

XML オーバービュー

2005-07-13

XMLコンソーシアム・エバンジェリスト
日本ユニシス(株)

小林 茂

本日の内容

- **Web(WWW , World Wide Web)**の発展
 - ブラウズからユビキタスネットワークまで
 - HTMLからXML , Webサービスの導入
- **XML(Extensible Markup Language)**とは
 - マーク付けについて
 - XML , 関連規格などについて

インターネットの登場 普及

● World Wide Web

- Webブラウザで、テキストを含むマルチメディア情報を扱う
- 情報発信から、B2C、B2Bの取引へ
- Webページの記述言語

HTML(Hyper Text Markup Language)



Webブラウザでの表示とHTML

Webブラウザでの表示
W3Cのホームページ

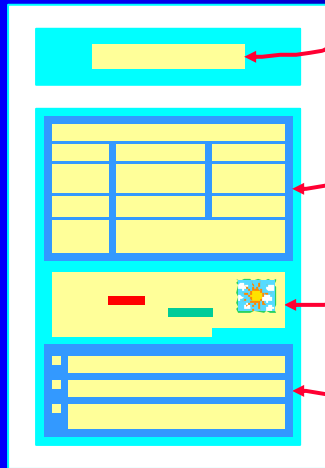
```

<html>
<head>
<title>タイトル</title>
<meta .../>
</head>
<body>
<h1><img .../></h1>
<h2>Leading .../></h2>
... 中略 ...
</body>
</html>
                
```

HTMLの記述

HTMLタグの例

HTMLでは、タグの名前と役割が決まっている

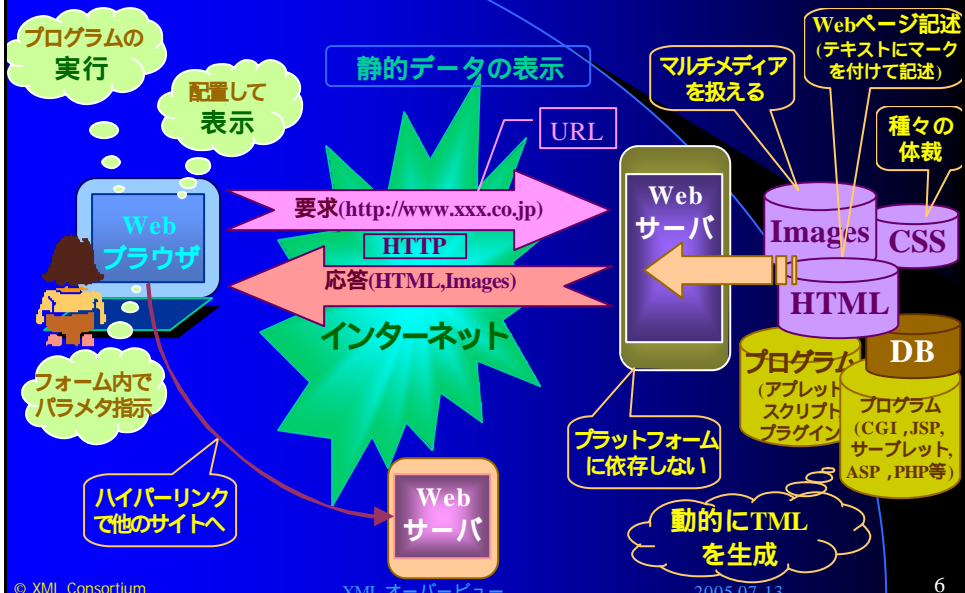


- h1** ヘッダ
- table** 表全体
- tr** 行(row)
- th, td** 表内セル
- p** 段落
- em** 強調
- a** ハイパーリンク
- img** 画像
- ul, ol** リスト
- li** リスト項目

Webページ

HTMLのタグ名と意味

Webページ表示の仕組み



複数アプリケーション連携

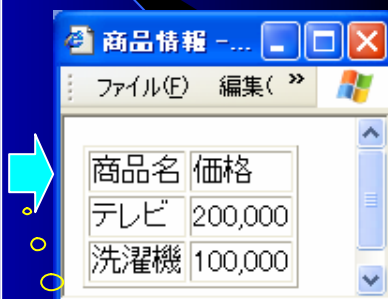


HTMLでの情報記述

table, tr, td, border
の使い方 意味は固定

```

<table border="1">
<tr>
  <td>商品名</td><td>価格</td>
</tr>
<tr>
  <td>テレビ</td><td>200,000</td>
</tr>
<tr>
  <td>洗濯機</td><td>100,000</td>
</tr>
</table>
  
