑時の接続数の分散を行うためである。更に、この 2 台のサーバは登録情報中の最新情報の抽出 (edb_latest; 履歴情報を含まない) の作成、公開可能な情報を抽出、学内公開用データベース (edb_university) および学外公開用データベース (edb_public) の XMLDB, RDB を作成する。以上の作業は常駐プログラム (edb_dwarf) がバックグラウンドで行う (図中、黒太線)。

情報登録は全て CGI (edb_edit) を経由して WWW を介して行う。また、閲覧専用の CGI として edb_browse が用意されている。whois/tcp ポートを利用したアクセスは、原則として公開用データベースにアクセスし、検索結果を指定された様式で出力する。その他、レプリカサーバは PostgreSQL のインタフェースを開放し、psql (PostgreSQL のユーザフロントエンド) や、EDB で用意した API ライ

¹厳密には、Latin 語文化圏のための表記。

²中文、ハングル語なども含む。

³機種依存コードエリアは利用しない。

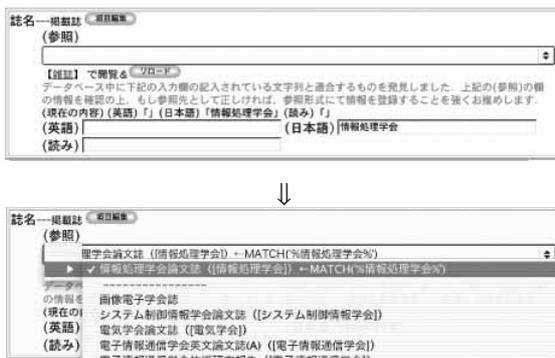


図 3. 既登録のデータをシステムが発見，プルダウンメニューにその雑誌名を呈示したところ。

ブラリにて公開用データベースにアクセスできる。

3.2 情報登録インタフェース

本システムでは参照型記述方式により情報記述の不確実性を排除したが，情報記述を識別子 EID により行わなくてはならない。一般に，EID による情報記述は利用者には受け入れがたく，また，誤入力の危険を伴う。本システムでは，EID ではなく参照先の情報リストを利用者に提示，選択を促す方法を採用した。例えば，著作情報の入力時に，参照実績のある個人情報リストを「個人名(所属)」の形式で表示し，選択を促す。また，テキスト入力に対してはテキストの適合情報を検索し，適合情報が存在する際にはその旨を利用者に通知，参照型記述方式での登録を促す。

図 3 は，雑誌名の欄に入力されたテキストに適合する雑誌の情報を発見したときの表示（実際の表示は注意を促す赤色）であり，下はプルダウンメニューにリストされた雑誌名である。

図 4 は，ある著作データ（図 5）の登録情報の閲覧画面である。XML 形式において属性 mapto="" の記述箇所が，参照型記述方式の該当部分である。これらは，図 4 において下線で示される箇所に対応している。また，ブラウザ上では参照データの閲覧ページへのリンクとなっており，参照情報を素早く確認できる。

4. EDB からの登録情報抽出モデル

本節では，EDB のシステムが提供するいくつかの登録情報抽出モデルについて述べる。

4.1 登録情報抽出インタフェース

現在，登録情報抽出インタフェースとして，低レベルの (1) PostgreSQL への直接接続と高レベルの (2) EDB ライブラリによる API (C 言語)，(3) whois/tcp (43 番) ポートを用いた接続が用意されている。

データの出力形式も，(1) では PostgreSQL(RDB) 形式と XML 形式，(2)，(3) では，XML 形式，参照型記述を補間した XML 形式， \LaTeX ($T_{\text{E}}\text{X}$) 形式，HTML 形式，CSV



図 4. 著作情報の表示。

形式，プレーンテキスト形式などを提供している。

4.2 WWW ページの作成

登録情報を利用して利用者自身が WWW ページを作成する場合，図 6 のモデルを利用する。業績は whois/tcp インタフェースへのアクセスによって，HTML 形式の定型のリストを得ることができる。利用者は HTML ヘッダや前書き，後書き等を補間し頁全体を容易に作成できる。基本的なスクリプトは既に提供されている。

4.3 各種刊行物の作成

刊行物の内容をデータベースから作成可能なものと不可能なものに分離し，図 7 のモデルによって処理を行う。

まず，刊行物の作成に必要な情報を各利用者 (objective person) が登録する。その後，プログラム edb.make (web A) が冊子に合わせた書式で \LaTeX 形式の出力ファイルを作成する。このファイルと，編集者が用意した冊子固有の \LaTeX ファイル群 (web B 上の *.cls, *.sty, *.tex) を混合し， \LaTeX 処理を施し PDF ファイルを作成する。

\LaTeX により作成されるドキュメントは，オフセット印刷でも遜色のない品質を達成でき，作成された PDF を最終出力イメージ (プレプリント) として利用者に提示できる。利用者はプレプリントをダウンロードし，各自の該当部分を確認しながら登録情報の編集作業を行う。

本モデルにおいて，プレプリントの作成作業は完全に自動化が可能であり，また情報の登録からプレプリントの作成までの時間を従来の手作業の編集作業に比べて飛躍的に短縮することができる (後述のシラバスでは，1 時間毎にプレプリントを更新した)。結果として，冊子作成に要する時間を短縮し，迅速な情報公開が可能である。

