

# XMLを用いた情報管理システム

## XML-Based Information Management Systems

野々村 克彦  
NONOMURA Katsuhiko

大盛 善啓  
OHMORI Yoshihiro

山本 晃司  
YAMAMOTO Koji

XML(eXtensible Markup Language)は、情報の関連づけの容易さやタグ単位の加工性といった特長を持つ。ここでは、XMLのそれらの特長を活用した二つの情報管理システム(マルチメディアアルバムシステムと技術文書分析システム)について述べる。マルチメディアアルバムシステムでは、当社の自動インデクシング技術とXMLマルチメディアフレームワークによりマルチメディアコンテンツの自動管理を可能にするとともに、高い拡張性を実現した。また、技術文書分析システムでは、当社のXML処理エンジンKnowledge Factoryの利用により、技術文書に対し多次元分析する機能を実現した。

Extensible Markup Language (XML) has two major features ; namely, reference expression to external information and tag-based expression for processing of information. This paper describes two XML-based information management systems that promote these features : a multimedia album system and a technical document analysis system. The multimedia album system can automatically manage multimedia contents by using automatic indexing techniques, and an XML multimedia framework within the system allows high extensibility. The technical document analysis system can process multidimensional analysis using an XML processing engine (Knowledge Factory).

### 1 まえがき

近年、情報の表現言語として、次の特長を持つXMLが注目されている。

- (1) 構造化表現により、情報の表現能力が高い。
- (2) スキーマを前提としないため、構造の変更ができる。
- (3) 外部データを参照により関連づけることができる。
- (4) タグ単位の細かい粒度での加工ができる。

ここでは、後者二つの特長を生かし、マルチメディアコンテンツの管理を自動化したマルチメディアアルバムシステムと、技術文書に対する多次元分析を実現した技術文書分析システム、の二つの事例について述べる。

### 2 マルチメディアアルバムシステム

近年のデジタル情報収集機器の普及に伴い、動画、静止画、サウンドなどのマルチメディアコンテンツを収集する機会が急激に増加している。一方で、収集されたマルチメディアコンテンツの管理は、例えばコンテンツの記録日時や内容に基づいて、ユーザーがあらかじめ用意したフォルダに保存するなどの手動で行うのが通常である。そのため、コンテンツの数が多くなると管理する手間と時間がかかるため、多くの場合はせっかく収集したコンテンツが、使われることなくハードディスクのどこかに保存されたままになっているのが現状である。

理想的には、収集したデータをブラックボックスに放り込

んでおけば自動的に整理されて、後から欲しいデータを簡単に探し出すことのできるシステムが望まれている。しかし、マルチメディアコンテンツはテキストコンテンツと異なり、内容の判断を自動化するには高度の技術を必要とするために、ユーザーが判断して分類しなければならない。更に、同時に収集したコンテンツを効率よく関連づける手段がなかったために、自動管理が難しかった。

このシステムは、入力されたデータから自動的にインデックス情報を抽出してXMLデータベース(DB)に登録しておく、ユーザーが見たい視点でデータを一覧できるインタフェースを用いてデータを検索することにより、あらかじめコンテンツの分類・整理をすることなく、マルチメディアコンテンツを自動管理できる。また、コンテンツ単位での管理だけではなく、ニュース番組のトピックのようにコンテンツの一部(シーン)単位での分類・管理も可能である。

#### 2.1 システム概要

このシステムの概要を図1に示す。ユーザーはデジタルスチルカメラ画像、テレビ(TV)番組などの動画、音楽などのマルチメディアコンテンツをマルチメディアアルバムシステムに登録する。

システムは、収集されたコンテンツから音声認識や映像解析を用いてインデックスとなるメタデータ<sup>(注1)</sup>を抽出したり、電子番組表からメタデータを取得したりして検索の手がかりとなる情報を自動的に蓄積する。ユーザーは、マルチメディアアルバムシステムを用いてメタデータを検索して、検索結

