

XML を利用した卒業論文管理システムの考察

坂本 健治*¹, 笥 宗徳*², 二石 芳裕*³, 渡辺 一衛*⁴

A consideration on graduation and master theses management system using XML

Kenji SAKAMOTO*¹, Munenori KAKEHI*²,
Yoshihiro FUTATSUIISHI*³, Ichie WATANABE*⁴

ABSTRACT: At Production System laboratory, the system which keeps graduation and master theses by electronic data (PDF etc.) was built in groupware. However, since all theses are not kept, in order that students may look for theses, students can spend time or cannot find theses after all. In this study we develop the system which supports the work (search, lending service, return etc.) about graduation theses and master theses. First, we analyze the present condition and enumerate required information items on work. In order to make easy to process enumerated information items, we define information structure. Finally, we build the database using XML (eXtensible Markup Language) according to the structure.

KEYWORDS: XML, Management System, System Analysis

(Received June 23, 2003)

1. 研究の背景と目的

2000 年度の修士論文[1]で構築されたグループウェアにおいて、論文を電子データ (PDF: Portable Document Format 形式など) で保管するシステムが提案された。しかし、論文の電子化あるいは PDF 形式への変換作業が滞りがちになり、全ての論文が検索できる状態にはなっていない。また、電子化されていない紙媒体の論文については、検索が難しく、探すのに時間がかかる。

一方、電子情報を管理する技術の 1 つとして XML (eXtensible Markup Language) が着目されている。

XML はインターネットの急速な普及に伴い、情報の共有化や一元管理[2]、情報交換のフォーマット (B2B や B2C) や電子政府に置ける電子申請書のフォーマットに利用されている。また、ソフトウェアの開発においては、インターフェイスとなる画面の背景やボタンの大きさ、Windows にインストールされている IME (Input Method Editor: 日本語入力のためのフロントエンドプロセッサ) の日本語入力の ON/OFF の設定情報などに幅広く用いられており、今後の応用や利用方法に大きな期待がもたれている。

そこで、本研究では XML を用い、研究室で行われている論文に関わる作業を支援するシステムを作成することを目的とする。具体的には、論文情報を取り出しやすくするために、情報の構造化を行い、構造化した論文情報を管理するために、XML を利用したデータベースを作成する。本研究では、研究室に提出された「学士論文」と「修士論文」をまとめて卒業論文と総称する。

学術論文の標準化については、科学技術情報流通技術基準 SIST08[3]があるが、本研究では研究室内の利用を中心に行っているため、タスクの分析をもとに情報項目を設定し、データベースの構築を行うことにした。

* 1 大学院情報処理専攻博士前期課程 現・大学院情報処理専攻博士後期課程

* 2 大学院情報処理専攻博士後期課程

* 3 成蹊大学客員研究員 現・株式会社グリーンフィールド

* 4 経営・情報工学科教授(watanabe@is.seikei.ac.jp)

Professor, Dept. of Industrial Engineering and
Information Science

2. 現状調査

論文の保管状況の調査として、提出された論文の調査、保管場所の調査、保管形態の調査を行った。その調査結果を表1に示す。保管場所には棚とサーバがあり、棚に保管されている論文は製本、クリップ、紙袋、バインダーの4つの形態がある。棚に保管されている論文の7割が製本の形態で保管されている。学士論文については全ての論文が保管されているが、修士論文については、所在がわからない論文が3本あった。学士論文は2000年度から、修士論文は2001年度からサーバ上に論文をWord形式で保管し、研究室内に限り閲覧できるようにしている。

次に、論文の利用者の調査として、論文の使用目的の調査と論文を取り扱う作業の調査を行った。主たる論文の使用目的には、以下の4点があった。

- 1) どのような研究がされてきているのかを調べる
- 2) 特定の研究分野の論文を読む
- 3) 論文の書き方を参考にする
- 4) 参考文献に引用する

1)は研究室の論文について詳しい人がいないと、探し出す論文がわからない場合である。2)から4)は、探す論文が決まっている場合である。論文を取り扱う作業内容は、

- ・検索 ・閲覧 ・貸し出し ・返却 ・登録

の5つから構成されていることがわかった。

さらに、論文を探す前作業として、

- ① 探し出す論文が何であるかを特定する

表1. 保管場所と保管形態の内訳

場所	形態	学士	修士	学士	修士
棚	製本	119	19	90	7
	クリップ			20	0
	紙袋			3	0
	バインダー			6	12
サーバ	Word	15	2	15	1
	PDF			0	1
どこにあるかわからない		0	3	0	3
合計		134	24	134	24

