

XDES - 多様な構造と流動的変化に対応できるデータエントリーシステムの構築

XDES - A Heterogeneous Extensible Data Entry System

杉本 典子[†] 金丸 玲子[†] 関 隆宏[‡] 石野 明[†] 竹田 正幸^{*} 廣川 佐千男[†]
 Sugimoto Noriko Kanamaru Reiko Seki Takahiro Ishino Akira Takeda Masayuki Hirokawa Sachio

1. はじめに

XML は、そのデータとしての扱いやすさと表現の柔軟性から、データ処理技術の理論と応用の両面で新たな発展の基盤となることが期待され、研究・開発が国内外で活発に行なわれている。しかし、現状では、XML が潜在的にもつ有益な能力が十分に活用されているとは言い難い。XML の柔軟な表現力を有効活用するためには、**多様で流動的なデータスキーム**に対応できるシステムの構築が必須である。

近年、各大学で必要性が高まっている大学評価システムは、流動的なデータスキームが必要となるデータベースの典型といえる。大学評価システムでは、大学戦略に従って学内教員や組織に関するデータの蓄積・公開を行なうとともに、学外組織や社会からの要求に従って情報の提出・公開が行われる。このようなシステムでは、扱うべきデータ項目の追加や再構成の作業が常時必要とされる。システムで扱うデータ形式もそれに対応して変化し続けなければならない。XML は、データ長やデータ形式の変更に柔軟に対応できる表現力があるので、このような性質のデータを扱うのに適している。

XML データベースを使うとしても、データを入力するユーザが何千人にもおよび、それぞれのユーザが共通の計算機環境を得ることが困難である場合、データをどのように収集するか、すなわち、**データエントリーシステム**が問題となる。従来の方法として、以下のものが挙げられる。

- ・データ入力対象者を階層化し、それぞれの層に取りまとめ担当者を配置することで分割統治により収集する。担当者はアンケート用紙や電子メールとして回収されるデータを、適切なフォーマットに編集してデータベースに入力する。
- ・データ入力対象者からアンケート用紙や電子メールとして回収されるデータを、データベース管理者、または業者委託等により適切なフォーマットに編集する。

これらの方法では、データの収集やデータベースへの入力を担当する者の負担が大きく、業者委託した場合は、そのために莫大な費用がかかるため、データ更新の頻度を年に1回とするなどの措置が必要となる。しかし、新鮮かつ正確な情報を提供するためには、変更が発生した後、即時対応することが望ましい。そのため、データベースシステムへのデータ入力、全ユーザがいつでも行えること、すなわち発生源入力が重要である。これについては、Web ペー

[†]九州大学 情報基盤センター Computing and Communications Center, Kyushu University
[‡]九州大学 大学評価情報室 Office for Information of University Evaluation, Kyushu University
^{*}九州大学 システム情報科学研究院 Graduate School of Information Science and Electrical Engineering, Kyushu University

ジにアクセスしてデータを入力する方法が広まっている。この方法では、パソコンの OS に依存せず、電子メール等で送信された際に発生する文字化けの問題も大部分は解消できる。また、Web ページのデータ入力フォームを利用することによって、データをすべて直接記述するのではなく、選択肢やチェックボックスによって、データを入力することが可能となるため、統一された正確なデータを収集することができる。しかし、データ項目それぞれに対して、データを編集する機能を提供する Web ページを作成するには莫大なコストがかかるという問題がある。さらに、運用開始後もデータスキーマの変更が発生するデータを扱う場合には、変更が発生するたびにデータ入力用 Web ページの変更も必要で、導入にも運用にもコストがかかるという問題がある。

本研究では、多様な構造と流動的なデータスキーマの変更に対応できるデータエントリーシステム XDES を開発した。本システムは、データスキーマの定義をマクロファイルという形式のテキストファイルとして管理し、CGI プログラムから分離した。スキーマ変更に伴う Web インターフェイスの修正作業は、CGI 等のプログラムの知識を持たなくても容易に行える。これにより、データ項目の追加や削除、入力形式の変更等の作業を短時間で行うことが可能となっている。データ項目の多様化と構造的変化が続いても、データエントリーシステムとしては、データスキーマ定義マクロを変更するだけよい。

本稿では、まず XDES の特徴と構成を述べる。次に XDES で実装した九州大学の大学評価システムにおける3年間の運用例を通じて得られたシステムの評価と今後の課題について述べる。