(注1) データについての情報を記述したデータ。

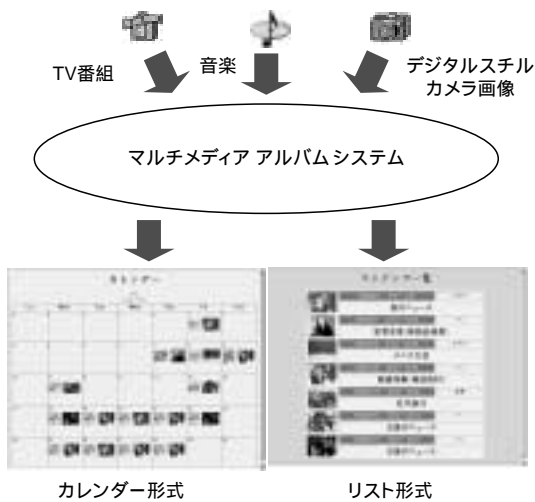


図1. システム概要 マルチメディアコンテンツは自動管理され、検索結果を好みの形式で見ることができる。  
Outline of multimedia album system

果をユーザーの好みに応じた形式で見ることができる。ユーザーはこの検索を繰り返すことにより、目的のデータに到達することができる。図1では、同じ検索結果をカレンダー形式やリスト形式で提示するようすを示している。

## 2.2 XMLマルチメディアフレームワーク

図2は、このシステムを実現するXMLマルチメディアフレームワークの概要であり、既存のXML DBをマルチメディアデータも扱えるように拡張したものである<sup>(1)(2)</sup>。

インデクシング部では、マルチメディアコンテンツを入力して、コンテンツをDB部に登録するとともに、コンテンツからインデックス情報となるメタデータもメタデータ抽出処理系により抽出して登録する。メタデータ抽出処理は主として、映像分割とキーワード抽出である。映像分割はカット点を自動検出し、連続するフレームの集合であるショットに分割し

て統合する処理を行う。キーワード抽出は映像に含まれるテロップ文字列の自動認識、及び音声認識を用いて行われる。抽出されたメタデータは電子番組表などから得られる情報とともに、DBに登録されるが、その記述フォーマットにはインデックス情報を記述する国際標準規格であるMPEG-7 (Moving Picture Experts Group 7) を利用している。

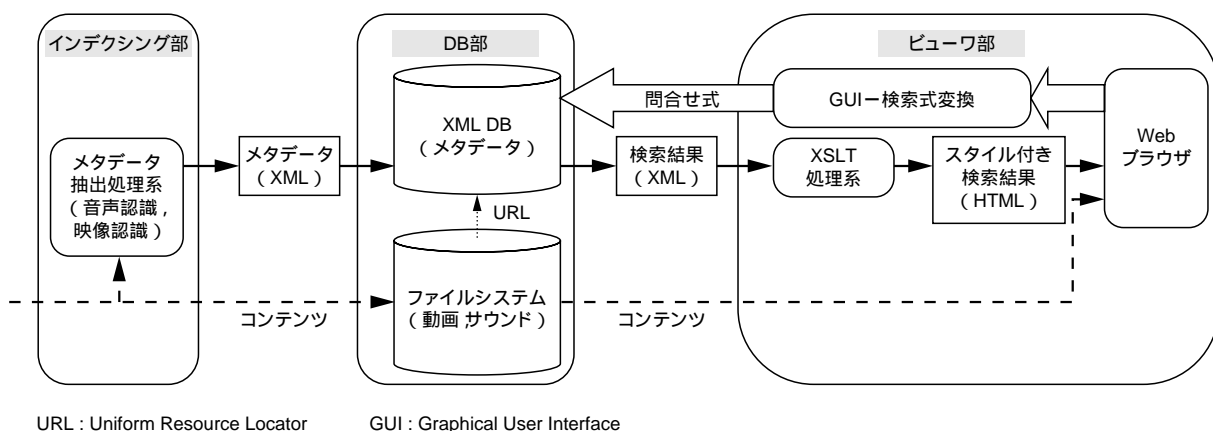
DB部は、XML DBとファイルシステムから構成される。XML DBではメタデータだけを管理し、コンテンツデータ自身はファイルシステムに保存して、XML DBには保存場所だけを登録する。

メタデータの検索は、XMLで記述された検索式を入力して、XML DBの検索機能を使用して検索し、結果をXML形式で出力する。

図3はビューワ部の外観を示すものであり、TVの録画映像を管理する例を表している。左ウインドウのキーワードはそれぞれ一つのXMLによる検索式に対応されており、ユー



図3. ビューワ外観 TVの録画映像を検索した例を示す。  
Example of viewer display



URL : Uniform Resource Locator

GUI : Graphical User Interface

図2. XMLマルチメディアフレームワーク XML multimedia framework モジュール間でのデータ交換と内部データ処理にXMLを採用することにより、様々な応用に適用できる。