- ② 探し出す論文が棚とサーバのどちらに保管されているのかを特定する

という2つの作業が必要であることがわかった。①は論文の使用目的の1)より、②は保管状況の調査よりわかったことである。①、②共に論文について詳しい人に聞くこととなる。

以上より図1に示す6つの作業に分類できることがわ

かった。この作業内容を以下に示す。

話を聞く：論文について詳しい人に、おおよその保管場所を尋ねる、どんな論文があるかを尋ねる

棚にある論文を検索する：棚にある論文を1つ1つ見て、目的の論文が見つかるまで繰り返す

論文を借りる：棚から探し出した論文を借りる時、研究室にある白板に論文題名または著者名、日付、自分の名前などを記入する、メモ用紙に記入して白板に磁石で貼るものもある

論文を棚に返す：借りる時に白板に記入した情報を消し、借りた論文を棚に戻す。記入したメモ用紙を白板からはがし、借りた論文を棚に戻す

サーバにある論文を検索する：サーバ上にある論文は、発表年度内の学士または修士のいずれかのフォルダに分かれ、そのフォルダの中に論文が入っているフォルダがある。そのフォルダを1つ1つ開いて、論文を探す

論文を登録する：サーバ上のフォルダを更新できる権限があるネットワーク管理者が、提出された論文をサーバに登録する。この作業は年1回4月に行っている

3. 分析・改善

前章に示した6つの作業について、仕事の構造による分析[4]とフローチャートによる分析を行った。

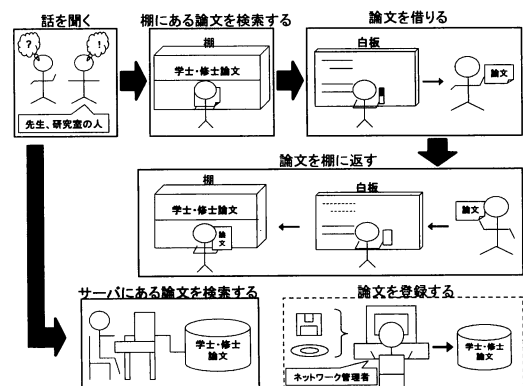


図1. 論文に関する作業の流れ

これらの分析により57個の問題点が列挙され、一部の作業の改善を行った。

3.1 仕事の構造による分析

図2は、「棚にある論文を検索する」作業についての仕事の構造分析の例である。分析範囲は“利用者（検索者）”が目的の論文を棚から探し出し、“探すべき論文を手にした利用者”になるまでである。

わかったことの例として、

- 論文題名と著者名は対で覚えていることが多いので、論文の検索キーは論文題名の場合がほとんどである
- 背表紙に論文題名が記載されていると、探すべき論文と異なる論文を棚から引き出さなくていいので、検索が早くできる

問題点の例として、

- 棚の前に行かないと検索が始められない
- 探すべき論文を違う論文を引き抜いた場合、棚に戻さなくてはならない

などがあつた。同様に、残りの5つの作業についても分析を行った。

3.2 フローチャートによる分析

前節で行った仕事の構造の分析の「変化」について、条件による作業手順の分岐や細かい作業の分析を行い、変化の内容での問題点を明確にした。

図3は、「棚にある論文を検索する」作業についてのフローチャートであり、丸数字と点線の囲みは前節で行った仕事の構造の分析の変化の項目番号である。わかったことの例として、