```



Web
ブラウザの
表示機能

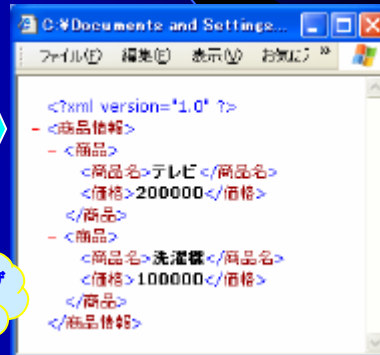
HTMLで規定

XMLでの情報記述

```

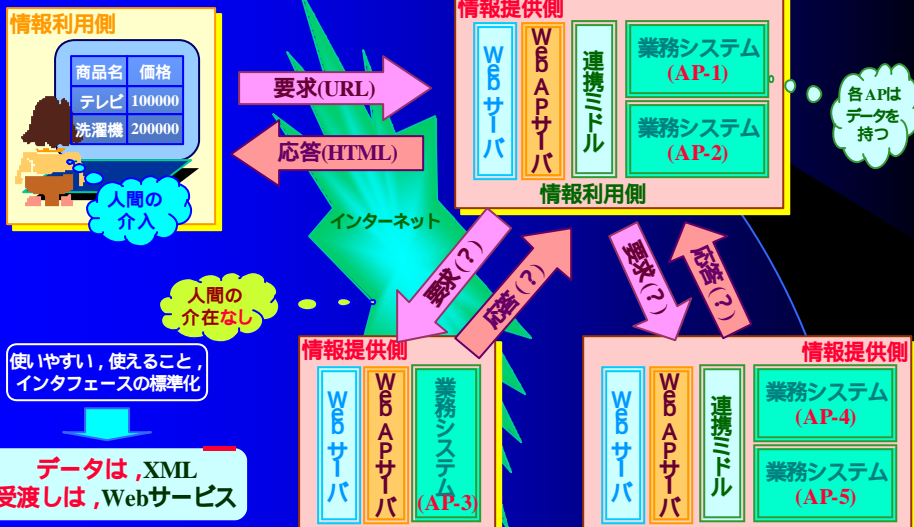
<?xml version="1.0"?>
<商品情報>
<商品>
<商品名>テレビ</商品名>
<価格>200000</価格>
</商品>
<商品>
<商品名>洗濯機</商品名>
<価格>100000</価格>
</商品>
</商品情報>
    
```

