

XML のプレゼンテーションと検索

○重元 康昌、藤澤 由美、宮崎 智、菅原 秀明

A Presentation and Search of XML

○Yasumasa Shigemoto, Yumi Fujisawa, Satoru Miyazaki,
Hideaki Sugawara

Abstract

It has been three years since the specification of XML was announced. In the meantime, recommendations have been also made for technologies for applying XML to applications. Here we summarize the merit of XML technology and introduce the application to an actual data system, i.e., databases of WFCC-MIRCEN World Data Centre for Microorganisms (WDCM).

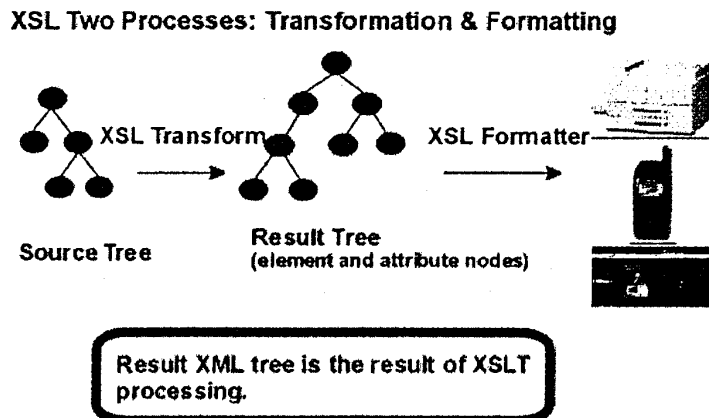
1. はじめに

XML 基本仕様の勧告後3年が経ち、応用技術についても勧告やドラフトが発表されつつある。ここではそれらの技術のメリットを具体的に述べ、適用事例としてWDCMシステムを紹介する。

2. XML のメリット

2.1 出力を用途に応じて柔軟に

XML 自身は表示方法を持たないがデータの出力形式を柔軟に変更できる。



<http://www.w3.org/TR/xsl/>より引用

図1. XML のフォーマットプロセス

出力制御例：抽出・追加・削除・ソート

出力フォーマット例：HTML・PDF (帳票)・SVG (2次元グラフィックス)・
X3D (3次元グラフィックス)・VoiceXML (音声)

2.2 データに意味付けが可能

XML のタグは自由に決めることができるため、データに意味付けを行うことができる。

```

<本>
  <タイトル>XML 入門</タイトル>   XML 入門の価格は？
  <著者>やまだたろう</著者>
  <価格>3000</価格>
</本>

```

図 2. XML と検索例

3. XML のプレゼンテーション技術

3.1 HTML・PDF・SVG への変換

XML をテキストとしてエディタなどで閲覧することは可能だが、ウェブからの情報発信に適した箇条書き、テーブル表示、色づけなどのために HTML へ変換することができる。また、帳票を出力するための PDF への変換やグラフなどグラフィカルに出力するための SVG への変換なども可能である。

3.2 XSL (eXtensible Stylesheet Language)

XML ファイルには表示に関する情報は通常入っていないため、表示に関する定義を記述したファイル (XSL) を指定する。このファイルの記述により追加・抽出・ソートなどの制御やさまざまなフォーマットへ変換を行うことが可能である。

3.3 サーバ側での変換

ウェブブラウザの中には XML に対応していないものもあるので、XML 由来の表示が利用者の環境に左右されないように、サーバでの変換処理が不可欠である。そこで、Apache XML Project の Cocoon というソフトウェアを用いることにより、サーバでの変換を行うことが必要であり、可能である。

4. XML の検索技術

4.1 DOM による検索

XML ファイルから XML パーサーというソフトウェアを用いることにより、ツリーオブジェクトを生成し、XML データをプログラムから扱えるようにする。ツリーオブジェクト生成後は、ツリーの順番やタグ名を指定してデータの取得や検索が可能である。