工学部研究報告 (平成 13~15 年度)⁴，学術研究要覧 (平成 13 年度)⁵，シラバス (平成 13~15 年度)⁶ がこの

⁴http://web.e.tokushima-u.ac.jp/book/bulletin200{1|2|3}.final/

⁵http://web.e.tokushima-u.ac.jp/book/ars4.final/

⁶http://web.e.tokushima-u.ac.jp/book/syllabus200{1|2|3}.final/

```

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE edb:article SYSTEM "http://web.ait230.tokushima-u.ac.jp/
dtds/article.dtd">
<edb:article xmlns:edb="http://web.ait230.tokushima-u.ac.jp/dtds/">
  <edb:base eid="20540" eoid="20579" mapto="0" mtime="1001911673"
operator="10729" avail="true" censor="0" owner="10729" read="inherit"
write="inherit" delete="inherit"/>
  <edb:article.kind mapto="10446"/>
  <edb:article.organization mapto="11119"/>
  <edb:article.organization mapto="11131"/>
  <edb:article.author mapto="10729"/>
  <edb:article.author mapto="10763"/>
  <edb:article.author mapto="10746"/>
  <edb:article.author mapto="10755"/>
  <edb:article.author mapto="10723"/>
  <edb:article.title>
    <edb:english>A Database Orienting Information Disclosure on Campus
Network</edb:english>
  </edb:article.title>
  <edb:article.publisher>
    <edb:english>I THET2001</edb:english>
  </edb:article.publisher>
  <edb:article.magazine>
    <edb:english>Proceedings of 2nd International Conference on Information
Technology Based Higher Education and Training</edb:english>
  </edb:article.magazine>
  <edb:article.city>
    <edb:english>Kumamoto</edb:english>
  <edb:japanese> 熊本市 </edb:japanese>
  <edb:pronounce> くまもとし </edb:pronounce>
  </edb:article.city>
  <edb:article.date>
    <edb:english>20010706</edb:english>
  </edb:article.date>
</edb:article>

```

図 5. 著作情報の XML 表現 .

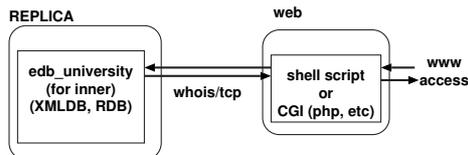


図 6. WWW 頁の作成モデル

方法で作成された .

4.4 各種調書の作成

大学においては、様々な調書の作成が要求される . 一般的には、これらの調書は個人の業績を調べるものが多いが、調査内容は似通っているものの書式は調査毎に異なっているのが通例である . しかし、データベース管理者が、調書にあわせたプログラム作成、調書のデザインなどを微調整することは負荷軽減という趣旨に反している .

本システムでは、図 8 に示すモデルを考え、汎用の登録フォーム作成プログラム (edb_form) を準備した . このプログラムは外部サーバ (web B) からフォーム定義 (form.def) やその他のファイル (person.list, staff.list, *.opt) を取得、web A 上で登録フォームを作成する . form.def の内容は WWW 頁であるが、フォーム入力欄や登録情報の検索結果表示等を内包できる . 本モデルではスタッフ (coordinator) が form.def の調整を行う . 調書の対象となる利用者は edb_form を利用して各自のデータを登録する . 登録されたデータは web A 内に蓄積され、定期的にスタッフがそれらを読みだし \LaTeX 処理などを施し最終調書のプリントアウトを作成する . 書式に合わせた調整は \LaTeX の

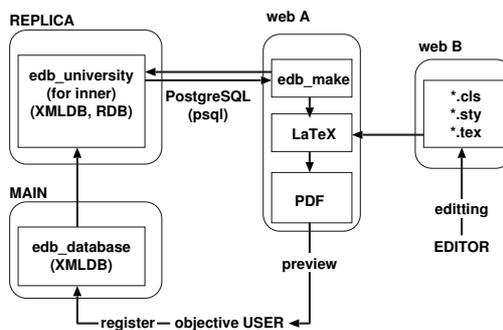


図 7. 刊行物編集モデル

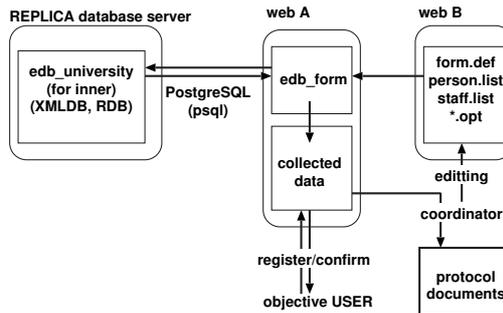


図 8. 調書作成モデル

スタイルファイルで行う .

このモデルを用いて、本学工学部は平成 13 年度着手の大学評価・分野別研究評価「工学系」の個人別研究活判定票の生成を行ったほか、現在、工学部内のいくつかの学科で進行中の JABEE 審査用の調書 (教員個人データ) の作成を行っている .

本モデルで作成された調書は、同じ処理を経て作成されるため統一感があり、また \LaTeX を利用していることから見た目も好評である . また、利用者によるデータ登録を除いて処理を自動化でき、一度システムが確立してしまえば、同様の調書作成にも応用できる .

5. むすび

本稿では徳島大学における教育・研究者情報データベースのコンセプトと構築、運用について述べた . 今後は登録情報の抽出、加工に関するいろいろなユーティリティの拡充、提供等を行っていく予定である .

なお、EDB のシステムのより詳細な情報は [7] にて公開しているの、そちらを参照して頂きたい .

参考文献

- [1] XML, <http://www.xml.org/>
- [2] Unicode, <http://www.unicode.org/>
- [3] FreeBSD, <http://www.freebsd.org/>
- [4] PostgreSQL, <http://www.postgresql.org/>
- [5] Apache, <http://www.apache.org/>
- [6] Laslie Lamport: "A Docuemnt Preparation System \LaTeX ," Addison-Wesley, 1986.
- [7] EDB, <http://web.ait230.tokushima-u.ac.jp/>