- スタイル情報の分離
- データの構造 意味が明確化
- プログラム処理が容易



Webブラウザ (IE) の表示機能

Webアプリケーションの連携

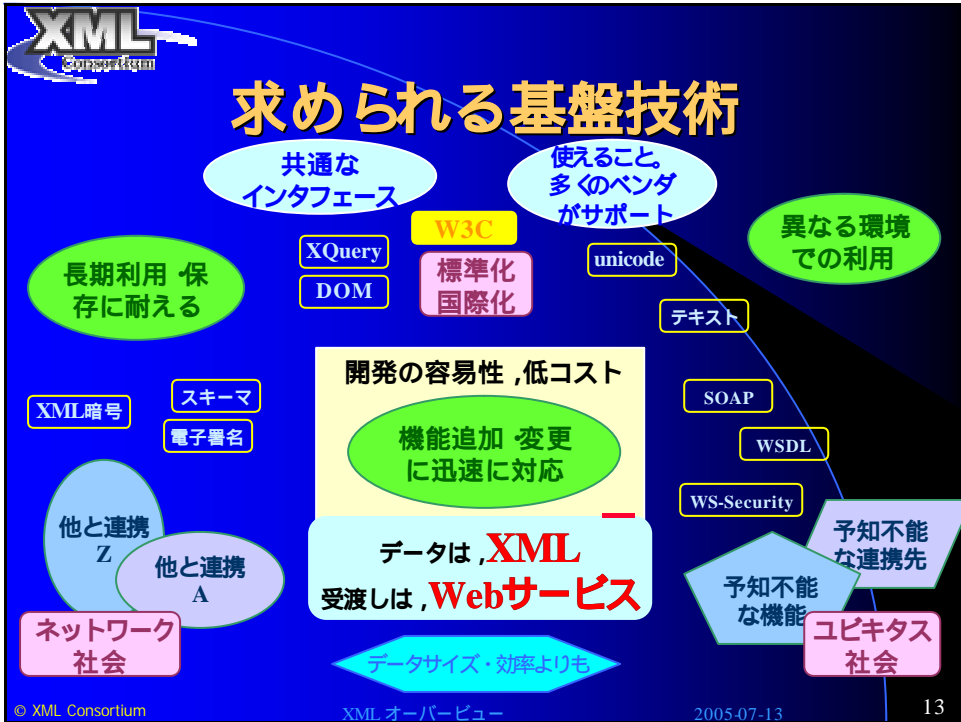


クライアントの多様化



ユビキタス社会へ





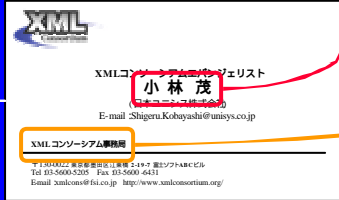
-
- ## XMLとは
- **XML 1.0 (Extensible Markup Language)**
 - W3C(World Wide Web Consortium)で規定するマーク付けの仕様
<http://www.w3.org/>
 - SGML(ISO/IEC規格)のサブセット
 - HTMLよりは**拡張性**がある
 - XML(SGML)は**メタ言語**, HTMLは1つの言語
 - 簡便性, 拡張性, インターネット利用を考慮
 - 関連する規格が多く規定されている
 - 多数ベンダが支持し, 安価なツールがある
- © XML Consortium XML オーバービュー 2005-07-13 14

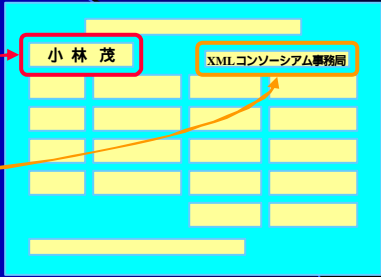
XML Consortium オフィス文書・ビジネス文書の利用

用紙

電子データ

伝票処理プログラム





印刷

用紙上での転記

人間が見て書き出す場合には、どのような形式でもOK

表示

PC上での切り貼り

人間が文字列を選択して、コピー&ペーストできるならば、どのような形式でもOK

取込


プログラムでの利用

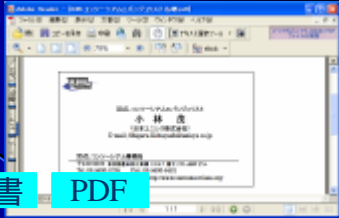
プログラムで入力側の名前の内容や会社の内容が分からなければならない!

プログラムでの利用に適したデータ形式が必要

© XML Consortium XML オーバービュー 15

XML Consortium 具体的な表示形式





ワープロ文書

PDF


文字情報、画像情報、図形情報、配置情報、体裁情報等の情報が、メーカーの固有データ形式にてファイル中に存在。

HTML

```

<div class="header">
  
</div>
<div class="main">
  <div style="font-weight:bold;">XMLコンソーシアムE
  <div style="font-size:large;font-weight:bold;">小林茂
  <div>(日本ユニシス株式会社)</div>
  <div>E-mail Shigeru.Kobayashi@unisys.co.jp</div>
</div>

```



© XML Consortium XML オーバービュー 2005-07-13 16

各種データ表現形式

注文伝票

品番	品名	単価	数量	金額
a01	鉛筆	200	3	600
b01	消しゴム	50	2	100

CSV形式

セパレータ方式

品番,品名,単価,数量,金額

a01,鉛筆,200,3,600

b01,消しゴム,50,2,100

固定長フィールド

a01鉛筆 0200000300600

b01消しゴム 0050000200100

HTMLデータ

```
<table>
<tr>
<td>品番</td>
<td>品名</td>
<td>単価</td>
<td>数量</td>
<td>金額</td>
</tr>
<tr>
<td>a01</td>
<td>鉛筆</td>
<td>200</td>
<td>3</td>
<td>600</td>
</tr>
<tr>
<td>a01</td>
<td>鉛筆</td>
<td>200</td>
<td>3</td>
<td>600</td>
</tr>
</table>
```

