

XML を用いた非文字資料デジタルアーカイブの構築

穂苅 正浩[†]

宇津木 敏人[†]

能登 正人[†]

神奈川大学工学部電気電子情報工学科[‡]

1. はじめに

近年、博物館や美術館の文化財や歴史資料、さらには各種映像資産などをデジタルデータで構築するデジタルアーカイブという言葉が使われるようになった [1]。従来の紙ベースの資料や写真は、年月により劣化し、情報が失われてしまうが、資料をデジタル化し、記録することにより、半永久的に劣化せずに保存可能な情報資料にすることができる。デジタルコンテンツのアーカイブ、つまりデジタルミュージアムの実現には、保存すべき資料を収集し、コンテンツの利用性を保ちながら蓄積・組織化し、利用者に提供することが求められる (図 1)。

デジタルコンテンツの保存のためには、コンテンツの構成方式、保存の履歴などコンテンツに関する情報、すなわちメタデータが必要とされる。本研究では XML (eXtensible Markup Language) によるメタデータを用いたアーカイブ構築のためのシステムを提案する。

2. メタデータ

メタデータとは、「データについての情報を記述したデータ」と定義される。コンテンツのデジタルアーカイブを実現するには、発見のための有用な属性以外に、コンテンツの構造や管理情報、保存の履歴、利用のための条件といったさまざまな情報が必要であり、これらはすべてメタデータである。ディ

ジタルデータにメタデータを埋め込むことにより、情報を持ったデータとなり、膨大な量の分散したデジタルデータの整理、さらには目的とするデータの発見のための検索が容易になる。

3. XML

3.1 XML による記述の利点

XML は文書やデータの意味や構造を記述するためのマークアップ言語の一つである [2][3]。マークアップ言語とは、「タグ」と呼ばれる特定の文字列で地の文に構造を埋め込んでいく言語のことである。他のマークアップ言語である HTML との違いは HTML は、Web ブラウザ上で文字を強調したり、段落をつけたりするタグが取り決められている。これに対して、XML はデータを作成する人がタグを自由に取り決めることができるようになっている。これを利用して、画像データについてのメタデータを属性とその値で表現することができる。つまり、XML による記述は人間がみてもわかりやすく、計算機でも処理しやすい。また自由度が高いのでデジタルコンテンツのメタデータとして扱うのに便利である。また、XML 形式という共通のフォーマットでデータを記述するので、他のシステム間での互換性も確保することができる。

3.2 XML データファイル形式

本研究で用いる XML ファイルのデータ構造と、本データ構造に従った XML ファイルの例を図 2 に示す。これは、'imagelist' タグをルートとした階層構造を持つ XML ファイルで、ルートタグの下に画像データを記述する 'filename' タグ、さらに下に画像の属性である 'title'、'creator'、'date'、'description' タグが定義されている。これら 4 つのタグはインターネット上の情報資源のための規則として利用されている DCMES (Dublin Core Metadata Element Set) による。さまざまな情報資源の記述のために共通で使うことのできる基本的な項目である。

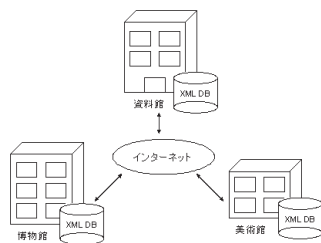


図 1: XML を用いたデジタルアーカイブの共有

Construction of Nonwritten Digital Archive using XML

[†] Masahiro Hokari, Toshihito Utsugi and Masato Noto

[‡] Department of Electrical, Electronics and Information Engineering, Kanagawa University

4. 提案システム

本研究では、デジタルコンテンツを収集するためのインターフェースと画像データ検索の部分を重

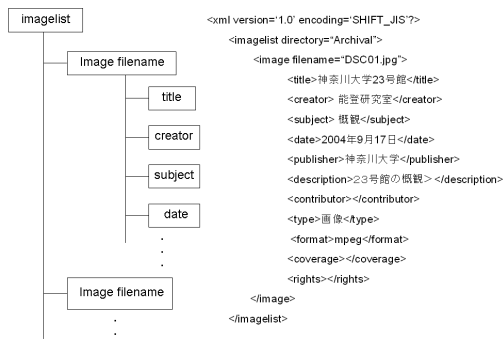


図 2: XML データファイルの文書構造と例

図 3: 非文字資料登録フォーム

視し、コンテンツの属性を記述することで自動的に XML に変換できるシステムを構築する。ここで扱う対象となる「非文字資料」とは、画像、映像等の文字になっていない資料のことである。このシステムにより利用者がメタデータを意識することなくデジタルコンテンツをコンテンツサーバーにアップロードすることができる。さらに検索機能により、ユーザの欲しい情報を検索でき、全国に拡散している非文字資料を体系的に組織化して閲覧することが可能となる。

5. システムの概要

5.1 データ入力システム

利用者はデータ登録時に、非文字データ、さらにはデータの属性を記述する。この属性は DCMES による。これらを自動的にタグに埋め込むことにより XML 形式に変換し、これにより情報を付加した XML データファイルを生成する。図 3 に入力システム画面を示す。

5.2 画像検索システム

ユーザが探したい非文字資料のタイトル、もしくは作者や年代などの情報やその資料に対する感想な

検索結果 (3件表示)

画像	タイトル	作者	日付	要約	その他
	23号館・撮観	穂刈	2004年6月17日	23号館の撮観	詳細情報
	23号館・入り口	吉岡	2004年12月17日	23号館・入り口の様子	詳細情報
	23号館・能登研究室 有川	有川	2004年9月17日	23号館・能登研究室の様子	詳細情報

図 4: 検索結果

どの情報を指定することにより、XML データベースに入力された資料を探し、ユーザに提示する。検索結果を図 4 に示す。

6. 考察

本研究では、仕様や形式の異なる非文字資料を XML 形式に統一しデータベース化することができた。これにより、データの操作性が向上し、デジタルアーカイブとしてどの環境からでもユーザが利用できるものとなった。

7. おわりに

本研究では、XML を用いて非文字資料のデジタルアーカイブの構築を行なった。今回は、デジタルアーカイブのインターフェース部分だけの紹介に留まったが、今後は、データのやり取りを P2P (peer to peer) により行ない、サーバーに掛ける負担を軽減させることや、既存の商用データベースとの併用、エージェント機能を導入し、管理者負担を軽減させ、ユーザにとって使いやすいものとするなどが挙げられる。

謝辞

本研究の一部は、神奈川大学 21 世紀 COE プログラム「人類文化研究のための非文字資料の体系化」の補助金による。

参考文献

- [1] <http://www.jdaa.gr.jp/>
- [2] 多田 裕人, 遠藤 教昭: オープンソースを用いた XML によるシラバスデータベースの試作, 情報処理学会研究報告, Vol. 18, No. 3, pp. 47-53 (2004).
- [3] 杉本重雄: デジタルコンテンツのアーカイブとメタデータ, 人工知能学会誌, Vol. 18, No. 3, pp. 217-223 (2003).