- 背表紙に情報がある場合は、検索するステップ数が増えるが、目的でない論文を引き出し、戻すステップが省ける

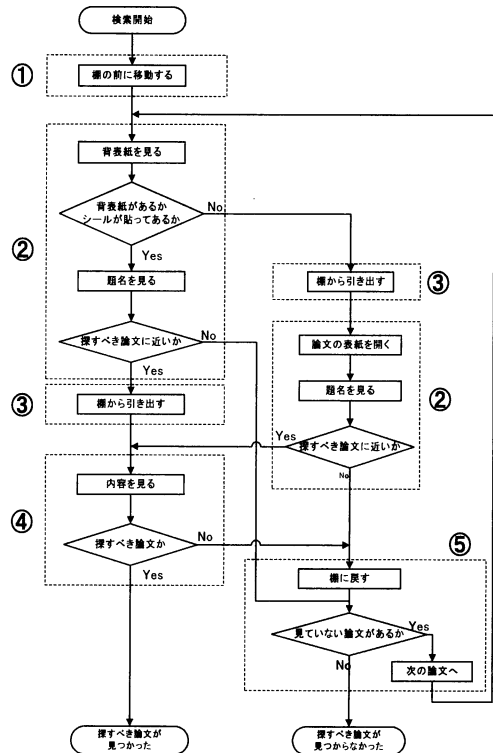


図3. フローチャートの例

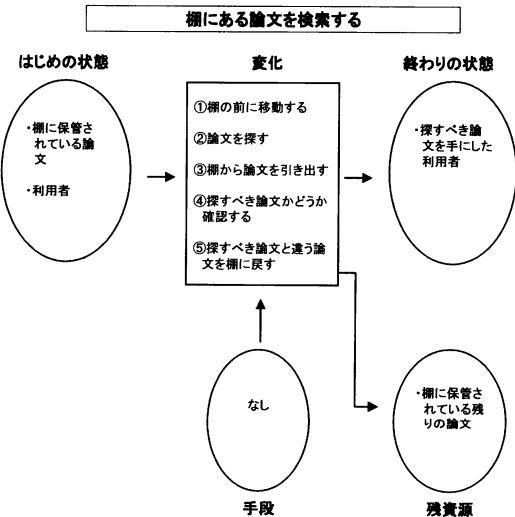


図2. 仕事の構造図の例 (物レベル)

問題点の例として、

- 棚にある論文の保管形態が異なるため、論文ごとに検索の仕方を変えなければならない
- 背表紙に題名などの情報が書いてあるシールがない場合、必要としない論文も棚から引き出し、確認しなければならない

などがあつた。同様に、残りの5つの作業についても分析を行った。

3.3 作業改善とシステム化

列挙された問題点を分類すると、以下の5項目に整理できた。

- 棚についての問題点
- サーバについての問題点
- 白板についての問題点
- 保管形態についての問題点
- 人の頭の中にある情報に関する問題点

改善のアイデアの例を以下に示す。

棚についての問題点：発表年度ごとに論文を整理し、年度内で通し番号を付けることで、通し番号のみを見て探すことができる

サーバの問題：論文にたどり着くまでのファイルパスを明確に示すことで、他のフォルダを開くことなく論文を見つけることができる

白板についての問題点：貸し出し情報を別に管理することで、白板上の他の情報と混ざること防ぐことができる。利用者自身が借りている論文のみの処理を扱えるようにすることで、間違えて他人が借りている論文情報を消すことを防ぐことができる。

保管形態についての問題点：背表紙のあるなしに関係なく表紙に番号を付け、棚から論文を少し引き出すという同じ手順で論文を探すことができる

人の頭の中にある情報に関する問題点：情報をデータベース化することで、いつでも取り出せるようにするこれらの改善をすることで、52個の問題点が解消できることがわかった。

3. 4 システムへの要望

システムを作成する上で、以下の要望があった。

- 1) 論文が掲載された学術誌の情報を取り入れる
- 2) 簡単に論文情報が入力できるようにする
- 3) 自宅のパソコンや携帯端末からもアクセス可能にする

本研究では、1)と2)は情報項目を新たに設けることにより採用し、3)はセキュリティ問題より外した。

4. 情報の構造化

論文に関する作業について、情報項目を列挙し、各々の作業を行う上で必要な情報とそうでない情報に分類し、XMLを利用したデータベースを構築するための情報の構造化を行う。改善前に必要な情報項目が改善後に欠落してしまうことを防ぐため、3章の分析・改善と同時に分析した。