タグ方式

XMLデータ

```
<注文>
<項目>
<品番>a01</品番>
<品名>鉛筆</品名>
<単価>200</単価>
<数量>3</数量>
<金額>600</金額>
</項目>
<項目>
<品番>b01</品番>
<品名>消しゴム</品名>
<単価>50</単価>
<数量>3</数量>
<金額>600</金額>
</項目>
</注文>
```

データ形式として

CSV

2003-04-01, コバヤシシゲル, 小林 茂, 日本ユニシス株式会社, Shigeru.Kobayashi@unisys.co.jp, ...

最新日付 :2003-04-01
読み :コバヤシシゲル
姓名 :小林 茂
会社 :日本ユニシス株式会社
Eメール :Shigeru.Kobayashi@unisys.co.jp
...

カンマで区切る

出現順序が重要, 利用側で順序を知っていること, サイズは小さい

固有形式

区切り文字等の形式が固有なで利用者に好都合。行単位にデータの意味と値を指定

マーク付けの方法を定める。タグ(要素, 属性)によって意味, 値が分かる。拡張性あり。サイズが大きい

利用者によって要素名, 属性名を定める

```
<ContactXML
  version="1.1" xmlns="http://www.xmlns.org/2002/ContactXML">
  <ContactXMLItem lastModifiedDate="2003-04-01">
    <PersonName>
      <PersonNameItem xml:lang="ja-JP">
        <FullName pronunciation="コバヤシシゲル">小林 茂</FullName>
      </PersonNameItem>
    </PersonName>
    <Occupation>
      <OccupationItem xml:lang="ja-JP">
        <OrganizationName>日本ユニシス株式会社</OrganizationName>
      </OccupationItem>
    </Occupation>
  </ContactXMLItem>
</ContactXML>
```

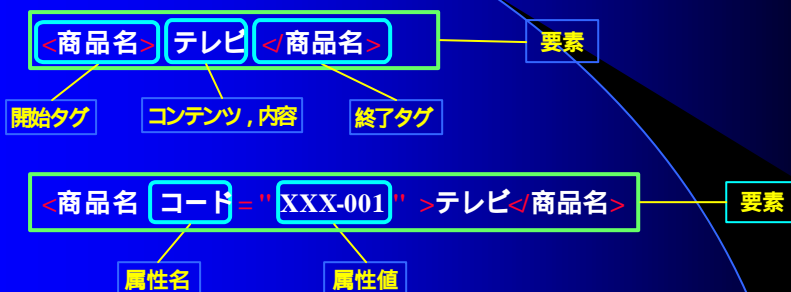
XML(ContactXML)

プログラムで汎用的に 利用しやすいデータ形式の要件

- データ形式が標準化され、仕様が公開されていること
 - ワープロ等専用ソフトに適したデータ形式は、利用しにくい
 - 未公開、複雑、余分な情報、改定されやすい
 - 形式が標準化されていないと、プログラムを作成しにくい
- 容易に利用でき、利用する環境が豊富なこと
- データの変更、利用方法の変更に対応しやすいこと



XML :マーク付け規則



- ◇ 最上位の要素は1つだけ(ルート要素)。コンテンツとして他の要素を含むことが可能
- ◇ 要素、属性の名前に指定できる文字種は限定
 - >1文字目は英字、和字など、下線
 - >2文字目以降は、英数字、和字など、ピリオド、ハイフン、下線、(コロン)
 - >全角英数字、半角カタカナ、中点、スラッシュなどは不可
 - >xm l, XMLなどで始まる名前は不可
- ◇ コンテンツに指定できる文字種は、ユニコードで規定する範囲

入れ子構造, 木構造, XML表現

伝票

顧客情報 小林茂

売上明細

商品情報

商品

コード XXX-001

名前 テレビ

数量 1

金額 1000000

商品の繰り返し

合計情報 1000000

コード	商品名	数量	金額
XXX-001	テレビ	1	100,000
合計			¥100,000

伝票

顧客情報

売上明細

商品情報

商品

商品

...