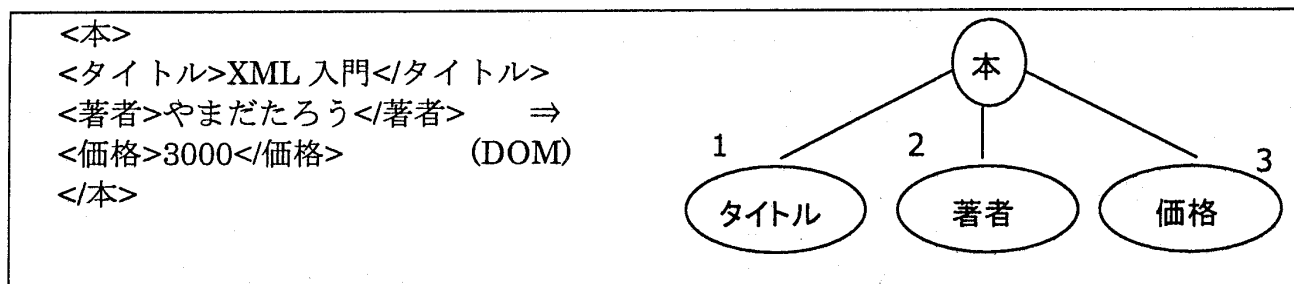


図 3. DOM によるツリー生成

4.2 XML 専用 DB

XML の木構造を考えた場合、XML をそのまま格納できるようなデータベースとして、オブジェクト指向データベースがあげられる。また、キーワード検索ソフトウェア(検索エンジン)のXML対応版として発展してXMLファイルからインデックスを作成するようなソフトウェアもあるが、まだこの分野のソフトウェアは大規模なデータベースを対象に考えた場合、パフォーマンスに限界があるようである。

4.3 RDB を用いた検索

XML の個々の要素をテーブルのカラムに保存する。検索は RDB に対して行うため、パフォーマンスなども含めて十分な機能を得ることができる。

また、XML として閲覧する際にはテーブルから組み立てる。RDB から XML ファイルの組み立ては動的 XML ファイル作成技術の XSP(extensible Server Pages) や ESQLEngine TagLibrary などの技術を使うことによって実現可能である。

