

XML (eXtensible Markup Language) は、データの意味を表すタグを目的に応じて定義することが可能な、国際標準のデータ記述言語である。タグにより、必要なデータをコンピュータが自動的に判定することが可能なため、多くの応用が考えられる。特に、共有/交換データをXMLで記述することは有効な手段であり、標準に基づきXMLデータをプログラムで処理するための各種のツールが提供されている。

XMLの応用は非常に広範囲にわたるが、ここではナレッジマネジメント、コンテンツ管理、システム間/企業間データ交換の三つの主要な応用を概観する。これらの分野では、様々な事例が報告され成果が出つつある。これらを通して、XMLの基本的長を述べるとともに、今後の応用の方向性について考察する。

Extensible Markup Language (XML) is an internationally standardized format to describe data. Users can define tags to represent semantics of data. Computers can recognize data by using the tags, so many applications can be designed in XML. In particular, XML is a highly effective format for shared or exchanged data. Many tools are available for handling XML in programming languages.

In this paper, we describe some typical applications of XML including knowledge management, contents management, and data exchange among systems and/or companies. Many efforts are being made to use XML in these fields, and some remarkable results can be seen. After reviewing these results, we outline the basic merits of XML and discuss future trends in these fields.

XMLとは

背景

XML⁽¹⁾⁽²⁾は、現在広く用いられているHTML(HyperText Markup Language)などと同様の、タグ付き構造化データ記述言語の一つである。

1989年にTim Berners-Leeにより開発されたHTMLは、インターネット上の標準記述言語として広く使われている。また、文書記述としてのSGML(Standard Generalized Markup Language)も、米国国防総省などを中心に使われている。HTMLの拡張と、複雑なSGMLの仕様の単純化という二つの流れが、98年にXML1.0としてW3C(World Wide Web Consortium)により標準化された。様々な製品が開発・提供され、多くの分野への適用が試みられて成果を上げつつある。

XMLの概要(囲み記事参照)

XMLは、タグ付きの構造化データ

である。各タグはそれぞれのデータ項目の意味を示し、文書の構造はタグの入れ子により表現することが可能である。テキスト形式のデータであるため可読性が高く、コンピュータによる柔軟な処理を実現することが可能になる。また、複数の目的でデータを共有する場合の記述方法として特に有効である。

同様のタグ付き記述方法であるHTMLと比べて、次のような特徴がある。

- コンテンツと表現形式の分離

HTMLでは、あらかじめ決められたタグで表示方法を指定する。

- タグの定義が可能

XMLでは、用途に応じてタグを定義することが可能である。

- 柔軟なリンク

HTMLでは、文書から文書への一方方向のリンクであるが、XMLでは、文書の一部へのリンクや双方方向のリンクなどが可能である。

現在、多くの製品が開発・提供され

る一方、スキーマ定義言語などの関連標準が検討されている状態であり、ダイナミックに技術が変化している状況である⁽³⁾⁽⁴⁾。

XMLの利用に関しては、現在次の三つが主要な分野である。

- (1) ナレッジマネジメント
 - (2) コンテンツ管理
 - (3) システム間/企業間データ交換
- 以下、これらについて個別に紹介する。

ナレッジマネジメント(KM)

KMとは⁽⁵⁾

組織内に存在する知識には、形式知と暗黙知があると言われている。形式知とは言葉で表現できる知識を意味しており、文書形式などで存在している。一方、暗黙知は言葉や文章で表現しにくいもので、ノウハウ、観点、思いなどが挙げられる。

KMシステムは、暗黙知-形式知変換を伴う知識創造のスパイラルプロ

セスを、うまく管理し支援する仕組みであると言える。したがって、狭義のKM(米国流)のように、文書管理を中心に共有・再利用を支援するということにとどまらず、知識創造のための支援も強く求められている。