2. システムの特徴と構成

XDES の柔軟性、多様性、頑健性は以下のような設計の単純さによる。

- A) 個別データとユーザの区別
- B) データスキーマ記述のためのマクロファイル
- C) 属性項目に対する弱い型の導入
- D) データ群から構成される階層構造については明示的なデータとしては扱わず、個別データにおける複数の属性として表現したこと

本システム開発の契機は大学評価システムであったが、筆者らが所属する九州大学には以前から教員データベースシステム[6]が稼動していた。しかし、新たなデータ項目の導入と組織構造の変化への対応が困難であったため、蓄積されたデータは移行しつつ、システムの抜本的な再構築のための分析を行った。維持・管理そして新たな機能の開発が困難となっていた一番大きな理由が、教員の個別データだけでなく組織構成データという2種類のデータの取り扱いであった。このため、例えば学府・研究院制度という多重組織階層の導入や、部局間の人事異動などへの対応作業

の負荷が大きかった。この経験を踏まえ、上記A) D)の方針に従いできるだけ単純なシステムとした。また、データ項目決定のための1年以上に渡る委員会における検討を通じ、必要最小限のデータスキーマ記述機能が特定でき、それをマクロ言語として定式化したものがB)である。C)はデータエントリ開発中に行った施行入力時のトラブルから必要となったものであり、Web インターフェイス上で選択ボタンやチェックボックスとしてエントリ方法が実装されるデータ項目は、その項目がとりうる値を選択肢内のデータやチェックボックスによる返り値に制限し、日付等の数値データに対しては、それが指定された半角または全角の数字をとることを要求するものである。この機能は、主に、XDES を他のシステムと連携させる際、データの整合性を検証するために必要となる。

XDES で扱うことができるデータ形式は、

- A) 固定フィールドをもつデータ
- B) 固定フィールドからなるレコードの繰り返しからなるデータ

C) A)及びB)の合成からなるデータ

の3通りであるが、大学評価システムで実際に使われているデータ形式は、A)とB)のみである。XDES は、データの形式、データエントリ方法、表示方法、蓄積形式などを各項目毎に記述したデータスキーマファイルを参照し、各ユーザに Web ブラウザからデータを入力し、インターネットを通じてデータベースサーバにデータを蓄積するためのインターフェイスを提供するシステムである。XDES の構成を図1に示す。

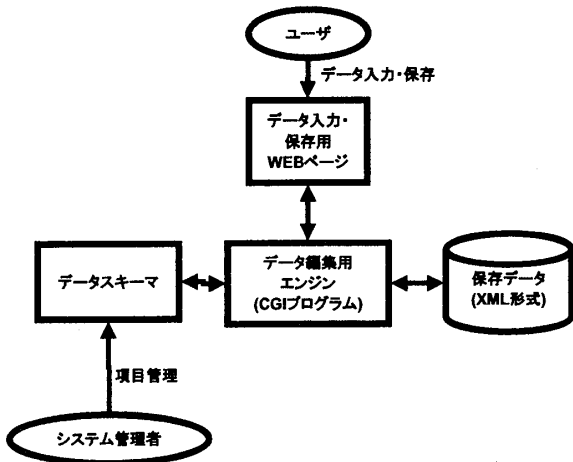


図1: XDES の構成

XDES の汎用性はデータスキーマ記述のためのマクロファイルによる。その例を図2に示す。データスキーマファイルは、50 から 200 行程度のテキストファイルであり、一般的なプログラムのように手続き的な内容を記述するものではなく、データ項目の ID、タイトル、Web 上での表示形式・入力形式等が宣言的にテキスト形式で記述された内部ファイルである。1つのデータ項目に対してファイルが1つ作成される。

インターネット上でユーザからデータの編集要求が発生すると、データ編集用エンジンは、このデータスキーマファイルを読み込み、ユーザに提供されているデータ項目編集画面にリンクするための目次を自動生成して Web ブラ