下の図は、RDB から XML ファイルを組み立てる際の実際の処理の流れである。

クライアントから①のような形式でアクセスすると、②のような XML 定義ファイルから③のような XML ファイルが動的に作成されるようすを示している。

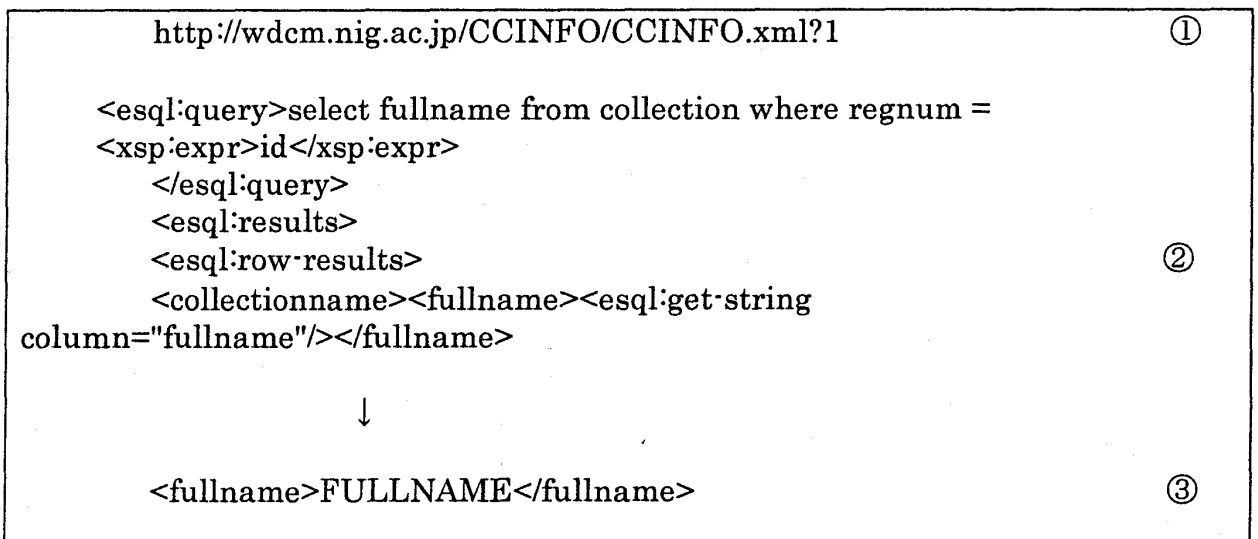


図4. RDB から XML ファイルの組み立て

5. 適用事例

国立遺伝学研究所の WDCM (<http://wdcm.nig.ac.jp/>) にて XML のさまざまな技術を用いてシステムを構築した。

データベースへの登録や検索はプログラミングを必要としたが、閲覧に関しては XML の技術を用いることによって開発コストを削減し、短期間での開発を達成した。

特に Web では XML から HTML への変換処理を行うことによってブラウザからデータの閲覧を行うことができるようにし、一方カタログや提出書類など帳票処理を必要とする情報は XML から PDF へ変換処理を行ったがこの変換処理は XSL ファイルを用意するだけの簡単な開発で実現することができた。また、RDB から XML への組み立てに関してもプログラミングで行うのではなく、 Cocoon の技術を使用することにより定義ファイルを作成するだけで実現することができた。

システム	適用技術
登録	Servlet, JDBC, RDB
全文キーワード検索	Namazu
フィールド検索	Servlet, JDBC, RDB
データの閲覧	XSP, ESQ L TagLibrary, XSL

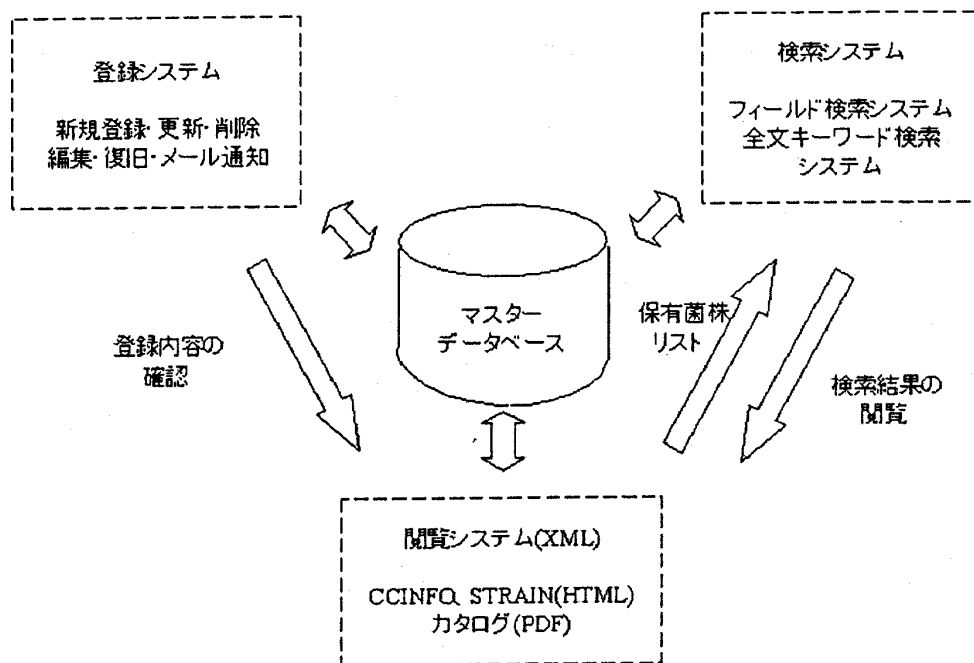


図 5. WDCM システム (XML 技術適用事例)

6. まとめ、今後の課題

WDCM システムに適用し実際に開発・運用することによって、XML 技術がさまざまなシステムに有用であることを証明した。今後は XML そのものを格納するためのデータベースやインデキシングツール、グラフィカルに表示したり音声での出力などの技術が進めば、さらにさまざまな分野で適用される技術であると思われる。

参考文献

- 1) WDCM HOME (<http://wdcm.nig.ac.jp/>)
- 2) Apache XML Project (<http://xml.apache.org/>)
- 3) The World Wide Web Consortium (<http://www.w3.org/>)

重元 康昌 富士通株式会社 (〒261-8588 千葉県美浜区中瀬 1-9-3 幕張システム
ラボラトリ)

藤澤 由美 国立遺伝学研究所 (〒411-8540 静岡県三島市谷田 1111)

宮崎 智 同上

菅原 秀明 同上

Yasumasa Shigemoto (yshige@ssd.se.fujitsu.co.jp) Fujitsu Limited
Yumi Fujisawa (yfujisaw@genes.nig.ac.jp) National Institute of Genetics
Satoru Miyazaki (smiyazak@genes.nig.ac.jp) National Institute of Genetics
Hideaki Sugawara (hsugawar@genes.nig.ac.jp) National Institute of Genetics