

Ⅱ-13 XMLのイントラネット土木技術情報データベースへの適用

石倉 正英

Masahide Ishikura

山本 隆彦

Takahiko Yamamoto

大岩 忠男

Tadao Oiwa

畑 久仁明

Kuniaki Hata

【抄録】論文や各種技術資料の社内共有や共有文書形式の標準化を目的とし、イントラネット上でのXML文書を介した技術情報DBの活用、DBとしてのXML文書の活用を図った事例を紹介する。XML文書自体がDB的な仕様を内包している為、専門のDBソフトを使わなくても簡易に検索等が可能であるなどのXMLの特徴を活かし、インデックスの部分をXML文書化して技術情報を管理・活用する方法を試行した。また、XML文書の特定のシステムへの非依存性、スケーラビリティの実証も行った。

【キーワード】XML、Web、イントラネット、データベース、技術情報

1. 背景と概要

近年、イントラネットインフラの充実化が進み、各部署に分散している技術情報を共有し戦略的に活用しようというニーズが生まれてきている。一方、建設省のCALS/E Cに代表される文書形式の標準化により、SGMLの流れを汲むXMLという文書形式がインターネット時代の標準的文書として脚光を浴びてきている。

XMLとは、特定のコンピュータ処理システムに依存せず、電子データのプログラム間での交換と永続的な保存に視点を置く仕様の作り方を定義した規格である。HTMLと組み合わせることにより、httpでの利用もできるため、イントラネット上での標準的文書と位置付けられる。

こういった背景を受け、本研究ではXML文書の特徴を活かし、イントラネット上に分散している技術情報データベース、特に对外発表技術論文のデータベースへの適用を試行し、XML文書による分散型技術情報データの管理手法を検討した。

2. XMLとは¹⁾

XML 1.0 の仕様は 1998 年 2 月に W3C(World Wide Web Consortium)の勧告として公開された。

XMLとは、その親となるSGMLと同様、文書の自動処理を容易にするために、文書の論理構造と物理構造を定義する手段を提供する「汎用的なデータ記述言語」の規格である。SGMLとの違いは、XML文書およびその関連情報をインターネット上で容易

に交換できることに主眼をおいて開発された点にある。すなわち、XMLはSGMLのサブセットであり、SGMLにおいて通常は使わない仕様を削除し、軽量化された規格である。実用面においては、インターネット技術に裏付けられた実装規格の整備が進んでいる。HTMLに代表されるインターネットでのデータ配信技術とSGMLの理論面が融合した「交換および共有される電子データ」のための規格である(図-1)。

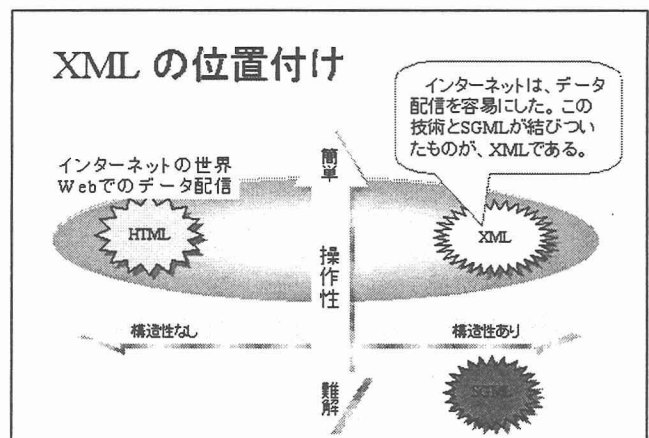


図-1 XMLの位置付け

以下にその概要と特徴についてまとめる。

- ・XMLとは、eXtensible Markup Language(拡張可能なマーク付け言語)の略で、インターネット技術に裏付けられた「交換及び共有される電子データ」のための規格である。
- ・XML文書はテキストデータであり、さらに、人間が直接に読めるような形式である。
- ・XML文書はその文章、単語、データにタグを付加す

ることによって、文書を構造化できる。

- ・XML 文書は冗長であるが、データ交換での利用において、CSV よりも優る点が多い。
- ・SGML のサブセットであり、XML 文書は必ずSGML 規格に適合する。