■ KMにおけるXMLの役割

組織内の情報には、図1で示すように、様々なものが存在する。代表的なものに、基幹システムで利用されている構造が明確な情報がある。一方、その対極にあるのが、Web上にあるような平テキスト情報である。しかし、組織内にある情報の多くは、日報情報、議事録、設計情報などのように、おおまかには構造はあるが、詳細に構造が定義できない、いわゆる半構造データである。

これら半構造データを構造化した枠組みに体系化してしまうと、自由度に欠けたものになり利用しづらいものになってしまう。また、これら情報を平テキスト情報として扱おうと、せっかくのタグ情報や構造情報があるにもかかわらず、それらを有効に利用することができないという問題点がある。

XMLはまさに、半構造データに適したデータ表現であると言える。以下、XMLをベースにしたKMのメリットを整理してみる。

- (1) 構造の変化に強い 情報の構造は日々変化するため、情報に対する属性追加、更新が自由に行える。
- (2) 操作できる プログラムなどから文書操作が容易になり高精度の検索、分析、分類などが可能になる。
- (3) 異種データ(モデル)の統合 複数の情報源を組み合わせた複合文書の作成が容易になる。また、文書、数値、メタデータ付き画像や音声などの異種データの統合も容易になる。
- (4) 情報の交換が容易 企業内の知識フロー(ナレッジチェーン)

XMLの概要

XMLの具体的な記述を下図左に示す。比較のために同じ内容のHTMLを右に示した。図は、注文リストの記述例である。XMLでは、目的に応じてタグを定義することが可能であり、データ項目の意味を表すのが一般的である。このタグにより、人間にもコンピュータにとっても可能性の高いデータ記述方法となっている。

HTMLと比べると、XMLは次の特長を持っている。

(1) コンテンツと表現の分離 XML

のデータには、表現の指定はない。XSLTと呼ばれる変換手段に表現を指定することで、表示データ(HTMLなど)を生成可能である。

(2) タグの定義が可能 上述のとおり、データ項目の意味を表すタグを定義することで、可能性の高いデータを記述できる。

(3) 柔軟なリンク 文書の一部へのリンクや双方向リンクの定義などが可能である。

```
<OrderList>
  <OrderNumber>0001</OrderNumber>
  <CustomerName>Taro Yamada</CustomerName>
  <Order>
    <ProductName>Note book PC</ProductName>
    <Quantity>100</Quantity>
    <ProductName>CD Drive</ProductName>
    <Quantity>50</Quantity>
  </Order>
</OrderList>
```

XMLの記述例

```
<HTML>
<HEAD>
<TITLE>Order List</TITLE>
</HEAD>
<BODY BGCOLOR="#FFFFFF" TEXT="#000000">
<TABLE BORDER="1"><TR>
<TD COLSPAN="100%">Customer:Taro Yamada</TD>
</TR>
<TR> <TD>製品名</TD> <TD>注文数</TD> </TR>
<TR> <TD>Note book PC</TD> <TD>100</TD> </TR>
<TR> <TD>CD Drive</TD> <TD>50</TD> </TR>
</BODY>
</HTML>
```