コード 名前 数量 金額

```

<伝票>
  <顧客情報>小林茂</顧客情報>
  <売上明細>
    <商品情報>
      <商品>
        <コード>XXX-001</コード>
        <名前>テレビ</名前>
        <数量>1</数量>
        <金額>1000000</金額>
      </商品>
      .....商品の繰り返し.....
    </商品情報>
  </売上明細>
</伝票>

```

XML文書の構造

XML文書 (ドキュメント)

```

<?xml version="1.0" encoding="UTF-16"?>
<!DOCTYPE 商品情報 [
  <ELEMENT 商品情報 (商品*)>
  <ELEMENT 商品 (#PCDATA)>
]>
<商品情報>
  <商品>.....<商品>
  <商品>.....<商品>
  <商品>.....<商品>
</商品情報>

```

親要素

子要素

XML宣言

省略可

DTD

省略可

名前・構造の定義

ルート要素

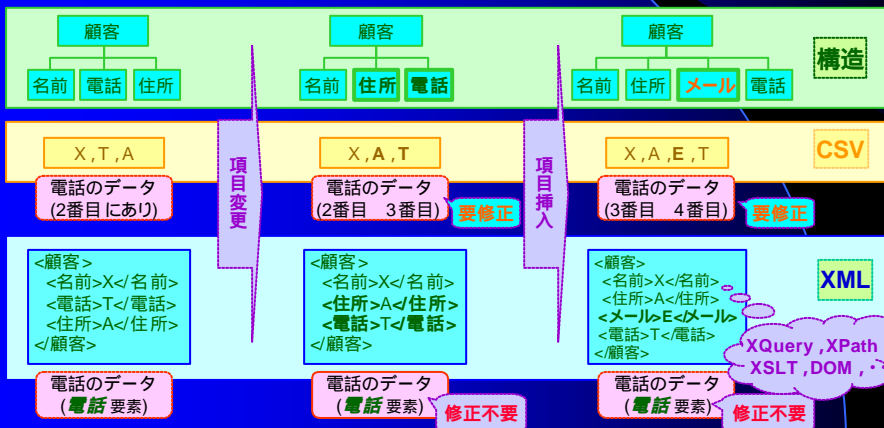
具体的なデータを指定する。ルート要素の中に入れ子する

- ◇XML宣言で文字符号化の方式(UTF-8, UTF-16, Shift_JIS)を指定
 - >省略可能。Shift_JISの場合には必須
- ◇DTD(文書型定義)は省略可能
- ◇ただ一つのルート要素
- ◇コンテンツ中に他の要素を指定可能

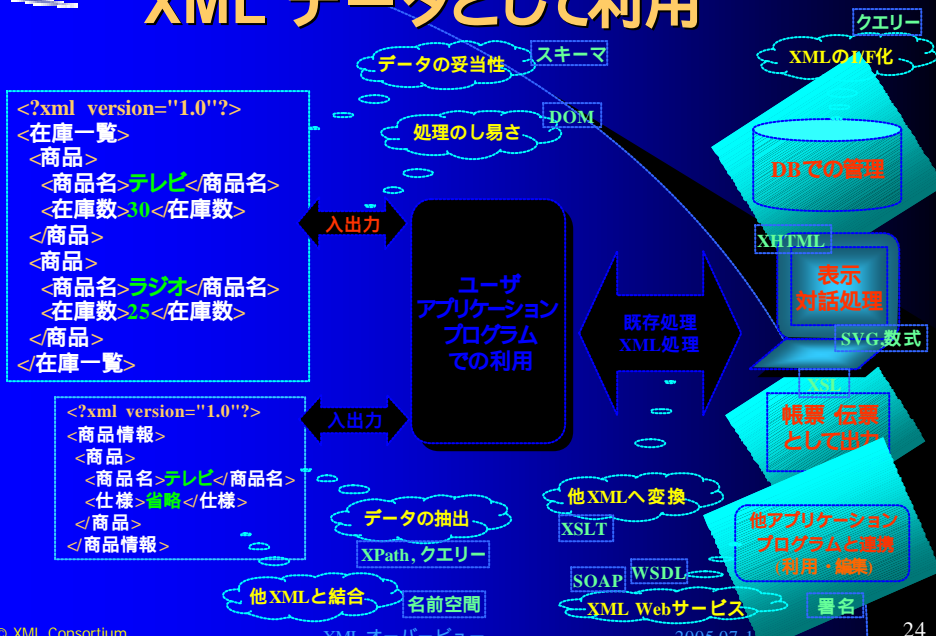
XMLの拡張性・柔軟性

- XMLの利点である拡張性、柔軟性
 - スキーマの変更・追加によるAPへの影響が少ない
 - CSVとXMLによる例示

電話のデータを扱う例
電話/住所の順序入替
メールの挿入による
APの影響は如何に!