3. 従来のCSSと Web クライアント／サーバ

XML 文書を用いたインターネット技術による分散型の技術情報データ管理システムを論じる前に、システムの基本部分となる「Web によるクライアント／サーバシステム」(以降 WCSS と呼ぶ)の利点、特に本研究のテーマに即した利点について触れる。

インターネット技術の発達により、イントラネットという形態で、企業内でも Web をインターフェースとした業務システムが標準化されつつある。本研究も例に漏れずこの WCSS をシステムの基盤として考えているが、ここで「従来のクライアント／サーバシステム」(以降 CSS と呼ぶ)との違い、優位な点について以下に述べる(図-2)。

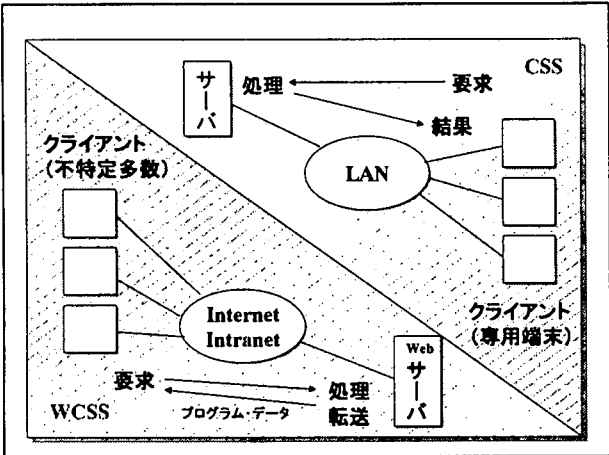


図-2 CSSとWCSS

従来型のCSSの特徴としては、

- (1)専用の端末、インターフェースによる操作
 - (2)セッション中のネットワークの独占
 - (3)全処理をサーバ側で行う
- 等が挙げられる。

一方、これに対するWCSSの特徴としては、

- (1)Web インターフェースによる操作
- (2)クライアントが不特定多数、かつ同時アクセスが可能
- (3)操作ごとにセッションを張りなおす
- (4)クライアント側でも処理を行うことが可能

等が挙げられる。

この特徴から、本研究に WCSS を採用した場合の利点について以下に論じる。

利用者が多数かつ同時にアクセスする状況を考慮すると、(3)操作ごとにセッションを張りなおす、即ち、ネットワークの断続的な使用が効果的である。何故なら、他のネットワーク処理への負荷の軽減が図れるとともに、閲覧時にはネットワークを通じた処理が発生しないという操作感の軽さが得られるからである。

Java 等を用いたクライアント側で展開されるアプリケーションの使用は、(4)クライアント側でも処理を行うことを可能にする。これは、サーバへの処理の集中を和らげ、昨今高速化してきているクライアントマシン（主に WindowsPC）の能力を最大限に利用できるという利点がある。

また、Web をインターフェースとすることにより、インターネットでの検索の感覚を活かすことができ、操作感の統一性が得られる。

4. 適用事例・技術論文データベース

それでは次に適用事例として本研究で取り上げた技術論文データベースについて触れる。

本研究で取り上げた技術論文データベースは、当研究所における対外向け発表論文の情報をデータベース化したものである(図-3)。

図-3 技術論文データベース

このデータベースは、部署内の管理情報としての価値もさることながら、他部署との情報共有という視点においても価値のある情報である。また、論文の電子化が進むに連れ、紙としての保管よりも、電子データとしての保管がなされるようになってきており、その保管場所や論文そのものの情報とのリン

クも考える必要がある。

4. 1. データのXML文書化

データをXML文書化するにあたって、最初にデータの構造化を行った(図-4)。全体を、論文情報/作成情報/発表情報/管理情報/本文情報の5項目に分け、各情報をツリー状に分類した。論文集に関しては主に投稿先で発行したものと、部署内でまとめ発行したものの2種類が存在しているので、それぞれを発表情報と管理情報とに分けて構造化を行った。

論文情報	論文名
作成情報	著者名・所属など
発表情報	投稿先・論文集情報など
管理情報	管理書・保管場所など
検索情報	キーワード・アブストラクトなど
本文情報	電子データ保管場所(URL)・ファイルタイプなど