ウザ上に表示する。図3は、図2に示すデータマクロファイルからデータ編集機能の「項目追加」によって自動生成されたデータエントリ用 Web ページである。

```

title 原著論文:
tablehead 発行年 論文題目:
tabledata 14 0:
;
term 0 textarea 3 80 論文題目:
term 1 check 主要 主要:
term 2 NonNegativeInteger input 3 貢献度:
term 3 input 80 著者氏名 (全員):
term 4 input 80 著者氏名 (修士学生):
term 5 input 80 著者氏名 (博士学生):
term 6 input 80 著者氏名 (研究生/学部学生(*)):
term 7 select 日本語 英語 その他: 使用言語:
term 8 input 50 使用言語 (その他):
term 9 textarea 2 80 学会又は雑誌等名:
term 10 select 国際誌 国内誌 学内誌 論文形態 (国際誌等の区分):
term 11 select Journal_Proceedings 論文形態 (Journal等の区分):
term 12 select 招待論文 一般論文 論文形態 (招待/一般の区分):
term 13 select 査読あり 査読なし 論文形態 (査読):
term 14 NonNegativeInteger select &year 50 発行年:
term 15 NonNegativeInteger select &month 発行月:
term 18 textarea 5 80 その他 (巻、号、ページ等):
term 17 NonNegativeInteger input 5 被引用回数:
    
```

図2: データスキーママクロファイル

図3: Web ページ

本システムでは、データ項目の追加、削除、データフィールドの追加、削除、データ入力形式の変更及び、Web 表示形式の変更等は、すべてデータスキーマファイルの内容を変更することによって行う。データスキーマファイルの内容が変更されると、該当する項目の Web ページ表示イメージが変わると同時に、データ編集のために準備されているデータの追加、削除、コピーして追加、修正、ソート、一覧/詳細表示等の機能が変更後のデータ項目に対して提供される(図4)。

XDES が扱うデータ形式は、XML で扱うことが可能なデータ形式の一部にすぎないが、表形式で表現されるデータ形式は全て扱うことが可能であり、さらに、通常の表形式では扱うことができないレコード繰り返し入れ子形式等を扱うことができる。XDES で使われているスキーマ定義ファイルは、それ自体を XML で表現することも可能であり、XML スキーマで表現することも可能である。しかし、XML スキーマは、表現力があると同時に記述形式が難解であり、プログラミングの知識を持たない管理者を想定している本システムの記述様式には適さないと考えられる。

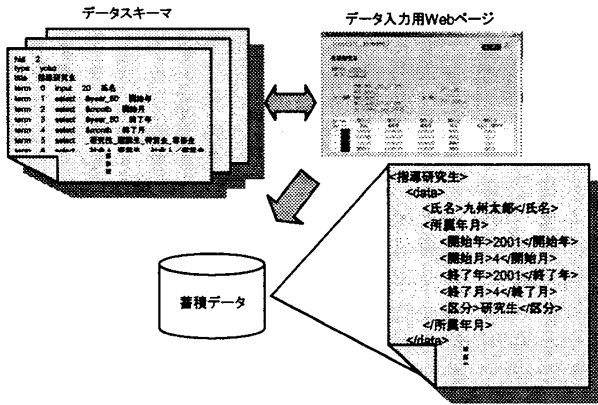


図 4: マクロ, 入力 Web ページ, 蓄積データ

3. XDES による大学評価システムの開発と運用

平成 16 年 4 月に行われた大学の独立行政法人への移行に伴って、各大学における社会への情報公開の重要性はますます高まっている。大学評価システムの開発は、大学評価活動に要する教職員の労力を軽減し、第三者評価への円滑な対応を行うために早急に対応すべき重要な課題となっている。九州大学では、大学評価システム開発の第一段階として、インターネットを使って各教員個人レベルでのデータを蓄積・公開するためのシステムを開発することとなった。本学で設定されたデータ項目は開発当初から数が多く、将来的にどのようなデータ分析に用いられるか未定のものが多い存在した。また、21 世紀 COE や大学評価・学位授与機構で運用が予定されている大学情報データベースで扱う項目が決定した時点で本システムの項目設定もそれに従って追加・変更されることが想定されていた。そのため、大学評価活動に有益なシステムを構築するためには、項目の種類や数、及びデータ長に制限がなく、それらに変更が生じた場合に迅速に対応できる柔軟性をもつデータベースシステムが必要であった。通常の関係データベースでは、前もってデータスキーマを特定しなければならないため、本システムには適していないと考えられる。RDB を使った既存の教員データベース[6]の修正では対応できないと判断し、柔軟な表現力のある XML データベースにより新規開発することとした。XDES の開発はこのような経緯で始まったが、2 章で述べたように大学評価に特化したものでなく、単純ではあるが汎用的な設計とした。