ザーは例えばテレビ局名などのキーワードを選択することによって簡単にDB部に検索式を発行できる。検索結果はXMLとして返され、ユーザーがあらかじめ選択しておいた表示形式に従って、XSLT( eXtensible Stylesheet Language Transformations )処理系によりスタイル付きHTML( HyperText Markup Language )に変換されて検索ツールの右ウィンドウに表示される。この例では、検索結果をカレンダー形式で表示している。XSLTによるスタイル定義ファイルは自由に変更できるので、容易にカスタマイズ可能である。

XMLマルチメディアフレームワークでは、モジュール( インデクシング部、DB部、ビューワ部 )間でのデータ交換にXMLを採用してモジュールを独立させ、拡張性を高めている。例えば、インデクシング部に音楽を登録するソフトウェアを簡単に追加できる。また、XMLというオープンな規格を用いているため、カスタマイズに特殊な予備知識を必要とせず、既に流通しているXML処理系を再利用でき、開発期間の短縮や開発コストの低減が可能である。

このシステムは、様々な分野のマルチメディア コンテンツ管理に応用できる。例えば、TVの録画番組管理、デジタルスチルカメラ画像の管理、健康データ管理などへの応用を既に試みている。今後は個人向けマルチメディア情報管理システムとして応用される予定である。

### 3 技術文書分析システム

#### 3.1 技術文書のXML化による効果と問題点

技術文書の分析作業の代表例として特許調査がある。特許調査では、検索システムで得られた関連特許について、カテゴリ、重要度、コメントなどの付加情報を加味したうえで特許マップを作成する。従来、その作業には大きな手間がかかっていた。

タグ単位の細かい粒度での加工ができるXMLを利用することで、図4に示すように論文、特許、設計情報などの技術文書の分析プロセスを効率化するシステム構築ができると考えられる。分析プロセスとして、編集、検索などの基本プロセスや、コメントの付加や様々な観点での分析( テキスト版OLAP( On-Line Analytical Processing ) )などの高度なプロセスがある。

これまでのシステム構成として、図5(a)(b)に示すように大きく二通りある。しかし、それぞれ次に示す問題点があり、分析プロセスを効率化するシステム構築は困難であった(表1)。

- (a) RDB( Relational DB )による技術文書管理では、テーブルとXMLとの間の高コストの変換処理が必要である。
- (b) 既存のXML DBによる技術文書管理では、加工処

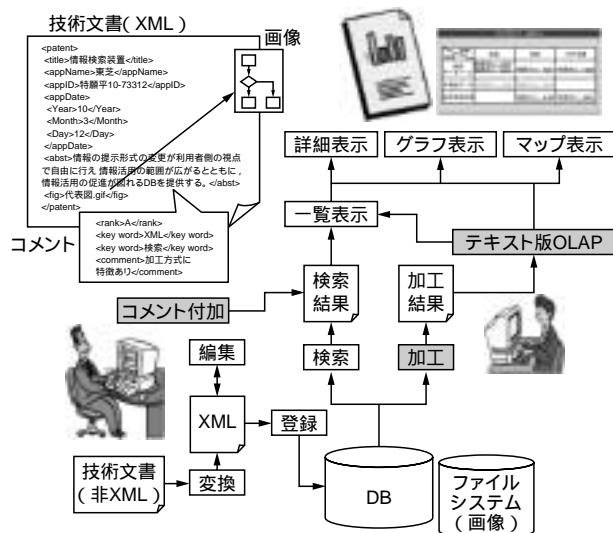
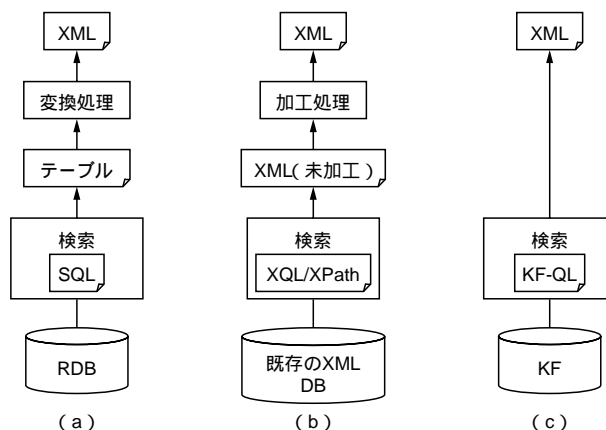