4. 1 情報項目の列挙と分類

論文に関する6つの作業について、3.1節で行った仕事の構造図に、「もの」が持っている情報項目を列挙した。その結果、全部で25項目が列挙された。しかし、この情報項目には、同じ情報を作業内容によって、言い換えている項目が含まれていた。例として、「借りたい論文題名」、「借りた論文題名」、「返却する論文題名」があるが、これらは全て「論文題名」である。このように、作業内容で変化する情報項目を統一すると、12項目まで絞り込むことができた。

図4は「棚にある論文を検索する」作業について情報項目を列挙し、仕事の構造図に整理したものである。はじめの状態の「もの」は「棚に保管されている論文」と「利用者」である。終わりの状態の「もの」は「探すべき論文を手にした利用者」となり、手段の「もの」はな

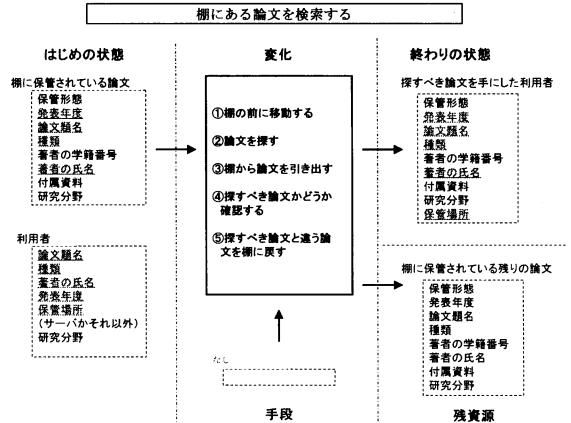


図4. 仕事の構造図の例（情報レベル）

く、残資源の「もの」は「棚に保管されている残りの論文」である。この作業は利用者が論文を検索するので、利用者が持っている情報を使って作業をする。利用者の持っている情報の中で、「保管場所」の情報は、論文が棚にあるのかサーバにあるのかで、検索する対象が異なる。この作業を始めるには保管場所として「棚」である必要がある。よって、「保管場所」はこの作業では必要な情報となる。論文を特定するための情報として、「論文題名」、「種類」、「著者の氏名」、「発表年度」の4つがある。しかし、全ての情報が揃っている必要はない。それぞれ単体でもいいが、組み合わせの情報で論文を特定できる。ただし、「種類」と「発表年度」の組み合わせでは、候補となる論文が多すぎるので特定できない。このようにして、「棚にある論文を検索する」作業で必要となる情報項目は「保管場所」、「論文題名」、「種類」、「著者の氏名」、「発表年度」となる。同様に、残りの5つの作業について、作業に必要な情報項目を検討すると、以下の情報項目が必要であることがわかった。

- 1) 研究分野
- 2) 論文題名
- 3) 種類
- 4) 著者の氏名
- 5) 発表年度
- 6) 保管場所
- 7) 利用者の氏名
- 8) 貸出日

4. 2 情報項目の構造化

作業を行う上で必要な情報を見ると、1)から5)までは論文を特定するために必要な情報であり、5)から8)は論文がどこにあるか特定するために必要な情報である。前者を「論文情報」、後者を「保管情報」とする。また、システムの要望にあった「自研究室の保管されている論文で参考文献として引用されている論文も読みたい」、「必要な参考文献がどの論文誌に載っているのかわかるとよい」を実現するために、「付属情報」と定義して「参考文献」や「記載文献」の項目を追加した。

