

# XMLってな～に？

- 初心者のためのXML入門 -

株式会社エー・ピー・エム  
開発部 古源 明広

# ある会社で

- 営業部門がグループウェアに以下の内容のワープロ文書を登録した

「昨日 A B C 建設会社から A20015 の注文がありました！」

# 社員はどう理解するか

「昨日 A B C 建設会社から A20015 の注文がありました！」

- A B C 建設会社・・・我が社の古くからのユーザー
- A20015・・・我が社が最近発売した新製品
- 注文数・・・この製品は通常企業に一つあればよいので、特に記述がなければ一つである

# なぜ理解できるか

「昨日 A B C 建設会社から A20015 の注文がありました！」

1. 社員は日本語がわかる
2. 社員は単語を会社独自のボキャブラリとして理解できる
  - A B C 建設会社は古くからのユーザー
  - A20015 は新商品
  - A20015 は複数入るような商品ではない

# それではコンピュータは？

「昨日ABC建設会社からA20015の注文がありました！」

- まず日本語の場合、単語に区切ることから難しい(^^;
- AI機能によりそれがクリアできたとしても、
- ABC建設会社が古くからのユーザーであること、A20015が新製品であること、注文数が1つであることを、
- このワープロ文章だけでコンピュータが  
⇒わかるはずがないですね！！

# じゃあHTMLなら？

- この文書をHTMLにしたらコンピュータはわかるかな・・・

```
<html><body>  
<font size=+2 color=red>  
<b>昨日 A B C 建設会社からA20015の注文がありました！</b>  
</font>  
</body></html>
```

⇒まったく同じ、わかりません！！

# 本当はこうやりたい！

「昨日 A B C 建設会社から A20015 の注文がありました！」

```
<html><body>
```

```
<order>
```

```
  <date>2001/5/24</date>
```

```
  <customer code="U000003"> A B C 建設会社  
  </customer>
```

```
  <product>A20015</product>
```

```
  <quantity>1</quantity>
```

```
</order>
```

```
</body></html>
```

# これができたら？

- ソフトウェアがデータ内部の要素の意味を理解して自動処理ができる  
＝データベース的処理が行える
- WEBやメールなどのインターネット技術を使って自由自在にデータコミュニケーションができる

# でもHTML3.2じゃできない

- HTMLが今日のWEBの世界的普及に果たした役割は大変大きいですが、ソフトウェアから見ると本質的にはワープロ文書とそう変わりない
- HTMLは見栄えアップのための文字装飾を行うタグはたくさん用意されているが、論理的な定義、文書の構造を定義するタグはあまり用意されていない
  - <html>、<title>、<body>、<h1>~<h6>など
- 構造を定義するためのタグを新たに作りたいたいと思っても、HTMLの仕様では、エンドユーザーが勝手に拡張することはできない

# それなら新しい言語を！

- それなら新しい言語を作ればよい
- でも言語の仕様を決めてしまうとHTMLの二の舞となる（＝拡張性がなくなる）
- うーむ、、、
- それなら言語を作る言語を作ればよい！

# メタ言語

- 言語を作る言語、これがメタ言語である
- HTMLはSGMLというメタ言語から生成された
- それならSGMLを使えばいいのでは？

# SGML (Standard Generalized Markup Language)

- 大変複雑で膨大なメタ言語(1986年ISO制定)
- 習得が困難で、とても標準化は無理
- しかもWEBに対応していない
- そこで、、、
- W3C内に、SGML W/Gが結成され  
(1996年6月)、新たなメタ言語の研究開発  
に取り掛かった

# XML誕生

- XML・・・eXtensible Markup Language
  - eXtensible（拡張可能な）
  - Markup Language（マーク付き言語）
- 1996年11月にXML仕様の第一草稿完成
  - 草稿全部で30ページ程度というその簡潔さに、世界中の（SGMLを知る）技術者が驚き、その可能性に気付いた⇒テクノロジーブレイクスルーと呼ばれる

# XMLの特徴

- WEB対応のシンプルなメタ言語である
- タグを自由に拡張できて、構造化ができる
  - データに理論的な意味を持たせることができる
- データを純粹にデータだけで持てる
  - 構造定義、表示フォーマットを分離
- テキストである
  - テキストだけでデータベースシステムを構築できる
  - 標準化、コスト大幅削減、データ永久保証が可能

# 何者かの定義

- XMLはSGMLの後継言語
  - SGMLのサブセットではなく新たな体系
  - WEB機能を付加
  - 文法上はSGMLのサブセットのところがある
- HTMLはSGMLの子供
  - HTMLはSGMLを使って開発された言語
- 従って、XMLはSGMLの縮小版ではなく、HTMLの拡張でもない、新たな体系である

# XMLデータ (注文.xml)

「昨日ABC建設会社からA20015の注文がありました！」

```
<?xml version="1.0" encoding="SHIFT-JIS" ?>  
<order>  
  <date>2001/5/24</date>  
  <customer code="U000003">ABC建設会社  
社  
  </customer>  
  <product>A20015</product>  
  <quantity>1</quantity>  
</order>
```