図-4 データの構造化

これをもとに、XML文書化したものが図-5である。和文と英文が存在するものに関しては、lang属性を用いて分類した。著者情報に関しては、主著・共著の別を著者順位属性によって、著作権の有無を著作権属性によって、また、当社の社員か否かを東亜建設属性によってそれぞれ分類した。キーワードに関しては、関連URLがあるものについては、引用URL属性を付けて、そのURLを記入できるようにした。本文情報に関しては、ファイルタイプと保管URL、およびその容量を持てるようにして、ダウンロードの補助的役割を担える情報を付加できるようにした。

このXML文書に対するDTD(Document Type Definition:文書型定義²⁾)を図-6に示す。

4. 2. XML文書によるデータベースイメージ

XMLを用いた技術論文データベースの閲覧を考えた場合、そのデータベースのイメージは2通り考えることができる。即ち、一つの論文に一つの情報データファイルという図書カード的なイメージと、複数の論文を一つの情報データファイルにまとめた一覧表的なイメージである。本研究では、あくまでもデータ本体は一論文に対し一情報ファイルという

図書カード的なイメージで構成し、閲覧時に目的によって表示方法を選択できるという方法をとった。

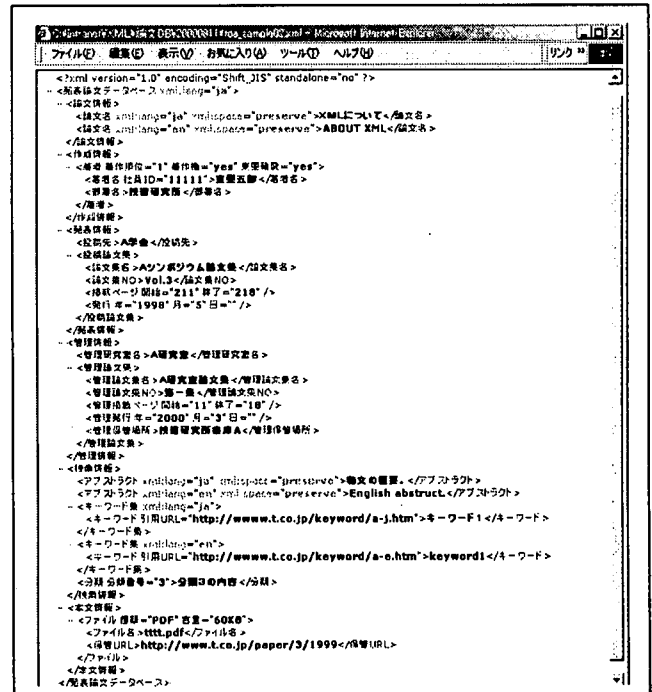


図-5 XML文書

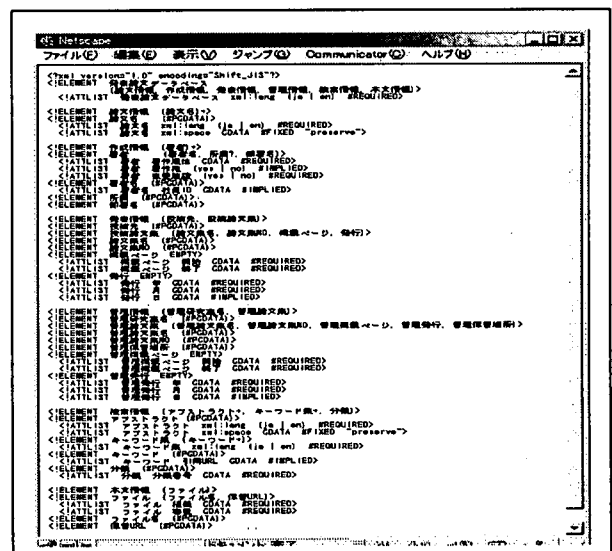


図-6 DTD

これは、文書情報が複数サーバにまたがっている場合に有利である。即ち、担当部署の管理下にある場所に情報データを置くことにより、データの分散管理が可能になり、登録時や閲覧時の負荷分散にも好影響を与えると考えられるからである。

もっとも、検索系を考えたときにあちこちに分散したXML文書を全文検索かけて行くような状況を予想すると、この散漫な管理方法は必ずしも有効とはならない。この場合は、検索系に関連する情報のみを別個に扱う(例えばデータベースエンジンを

用いるなど) 手法が有効であろう。インターネットの URL 参照の技術を利用したキーワードの一括管理や、データベースエンジンを用いた方法など、今後模索する必要があるだろう。