論文情報を構造化した結果を図5に示す。「論文情報」

の中に“著者の氏名”があるが、著者は学生であり、個人を特定するための固有番号、つまり、学籍番号が必要である。著者の氏名と対等な情報項目なので、“著者”という項目を新たに追加し、その下に“学籍番号”と“氏名”を定義した。自研究室の論文以外に他の研究室から提出された論文と、他の研究室に提出した論文も存在することから、それらの論文を保管するために“提出先”と“受け入れ先”を定義した。「保管情報」の“保管場所”の「場所1」は棚かサーバか、場所2は棚の場合は棚番地内の通し番号、サーバの場合は論文が入っているフォルダ名である。“保管状況”の“保管形態”は現状調査からわかったもので、製本、クリップ、紙袋、バインダー、Word、PDFのいずれかである。“現在の状態”において、“利用者の氏名”と“貸出日”にデータがある時は貸出中、ない時は閲覧が可能な状態を表わしている。“付属資料”は論文を読むときに、より詳しく理解したり、レジュメの書き方の参考にしたりなど利用用途が広いので、「保管情報」に追加した。

5. データベースの実装

XMLとは、“eXtensible Markup Language”の略で、標準化された拡張可能なマーク付け言語である。開始

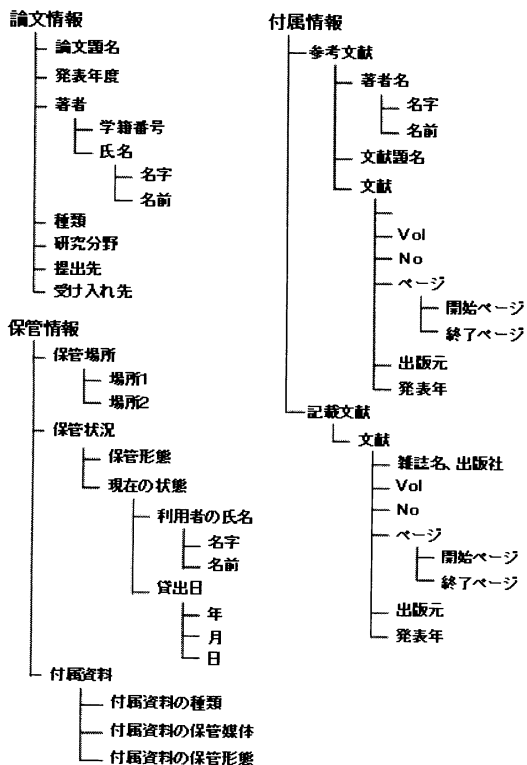


図5. 論文情報の構造

タグと終了タグでデータを挟んだものを要素という。XML文書は親要素と子要素からなる階層構造で表現される。XML文書は、

・XML宣言 ・要素宣言 ・文書本体

の3つからなる。特に要素宣言が不要なXML文書を整形形式文書（well-formed XML document）という。本論文で作成するシステムでは、外部とデータの交換をする処理が発生せず、システム内でのみデータを扱うので、データ要素宣言を必要としない整形形式文書で作成する。

4. 2節で行った論文情報の構造化を基に、XMLによるデータベースの実装を行う。

- 1) 情報項目をそのままタグ名に置き換える
- 2) 構造に従って、親要素に子要素を入れる
- 3) 同じレベルの階層構造（子要素）が複数存在する場合、その同レベルの子要素をまとめるための親要素を作成する

学士論文の場合、共同で研究することがあり、著者＝学籍番号－氏名の構造が共同研究者分だけ繰り返される。よって、著者をまとめる要素として“著者リスト”を親要素として定義する。“著者リスト”を開くことで、その下の階層にある全ての著者を見ることが出来る。この関係を図6に示す。同様に、同じ構造が繰り返される理由から、研究分野リスト、提出先リスト、受け入れ先リスト、付属資料リスト、参考文献リスト、記載文献リストを定義した。

6. システム設計・構築

6. 1 システムに必要な機能

3. 3節より、システムには“論文の検索”、“論文の貸し出し管理”、“論文の返却管理”が必要であることがわかった。さらに、システムには以下の機能が必要であることがわかった。

- ①特定の論文の情報を見るための詳細表示
- ②データベースに情報を登録または修正および追加するためのデータベース登録および修正
- ③システムを使ってもらう時に必要なユーザ名とパスワードを設定するユーザのアカウント登録および修正