# XML 構造定義 (DTD, XMLスキーマ)

【注文.dtd】

```
<!ELEMENT order  
  (date,customer,product,quantity)>  
<!ELEMENT date      (#PCDATA)>  
<!ELEMENT customer  (#PCDATA)>  
<!ELEMENT product   (#PCDATA)>  
<!ELEMENT quantity  (#PCDATA)>  
<!ATTLIST customer  code CDATA #REQUIRED>
```

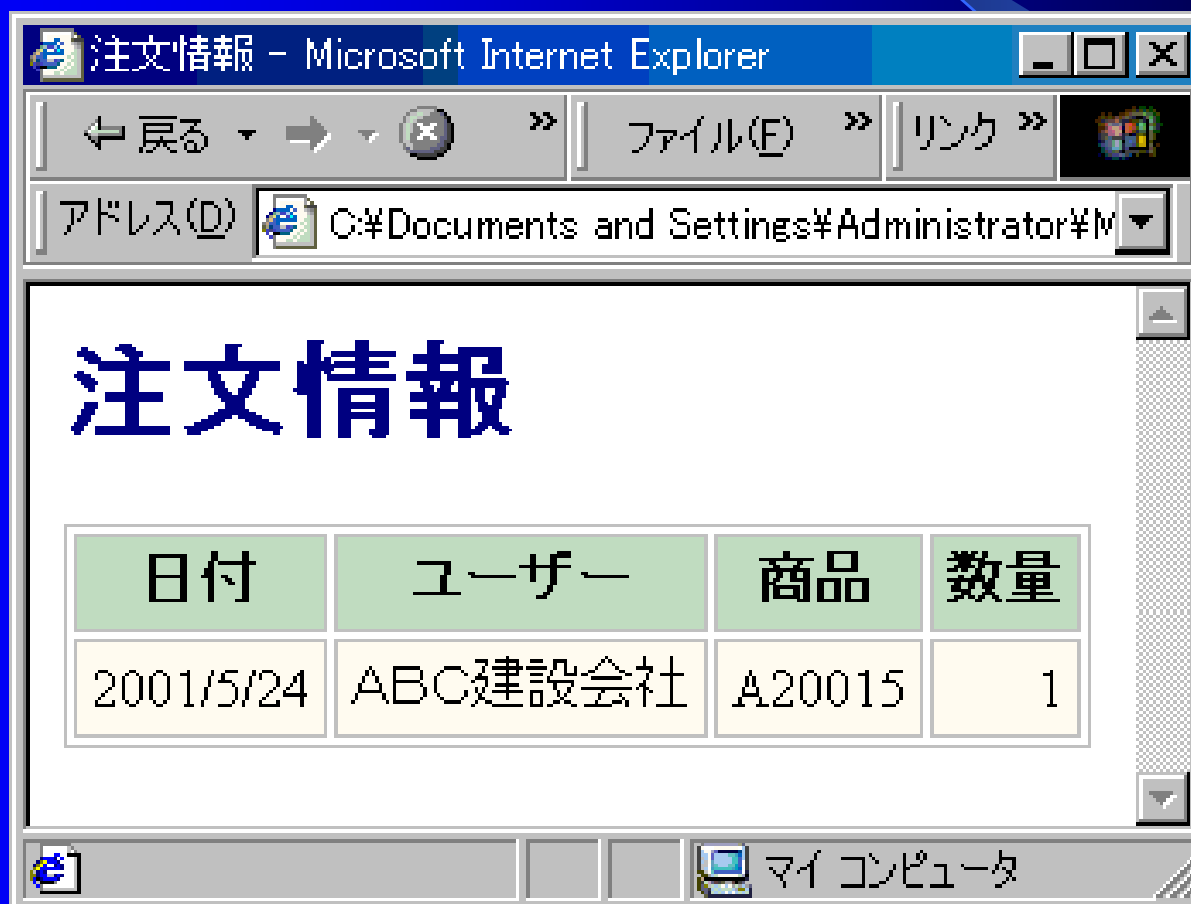
※現在はXMLスキーマを使うのが主流

# XMLスタイルシート (注文.xsl)

```
<?xml version="1.0" encoding="SHIFT-JIS" ?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD-xsl">
<xsl:template match="/"><html><body><table border="1">
<tr><th>日付</th><th>ユーザー</th>
<th>商品</th><th>数量</th></tr>
<xsl:apply-templates select="order" />
</table></body></html></xsl:template>
<xsl:template match="order"><tr>
<td><xsl:value-of select="date" /></td>
<td><xsl:value-of select="customer" /></td>
<td><xsl:value-of select="product" /></td>
<td align="right"><xsl:value-of select="quantity" /></td></tr>
</xsl:template></xsl:stylesheet>
```