HTMLの記述例

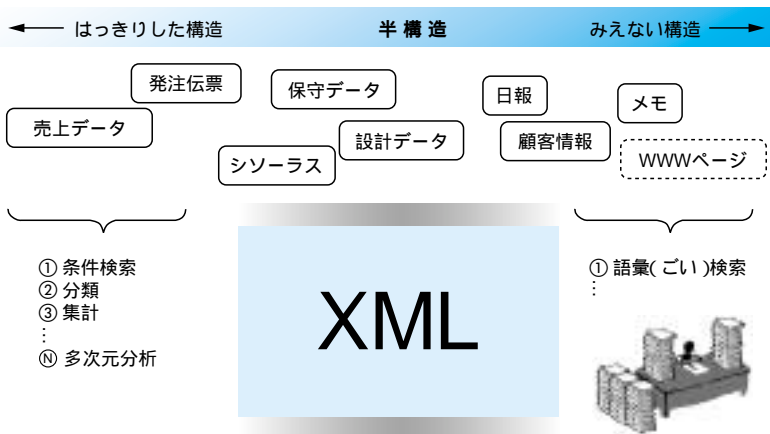


図1. 組織のKMで扱う情報 日報などのデータをXMLデータとすることで、様々な用途での柔軟な利用が可能となる。

Information handled by knowledge management (KM) in organization

環境が、標準交換フォーマットとして実現できる。

■ XMLによるKMの技術ポイント

- デジタル文書統合技術
- 組織内には紙ベースの文書も数多

く存在する。これらのデータをどのように扱うかが問題になっている。OCR(光学式文字読取り装置)などを用いてデジタル化を行い、それらを有効に活用できるプラットフォームが重要になる。つまり、文書のデジタル化から、ワークフローに沿った処理を可能とする各種コンポーネントの確立が必要である。

・自動タギング技術

ワードプロセッサなどで作成されたデジタルドキュメントから、そのスタイルファイルに沿ってXMLのタグを生成する技術がある。

また、あるタグに挟まれた平テキスト情報から、自然言語処理、テキストマイニング技術を使い、概念、キーワードなどを抽出して新たにタグ情報を生成する技術が研究されている。

・データ格納・検索技術

XMLデータを効率よく格納し、高速に検索するためのデータ管理技術が必要になる。1文書を1ファイルで格納する簡易な方式もあるが、大量の文書には対応できない。

現在の主流は、RDB(Relational DataBase)にXMLデータを格納する方式である。RDBに格納できれば、基幹系データと統合して扱うことができるメリットがある。しかし、RDBは表形式のデータモデルであるのに対して、XMLは木構造のデータモデルである。RDBにマッピングする方法として色々提案されているが、検索効率が悪化したり、データ構造が固定化されたりする欠点がある。

XMLの構造モデルにあったXMLネイティブなDBも市販され始めている。

・分析技術

数値情報は、RDBやデータウェアハウスに格納され、多次元分析など、いわゆるOLAP(On-Line Analytical Processing)が行われている。自然言語処理、テキストマイニングなどを組み合わせることによりいろいろな視点からの分析が可能になる。われわれ

はこれをテキスト版OLAPと呼んでいる。

・XMLによるKMの適用領域

KMは現在、“研究部門における知識共有”、“コールセンターなどにおける顧客知識共有”、“品質管理のための知識共有”など幅広く実施されている。特に、XML化しやすい技術情報(特許、論文など)やクレーム・アンケート情報などから立ち上がり、多くの情報がXML化されていくと予想される。また、設計情報に関して、インターネット上での協調設計の枠組みの中に組み込まれていくと思われる。

■ コンテンツ管理

■ 企業におけるコンテンツ管理

企業や組織において扱うべきデータの量は、インターネットの普及とともに近年急速に伸びており、9か月で2倍になるとの報告もある。企業活動において、データの維持・管理費用は大きな課題である⁽⁶⁾。

一方、インターネットや携帯電話など多様化するアクセス手段を活用し、自社の情報やサービスを積極的に顧客にPRすることは、ビジネスを有利に進めるうえで不可欠なこととなっている。

このように、膨大なデータを、多様化するアクセス手段に合わせて、顧

客や従業員が、効率的に利用できるようにすることは、企業にとって極めて大きな課題であると同時に、これによる業務の効率化、付加価値の増加など、他社差異化のための大きな力ともなりうる。

一方、政府においては、電子政府設立に注力しており、ここでもコンピュータによるデータの流通を実現することが大きなポイントである⁽⁷⁾。

■ コンテンツ管理とXML関連技術

XMLを利用して、企業内のデータを管理することにより、様々なアクセス手段に応じた提供が容易となる(図2)。

XSLT(XSL(eXtensible Stylesheet Language) Transformations)はもっとも基本的なXMLデータの変換技術である。あるスキーマのXMLデータを別のスキーマのXMLデータに変換したり、HTMLに変換することが容易に可能である。