5. 検索・表示系

本研究で用いたシステム構成としては、Web サーバに Sun の UNIX ワークステーションと Compaq の NT ワークステーションを用い、クライアントには、Windows98 上の IE5.5 を用いた。

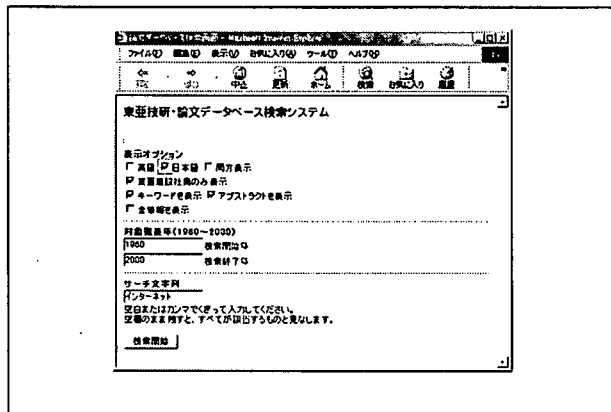


図-7 検索画面イメージ

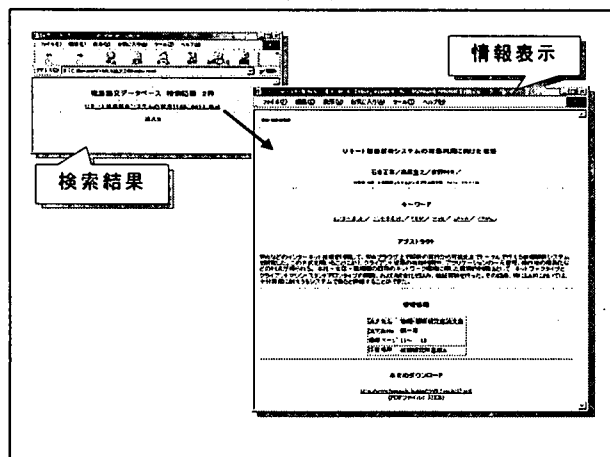


図-8 表示例

図-7 に検索画面イメージを示す。XML 文書の特徴として、タグごとのデータ管理に卓越しているので、表示・非表示の切り替えがしやすい。検索は DOM (Document Object Model³⁾) を用いて行っている。

図-8 に表示例を示す。このように論文情報データが複数サーバにまたがっている場合でも、同一のスタイルで表示可能である。また電子化されている本文に関しては、必要に応じてそのページからダウンロード可能である。図-8 の例のように PDF 形式のファイルであれば、ダウンロードせずとも、Web 上

で表示し、高品質な印刷物として即座に入手することが可能である。

6. まとめと今後の展開

本研究を通じ、XML の利便性として実感できたこととしては、

- (1)データそれ自体の内容を読むことができる
- (2)データベースエンジンを起動することなどなく検索結果を表示できる (操作感の軽さ)
- (3)分散されたデータを仮想的に統合できる
- (4)特定のシステムへの非依存性

などである。特に(3)に関しては、インターネットの技術をフルに活用でき、また、論文それ自体をダウンロードという形態で即座に入手できるということは、イントラネットというクローズした環境下では利便性が非常に高いと言える。

本研究はまだ開始されたばかりであり、今後の展開や模索/試行の枚挙に暇はないが、主なところをまとめておくと、

- (1)データの登録インターフェースを含めた一つのシステムとしての開発
- (2)URL 参照の技術を用いた分散管理型のデータベースシステムの模索
- (3)データベースエンジンを検索系に取り込んだシステムの模索
- (4)データの転送部を機能として盛り込んだ、ネットワークに負担をかけないクライアントアプリケーションの開発
- (5)他の技術情報データや数値解析分野への応用等である。

インターネット技術の根幹である「リンク」即ち「URL 参照」の技術をキーにこれらの模索・試行・開発を進めていこうと考える。

参考文献

- 1) 山本隆彦、桜井悦雄：建築生産情報の XML 化に関する研究-2、日本建築学会 第 16 回建築生産シンポジウム 2000
- 2) Robert Eckstein：XML デスクトップリファレンス、オライリー・ジャパン、2000.3.
- 3) Msdn online Web Workshop：<http://www.asia.microsoft.com/japan/developer/workshop/>