XML データとして利用



XML名前空間の指定

•複数のXMLを同時に扱う場合、名前空間を付加して区別する

接頭辞: **xsl** | local-name: **template** → QName: **xsl:template** → `xmlns:xsl="名前を示す空間の文字列"`

→ `xmlns: "名前を示す空間の文字列"`

XSLT

```

<xsl:stylesheet
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns="http://www.w3.org/1999/xhtml"
  ... 中略 ... >
  <xsl:template match="価格">
    <td>
      <xsl:value-of select="."/>
    </td>
  </xsl:template >
  ... 後略 ...

```

XSLTで規定する語彙 → `<xsl:stylesheet>`, `<xsl:template>`

XHTMLで規定する語彙 → `<td>`

XMLからHTMLへの変換例

XSLTではXMLで表現する

XML情報の可視化

XML

```

<商品>
  <商品名>テレビ</商品名>
  <価格>200000</価格>
</商品>

```

カンマ無し

- スタイル情報の付加が必要
- ブラウザする処理系にあわせた言葉に変換する

HTML

```

<tr>
  <td>商品名</td>
  <td>価格</td>
</tr>
<tr>
  <td>テレビ</td>
  <td>200,000</td>
</tr>

```

XSLTで規定

XMLからHTMLへの変換



XMLの特徴

(Extensible Markup Language)

- **平易なテキスト形式。標準規格。**
 - XMLに関連する規格が多い。
 - 長持ちする。人にも分かる。
- **データに構造・意味を持つ。**
 - 拡張性,保守性が高い。
 - 他の用途に利用することが容易である。
- **ソフトウェアベンダによる支援が多い。**
 - 大手ベンダもサポートを表明している。
 - フリーなXMLパーサ, ツールが提供されている。
 - 使い勝手のよい開発環境が提供されている。
- **Webベースで利用が可能である。**
 - 安価なところからスモールスタートで始められる。
- **皆が使い始めている。もはや机上の産物ではない。**

何故XMLなのか

新機能・機能変更の要求
異なる環境での利用
何と結びつかず予め予知できない

- **IT環境の変化, ビジネス環境の変化に即対応可能**
 - 標準化されている為, 考え方, プログラミング等が, OS, フレームワーク上で統一
 - ITベンダが標準的に対応し, 安価に利用可能
 - データとして長期保存に耐え, 再利用が容易

他のアプリケーション固有なデータ形式に比べ, データサイズ, 処理効率の面で劣るにもかかわらず, 利用されている

XMLの意義

- XMLの特徴を活かし、業務システムがより短期間に、低コストで開発できる。
- 異なるシステム間で連携しやすい。予期しない利用法で、経営情報に活かす可能性を秘めている。
- 自社にとってのメリットに加え、外部へXMLでのインタフェースを提供することにより、Win-Winの関係を達成できる。

XML / Webサービス適用による 新たなビジネス展開



まとめ

- **Webの発展**
 - 単純な閲覧から , 複雑な連携
 - シンプルクライアントから , リッチクライアントへ
 - モバイル・アクセス , ユビキタス・ネットワークへ
- **情報記述としてのXML**
 - XMLはデータ記述に関するインフラ
 - XMLは目的ではなく , IT技法のひとつ
 - XMLは難しくない。既に使われている
- **使えるところから始めよう**

付録 参考情報

- **関連サイト**
 - **XMLコンソーシアム**
<http://www.xmlconsortium.org/>
 - **W3C 各種規格の公開**
<http://www.w3.org/TR/>
 - **@IT XML & SOA (Web Services)**
XMLとSOA (Webサービス) を理解し、ビジネスに活用するためのフォーラム
<http://www.atmarkit.co.jp/fxml/>
- **書籍**
 - **「リアルWebサービス」**
XMLコンソーシアム監修 , 秀和システム発行 , 1,600円 , 2005-04-15