XDES の開発は、九州大学で平成 14 年 4 月開始され [1,2], 平成 15 年 11 月から全学的に運用を開始している大学評価システムのデータ入力システムとして現在稼働中である (図 5)。九州大学の大学評価システムで扱うデータは、大項目数が 60 (個人基礎情報関連項目数: 7, 教育関連項目数: 19, 研究関連項目数: 19, 社会連携活動関連項目数: 7, 学内運営関連項目数: 1, 競争資金関連項目数: 7), 1 教員が入力すべきデータの種類の総数が 753 と大規模なものである。入力されたデータは、データベースサーバに蓄積されるとともに、学外に公開することが定められている項目は、学外公開用 XML に自動的に変換され、「九州大学研究者情報」として、学外へ情報公開を行っている。

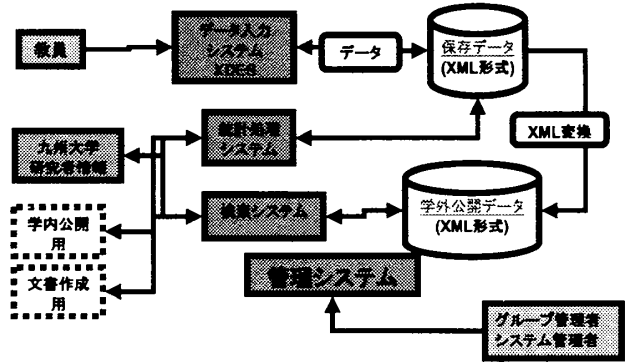


図 5: 大学評価システムの構成

システムの設計, 試験運用, 本稼働後のすべての過程においてデータスキーマは変更されている。しかし, XDES の変更作業は 1 人の管理者が最大でも 1 日, システムを停止させることによって約 2300 名の教員がそれぞれにデータを入力できる環境を提供している。

大学評価システムの設計が開始された平成 12 年度からの年度毎に発生した XDES に関するデータスキーマ変更件数を表 1 に示す。大学評価システムは、平成 12 年 12 月から設計が行われた。表 1 では、平成 13 年度の変更件数は 3 件となっているが、この段階ではほとんどの項目に対して全面的な修正を行っている段階であるので、具体的な修正箇所をカウントすることが困難であった。そのため、大掛かりな全面修正の数 3 件のみを記載している。平成 14 年からは、システムの本格的な構築を行うとともに、段階的な試験運用を行って、学内教員からシステムやデータ項目についての意見を収集し、それらに基づいてシステムを改善した。平成 15 年には全学運用が開始され、それに伴って学内教員や組織からの要望に答える形での修正が発生している。平成 16 年度には、検討中項目として保留していた社会連携活動、国際交流等に関する項目のデータ入力を開始し、それに伴って項目の全面修正したり、新規項目を追加したりすることが必要となった。

平成 13 年度	3
平成 14 年度	437
平成 15 年度	80
平成 16 年度	8
変更総数	528

表 1: 年度毎修正件数

表 2 では、修正内容の内訳を示している。もっとも大掛かりな修正は、項目の統合や分割であるが、これらは件数は少ないものの、実際に試験運用を行ってから要求されることがほとんどであった。ラベル、タイトル、画面表示変更は、単に Web ページで表示される文の変更であり、比較的容易な修正作業である。入力形式変更の中には、自由記述であったものを選択ボタン二つで表現したり、選択ボタンで入力していたものをチェックボックスによる入力に変更したりするものが含まれている。

画面表示変更	174
項目削除	12
項目統合	4
項目分割	1
新規項目	15
入力形式変更	260
全面修正	7
タイトル変更	35
ラベル変更	10
レイアウト変更	10

表 2 : 修正内容内訳

このように、システムの設計、試験運用、本稼働後のすべてのプロセスにおいてデータスキーマは変更されているが、XDES の変更作業は1人の管理者が最大でも1日、システムを停止させることによって約 2300 名の教員がそれぞれにデータを入力できる環境を提供している。