XMLデータに関するDB機能も重要な技術である。前章で述べたように様々な方式があり、コンテンツ管理の側面からは目的に応じた方式の選択が必要となる。

ワークフローなど、各種ソフトウェアのI/F(InterFace)としてのXMLも注目されている。XMLを使うことで、ソフトウェア間でのデータの相互

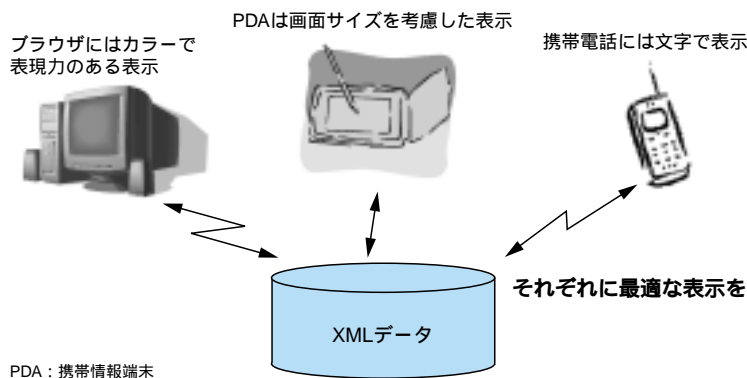


図2 .XMLによるコンテンツ管理 コンテンツを出入力デバイスに合わせて容易に変換・出力することができる。

Contents management using XML

利用が容易となり、企業内データの効率的な活用が可能となる。

■ システム間 / 企業間データ交換

■ 企業環境の変化とシステム要件

グローバル競争や低価格化競争など、厳しさを増す一方の状況のなかで、多くの企業が経営方針の見直しを迫られている。コアコンピタンスの追求と、最適なパートナーとの連携によるユーザーニーズへの対応は、有効な手段であるとされている。

企業における情報システムも、このような企業活動を支える手段としての利用が試みられている。特に、社内外のシステム間連携は、もっとも基

本的かつ重要な機能として、実現への取組みが行われている。古くからのシステム間 / 企業間データ交換に加えて、ビジネス環境の変化に応じて、インターネット上で最適なパートナーと動的に連携を実現する機能がポイントとなっている。

■ データ交換における XML

データ交換においてもっとも基本的かつ重要なことは、データを互いのシステムが理解可能なことである。

XML ではタグにより自システムに必要な項目を探ることができるので、順番が違っていたり、項目数に違いがあってもデータ交換が可能であり、有効な解決手段となりうる⁽⁸⁾。実際の適用に際しては、次のような技術的ポイン

トがある(囲み記事参照)。

・ タグの標準化

データ交換に使うタグの名称とその定義について、業界ごとの標準化が活発に行われている。

・ 交換プロトコル

交換手順の標準化は、見積依頼 見積回答 発注 発注確認 ……といった、一連の手続きを標準化する。タグと同様、業界ごとの活動が中心である。

・ その他

業務取引を行ううえで十分な信頼性を確保するため、セキュリティ技術や、交換データの圧縮技術、XML データと基幹系 DB との間での相互変換や基幹系処理を連動させる技術などが研究・開発されている。

■ XML 適用例

2005年には、企業間の電子商取引の国内規模が110兆円を超え、その多くがXMLを使ったものになると予想されている⁽⁹⁾。

ロゼッタネット(RosettaNet)⁽¹⁰⁾は、代表的な取組みの一つであり、98年に米国でパソコンの組立ての効率化を目指してスタートした。部品の調達、物流、決済などの手順をPIP(Partner Interface Process)として定め、それに基づきシステム間で連携を行う。自動車や金融など、ほかの業界もロゼッタネットでの活動を計画している。

ebXML(electronicbusiness XML)⁽¹¹⁾は、中小企業での取引も含めた検討を行っている。PIPに相当する取引手順の取決めや、企業情報の登録・検索などに関する標準も検討している。ebXMLとロゼッタネットは、お互いに歩み寄りの方向にあり、企業間データ交換のスタンダードになる可能性が高い。