図4. 技術文書におけるプロセス 四角のブロックは基本プロセスを、網掛けした四角のブロックは高度プロセスを示す。  
Processing of technical document



SQL : Structured Query Language  
XQL : XML Query Language  
XPath : XML Path Language

図5. システム構成 DBとクエリによるシステム構成の違いを示す。  
System configurations

表1. システム構成の比較  
Comparison of system configurations of Fig. 5 (a), (b), and (c)

システム構成	(a)	(b)	(c)
DB	RDB	既存のXML DB	KF
モデル	テーブル (表現レベル低い)	ツリー (表現レベル高い)	ツリー (表現レベル高い)
管理対象	テーブル (XMLの部分情報)	個々のXML	XMLの集合
クエリ言語	SQL (加工レベル高い)	XQL, XPath (加工レベル低い)	KF-QL (加工レベル高い)
問題点	情報の欠落あり、 若しくは検索精度 が落ちる	加工処理をプログラ ミングする必要あり	



図6. 特許分析における画面例 クエリを再利用して分析を進めていく過程を示している。  
Example of user interface for patent analysis

理はプログラミングせざるをえない。

### 3.2 XML処理エンジンによる技術文書管理システム

そこで、図5(c)に示す高速で高度な加工が可能なXML処理エンジンKnowledge Factory(以下、KFと略記)この特集のp.11に掲載)とそのクエリ言語KF-QL(Query Language)を利用することで、従来システムの問題点が解決され、主に三つの機能を実現することができた。

- (1) テキスト版OLAP機能 KF-QLは複数の情報を結合する機能や集計する機能を持つ。この機能により、XML形式の技術文書と概念辞書とを組み合わせ、複数文書を分類し、様々な観点でのマップを作り出すことができる。更に、KFが提供するクエリへのリンク機能により、様々な画面へ遷移する多次元分析ツールを構築することができる。
- (2) コメント付加・再利用機能 KFは複数の情報をまとめて一つのツリーで管理できる。この機能により、付加されたコメントをオリジナルの技術文書とは独立に管理し、検索や加工で再利用することができる。
- (3) 分析プロセス共有・再利用機能 KF-QLはXML表現であるので、クエリ自体をKFに格納することができる。技術文書で頻繁になされる分析プロセスをパラメータ付きのクエリで蓄積しておくことで、エンドユーザーは分析プロセスを共有・再利用することができる。

特許分析における画面の例を図6に示す。蓄積されているクエリから“マップ”に関連するクエリ一覧を求めることで、分類カテゴリ“技術×機能”を軸とした特許マップを容易に作成できる。更に、“機能”のサブ分類カテゴリである“検索機能”を軸とした特許マップへのドリルダウンや詳細表示が可能である。

このシステムはKF上にシステム構築したことで、開発コストを削減でき、メンテナンス性と拡張性が向上した。また、KFが提供する細かい粒度での検索、加工技術により、検索

精度と検索処理速度が向上した。

今後は、自然言語処理の組み込みなどナレッジマネジメントシステムとしての機能強化を予定している。

## 4 あとがき

XMLを用いた情報管理システムとして、マルチメディアアルバムシステムと技術文書分析システムについて述べた。当社のXMLによるマルチメディアフレームワークとXML検索技術(KF)により、マルチメディアコンテンツ自動管理やテキスト版OLAPを実現し、XMLの付加価値を高めることができた。

## 文献

- (1) 山本晃司,ほか.“XMLを用いた後整理型マルチメディア・アルバム・システム”.インタラクショナル2001,2001,p.143-144.
- (2) Ohmori, Y., et al. “An XML Based Multimedia Data Acquisition and Retrieval With Wearable Computers”. IWSAWC 2001, 2001, p.272-277.



野々村 克彦 NONOMURA Katsuhiko  
研究開発センター 知識メディアラボラトリー研究主務。XMLベースのナレッジマネジメントシステムの研究・開発に従事。Knowledge Media Lab.



大盛 善啓 OHMORI Yoshihiro  
研究開発センター マルチメディアラボラトリー。状況依存型インタフェースの研究・開発に従事。Multimedia Lab.



山本 晃司 YAMAMOTO Koji  
研究開発センター マルチメディアラボラトリー。映像構造化技術の研究・開発に従事。Multimedia Lab.