4. 関連研究

本システムはあくまでもデータエントリのためのシステムであり、統計処理システムや検索システムと連携して利用されることを想定している[1,4,5,6,8]。また、これらのサブシステムの連携やユーザのアカウント、権限、及び所属グループの管理機能は、ユーザ管理システムによって提供される必要がある。大学評価システムの応用では、ユーザ管理の方法は、平成 10 年から、教員の研究、教育及び社会連携活動についての情報をホームページで公開する目的で運用されていた「教官の研究教育活動等報告書データベース」[6] で用いられてきた方法に基づき、部局や部門毎に数名ずつ配置されているグループ管理者が、その部局や部門に所属する教員の ID 発行、パスワード管理、代行入力、及び公開データの内容の審査を行う管理体制を引き続き採用した。このシステムは、教員によるデータの逐次更新と、公開データの内容チェックやユーザアカウント管理を部局毎に行うための機能提供に重点を置いた設計となっており、実際に九州大学全部局で活発に活用されるシステムとして浸透させたことによる功績は大きい。しかし、このシステムでは、関係データベースをつかっており、データの形式がシステムの全体に深く関連しているため、データスキーマの変更に対応するためには、システム修正に時間がかかるという問題があった。また、教員データを管理する部局毎に階層化して管理していたため、複数のグループに所属したり、学内での異動が発生した際に、所属変更作業がスムーズに行えないこともあった。

大学評価システムとしては、他に大阪大学、徳島大学、九州工業大学などで開発されているものがある。また、学務情報などを幅広く扱うシステムとして、熊本大学の SOSEKI がある。このような稼働中のシステムとしても、あるいは研究の方向としても、ユーザーインターフェースとしては Web、データベースとしても RDB から XML に向かいつつある。

5. まとめと今後の課題

多様な構造と流動的なデータスキーマの変更に対応できるデータエントリシステム XDES について述べた。デ

ータスキーマを独立したマクロファイルとして管理することにより、データ項目の追加や削除、入力形式の変更等の作業を短時間でできる汎用データエントリシステムとなっている。また実用例として XDES により実現された大学評価システムの運用状況を紹介した。

従来のシステムと比べ、運用のコストは格段に低減しているが、項目変更作業を完全に自動化するまでに至っておらず人手による個別作業が必要となることもある。その一つに、蓄積データの形式変換の問題がある。すでにデータが蓄積された項目の構造を変更しなければならない場合、これまでは、個別に変換プログラムを作成して対応してきた。この作業を自動化することによって、メンテナンス作業をさらに軽減することができる。具体的には、マクロの XML 化により対応できると考えている。マクロの XML 化は、それ以外に、データスキーマ群の管理を、XDES で実現するメタ・データベースを可能とする。現時点では、マクロの編集はテキストエディタで行っているが、その作業が XDES で実現できることになりより汎用性の高いものとなる。その他にも、データスキーマから蓄積データの XML スキーマを自動生成する問題や、XML データからデータスキーマを自動生成する問題、XML データの妥当性検証やデータベース自動生成等の技術開発としても、実用的な問題への応用としても多くの展開が期待される[7]。

参考文献

- [1] 石野明, 竹田正幸: ストリーム指向の XQuery 処理システムについて, 情報処理学会 第 49 回デジタル・ドキュメント研究会資料 (2005)。
- [2] 杉本典子, 金丸玲子, 池田大輔, 竹田正幸, 井上仁, 廣川佐千男: 九州大学自己点検・評価関連情報システム, 情報処理学会第 41 回デジタル・ドキュメント研究会資料 (2003)。
- [3] 杉本典子, 廣川佐千男: データベース入力サーバ装置及データベース入力システム, 特願 2003-76817 (2003)。
- [4] 関隆宏, 廣川佐千男: 教員連想検索システム, 情報処理学会第 48 回デジタル・ドキュメント研究会資料 (2005)。
- [5] 竹田正幸, 宮本哲, 石野明, 辻寿嗣: 高速一方向逐次処理技術に基づく XML 文書の検索と変換, 情報処理学会第 41 回デジタル・ドキュメント研究会資料 (2003)。
- [6] 竹田正幸, 石野明, 辻寿嗣, 宮本哲: ストリーム指向の高速 XML データ処理技法について, データベースと Web 情報システムに関するシンポジウム DBWeb2003 (2003)。
- [7] Volker Turau: A framework for automatic generation of web-based data entry applications based on XML, Proc. 2002 ACM Symposium on Applied Computing, 1121-1126 (2002)。
- [8] 辻寿嗣, 石野明, 竹田正幸: 高速正則生垣パターン照合アルゴリズムに基づく XML テキスト変換, 電子情報通信学会コンピュータシミュレーション研究会資料 (2003)。
- [9] 野中裕介, 井上創造, 秦野克彦, 原田努, 乃村能成, 岩井原瑞穂, 峯恒憲, 牛島和夫: 教官の研究教育活動等報告書データベースシステムの開発と運用, 電子情報通信学会論文誌 D-I Vol. J84-D-I No.6, 974-986 (2001)。