■ Web サービス⁽¹²⁾

企業間でのデータ交換を積極的に活用し、インターネットを通じてもっとも優れたパートナーと連携し、競争力

XMLの標準

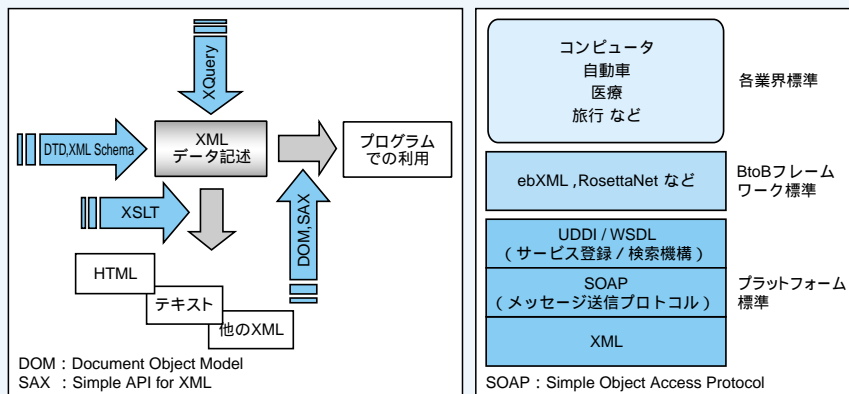
XMLに関連する標準はとにかく多い。「XML自身はW3CでXML1.0として標準が定められているのになぜ？」という疑問が生ずる。

現在、XMLに関連した標準活動が数多く行われているが、大きく二つに分けることができる。

一つは、XMLデータを処理するための基本的な技術に関するものである(下図左)。スキーマの定義や変換、プログラムでの取込みなどに関する技術

の標準化が行われている。また、データとしてもモバイルでの利用を考慮したWAP(Wireless Application Protocol)など、プリミティブなレベルでの標準化が検討されている。

もう一つは、企業間データ変換のための標準化(下図右)で、メッセージのやり取りの仕組みから業界ごとのXMLを用いた記述方法(どのようなタグを使うか)まで、多くの標準化活動が行われている。



を獲得しようとするのが、Web サービスである(図3)。

各企業は、プロフィールをUDDI (Universal Description, Discovery and Integration)に登録する。企業自身に関する情報のほか、提供可能な製品/サービスに関する情報もWSDL(Web Service Description Language)なる記述形式で記述される。サービス利用者はUDDIからもっとも望ましいパートナーを検索することが可能となる。