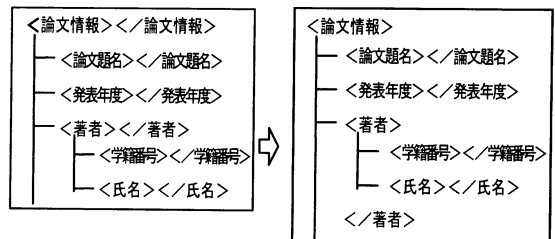


図6. タグ付けによる構造の変化

よって、システムに必要な機能は

- 1) 論文検索 2) 貸し出し管理
- 3) 返却管理 4) 論文情報の詳細表示
- 5) 論文登録・修正 6) アカウント登録・修正

である。1)～4)の作業は、通常のユーザが使うことができる機能であり、5)と6)は、システムの誤動作や故障を防ぐために、システム管理者が使う機能とした。

6.2 システムの構成

作業改善後の作業の流れに沿ったシステムの構成を図7に示す。機能と第3章のシステム導入後の作業の流れから、機能1), 2), 4)を行う「検索・貸し出し処理」、機能3)を行う「返却処理」、および機能5)と6)を行う「システムメンテナンス処理」の3つに分かれる。

また、「返却処理」の場合、細かい検索は必要ないが、どの論文を返却するか特定しなければならない。現在借

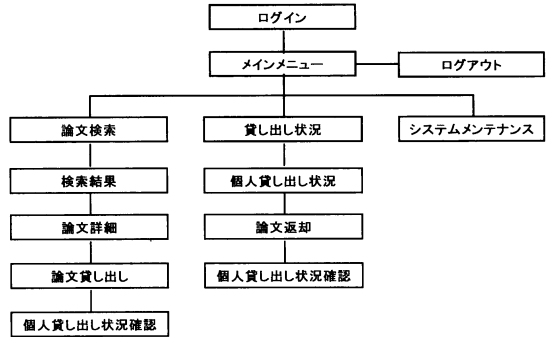


図7. システム構成図

りている論文の一覧が表示されれば、ユーザ自身の中から選択するだけでよいと考え、貸し出し状況確認の機能を追加した。

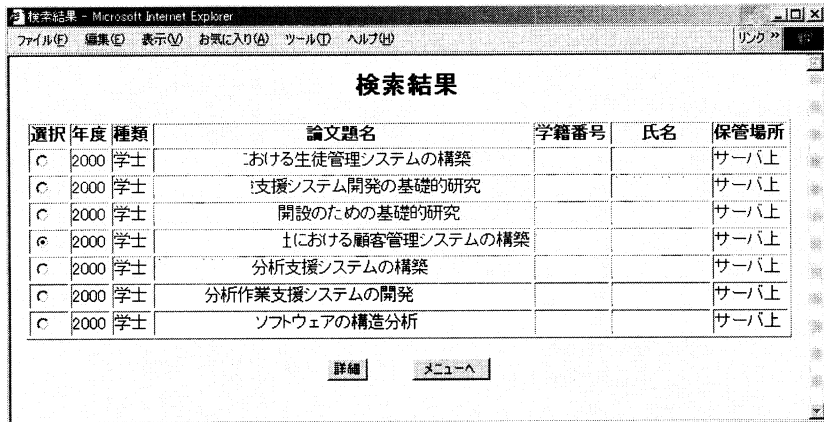


図8. クライアント側の検索結果の表示

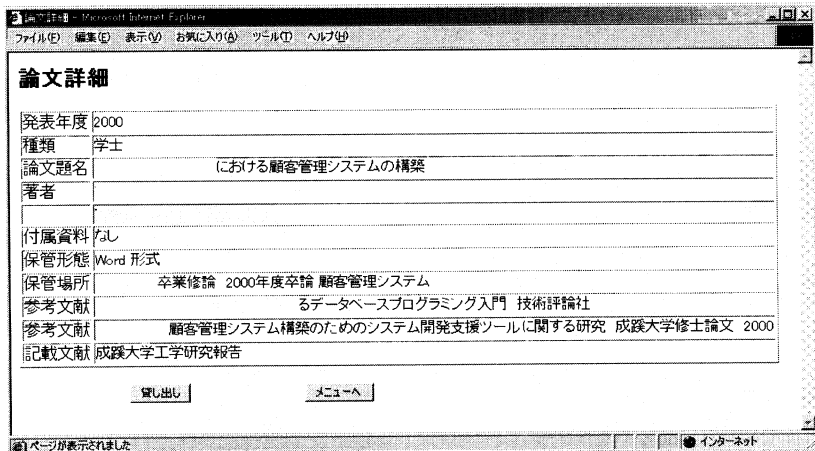


図9. 論文の詳細の表示

6. 3 構築ツールの選択

システムを構築する際にツールとして、「横浜ベイキット」[5][6]を使用する。特徴として、以下の項目があげられる。

- ・サーバソフトとXML文書を扱う言語(Xi: Extend it!)が一体になっている。
- ・XML文書の構造を変更せずに扱える。
- ・Xiの中に“Ninja”というAPI群があり、ファイルの入出力や数値演算などができる。

などがあげられる。

システムを構築するために使用したツールは以下の3つである。

- ①Baykit XML Server(BXS) (Ver. 1.0.26) : サーバソフト
- ②Baykit Extends it!(Bxi) (Ver. 1.1.4) : BaykitによるXiの処理系
- ③Extend it!(Xi) : XML形式のプログラミング言語

6. 4 システム構築

論文を特定するための基本情報を論文題名と著者の氏名の両方で検索しようとしたが、応答時間とサーバにかかる負荷を考えて、システム内部の情報として、論文に固有番号(ID)を付けることにした。その結果、検索結果から詳細を出力する時に、ID番号を探す処理だけでよくなった。ID番号は論文タグの属性として定義した。論文の検索の結果はXML文書で返され、クライアント側では、図8に示す一覧表の形式で表示される。左側のオプションボタンで選択し、「詳細」を押すことで、図9に示す特定の論文の詳細を見ることができる。このとき、サーバ側ではどの論文が選択されたかがわからないという問題が発生した。そこで、オプションボタンが押されたときに、一覧表において上から何番目の論文が選択されたかを検出するために、JavaScriptを使って、何番目のオプションボタンが押されたかを計算し、その数字を表示していないテキストフィールドに数字を書き込む。こうすることで、サーバ側では、検索結果のXML文書をテキストフィールドに書き込まれた数字の分だけ論文タグを読み飛ばすことで論文の詳細を表示できるようにした。

7. 結論と今後の課題

7. 1 結論

- ・仕事の構造の分析とフローチャートによる作業改善ができた。
- ・仕事の構造の分析を情報レベルで整理することによって、作業に必要な情報項目の分類ができた。
- ・情報の構造化からXML文書を利用したデータベースが構築できた。

- ・棚にある論文のロケーションを整えることで、論文の保管場所の情報が取り出しやすくなった。

- ・先生や院生に論文のことを聞かずに、論文を探し出せるようになった。

7. 2 今後の課題

- ・他の研究室の論文管理に適用し、欠落している作業や情報がないかを確認する。

- ・サーバ上にある論文のファイル名を英数半角文字に置き換えることで、検索結果から直接論文を開ける、またはダウンロードできるようにする。

- ・論文の内容(中身)についても構造化し、論文で使われている手法や分析方法での検索をできるようにする。

- ・セキュリティの問題を解消して、研究室以外からのアクセスを可能にする。

[参考文献]

- [1] 白崎誠, “小規模業務システムにおける情報共有アプリケーションの利用に関する基礎的研究”, 成蹊大学修士論文, 2001
- [2] 西村龍文, “ドキュメント管理システムの開発と運用事例”, テクニカルコミュニケーションシンポジウム2001論文集, pp.47-51, 2001
- [3] 科学技術情報流通技術基準 ハンドブック 2003 SIST08 学術論文の構成とその要素, <http://www.jst.go.jp/SIST/>
- [4] 中村善太郎, 「新改善技術 もの・こと分析」, 日本能率協会マネジメントセンター, pp.105-222, 1992.
- [5] 横浜ベイキット, <http://www.baykit.org/>
- [6] XML PRESS Vol.6 「特集1 ベイキット完全マニュアル」, 技術評論社, pp.2-61, 2002