Webサービスでは、各企業は優れたパートナーを検索できると同時に、サービス提供者としてビジネスチャンスの増加も期待できる。

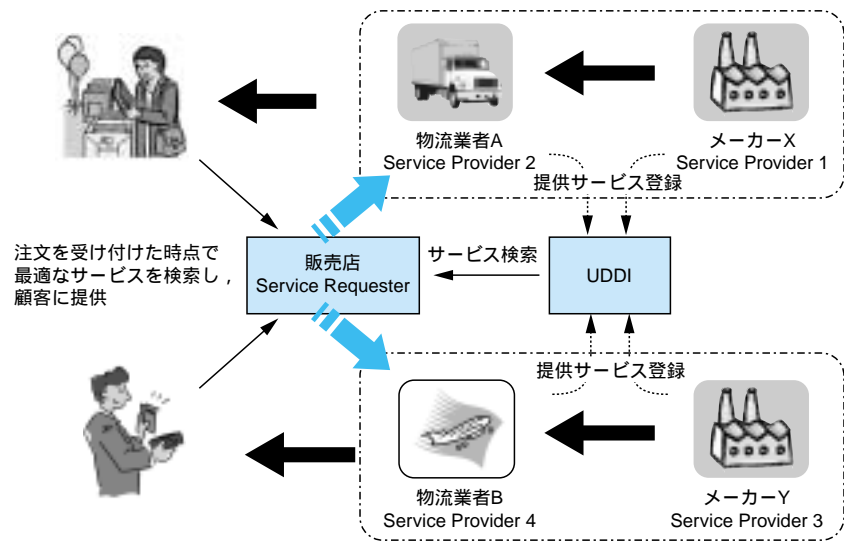


図3 . Web サービスのイメージ
Example of Web services

今後の展望

ここでは、XMLの概要とその応用について全体動向を概観した。最後に、今後の展望について考察し、まとめたい。

技術的にもっとも大きな課題の一つが、XMLデータのスキーマ記述の標準化である。DTD(Document Type Definition)が使われてきたが、データの型を定義できないなど多くの課題がある。これに対して、XML Schemaなるスキーマ記述言語が標準化されようとしている。しかし、非常に複雑な仕様であると同時に、理論的な記述能力の不足が指摘されている。スキーマ記述言語は、システム仕様に大きな影響を与えるため、その動向が注目される。

KMやコンテンツ管理での利用においては、業務の中で発生したデータをいつ、どのようにXMLデータとするのが、もっとも大きな課題であろう。前述した自動タギングの技術が、どの程度有効に機能するかがかぎとなる。

ビジネスの側面でのポイントは、セキュリティ確保の方式や業務レベルでのトランザクションの管理方法など、

商取引を安全かつ確実にを行うための方式の整備であろう。Webサービスにおいては、ビジネスモデルの検討が技術と同様に重要になるであろう。

以上、XMLについて技術及び応用に関する動向を説明した。XMLにより、システムに依存しないユニバーサルなデータが出現したことになる。XMLがビジネス活動において、データベースやワークフローなど、すべての応用の基盤として利用される日が遠からずやってくる。これをビジネスにいかに関活用するか、その実現方法について検討・評価しておくことが重要であろう。

文献

- (1) Abiteboul, S., et. al. Data on the Web : From Relations to Semistructured Data and XML. Morgan Kaufmann Publishers, 2000, 259p.
- (2) XML/SGML サロン . 標準 XML 完全解説 . 技術評論社 . 1998 , 354p .
- (3) Liefke, H. XMill : An Efficient Compressor for XML data, Proc. Of the ACM SIGMOD2000, 2000, p.153 - 154.
- (4) Widom, J. Data Management for XML : Research Directions, IEEE Data Engineering Bulletin, Special Issue on XML. 1999, p.44 - 52.
- (5) 南正名, ほか . ナレッジマネジメントの技術動向 . 東芝レビュー . 56 , 5 , 2001 , p.2 - 7 .
- (6) 喜連川優監修, ストレージ・マネジメント研究会著 . ストレージ・マネジメント . 日経BP企画 , 2001 , 222p .

- (7) 今村誠, ほか . 電子政府におけるXML利用技術の動向 . 情報処理学会誌 . 42, 7 , 2001 , p.654 - 658 .
- (8) 渡辺榮一 . XML 企業間連携システム . オーム社 , 2000 , 302p .
- (9) 電子商取引推進協議会 (ECOM) / アクセンチュア “平成12年度電子商取引に関する市場規模・実態調査” . 2000.
- (10) <http://www.rosettanet.org>
- (11) <http://www.ebxml.org>
- (12) 今村誠, ほか . Web Servicesによる動的な電子商取引の実現 . 情報処理学会誌 . 42, 7 , 2001 , p.643 - 647 .



原嶋 秀次
HARASHIMA Shuji

e-ソリューション社 SI技術開発センター SI技術担当主務。データベースシステムの研究・開発に従事。情報処理学会、電子情報通信学会、ACM会員。
Systems Integration Technology Center



末田 直道
SUEDA Naomichi, Ph.D.

研究開発センター 知識メディアラボラトリー研究主幹、工博。人工知能、ナレッジマネジメント、XMLデータベースの研究・開発に従事。情報処理学会、人工知能学会、情報知識学会、AAAI会員。
Knowledge Media Lab.