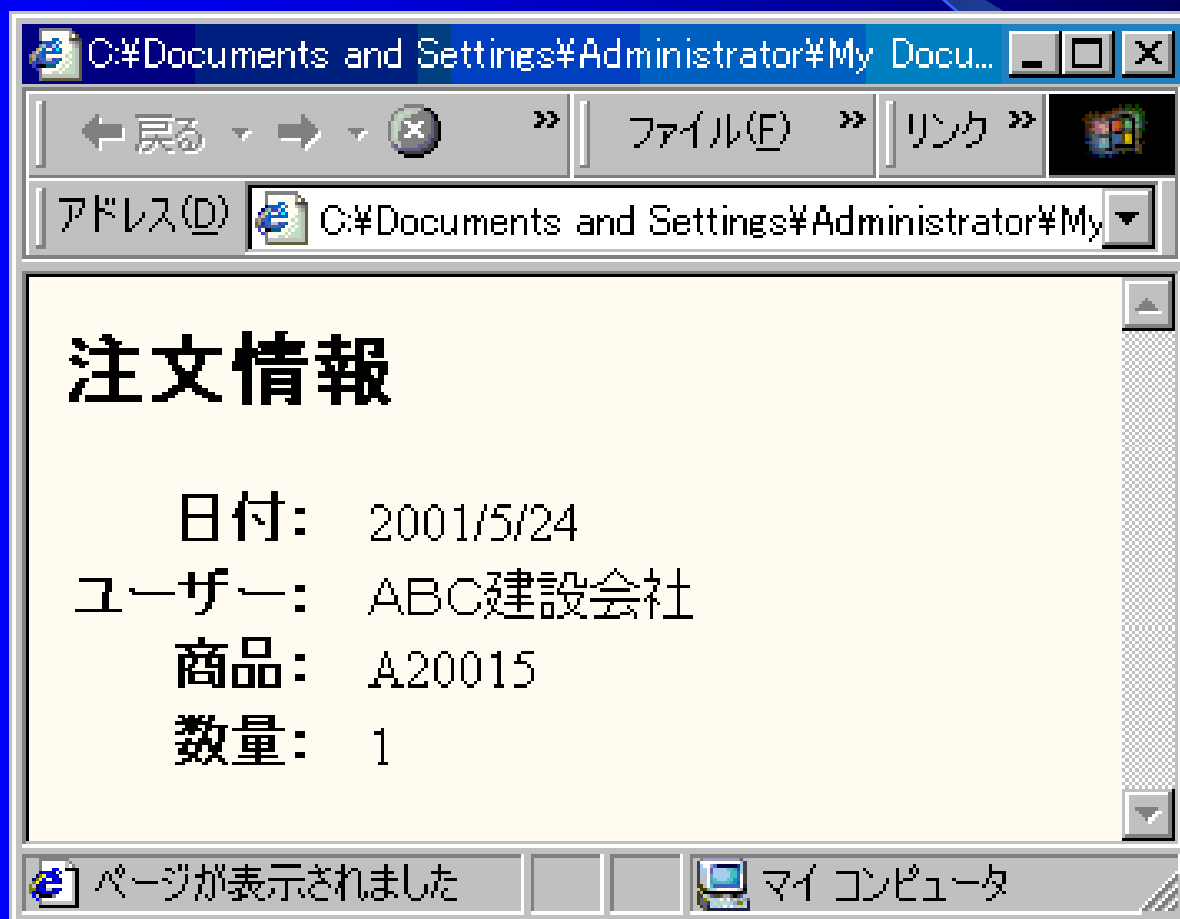
# このスタイルシートを使う と・・・

こういう表現ができる



# 違うスタイルシートを使う と・・・

違う表現に一瞬で切り替えられる



# コンピュータにも人間にも

- 構造定義 (.dtd,XMLスキーマ) を使えばコンピュータソフトウェアが理解できる
- スタイルシート (.xsl) を使えば人間が理解できる
- しかも同じデータリソース (.xml) を使って!

⇒究極のデータ活用形

# XMLは情報交換に強い

- XMLは文書であり、データベースである
- 構造定義の標準化さえなされていれば、情報交換に大変強い
- ソフトウェア同士の論理的コミュニケーションができる
- 例えば、、、

# XMLと医療情報（その1）

- 患者情報ほど重要なものはない
  - 病歴、アレルギー情報、現在投与中の薬、血液型など、命にかかわる情報ばかり
- 患者情報システムの現状
  - 患者情報をデジタルで持つ病院は多いが、他の病院の患者情報システムとのデジタル的なデータ交換は標準化・日常化されているとはいえない

# XMLと医療情報（その2）

- 現状のデータ交換の方法
  - 現在、患者情報の交換がなされているとすれば、システムのハードコピー画像、フリーテキスト、HTML、PDFなどのデジタルオブジェクトをメールする、（印刷して）郵送、FAXするなどの方法で送信することになる
  - そうすると、受け取り側の病院では、そのデータを手入力またはカット&ペーストでその病院の患者管理システムにインプットするしかない

## XMLと医療情報（その3）

- 手入力は面倒くさい！
- インプットミスが起こるかもしれない！
  - 投与してはいけない薬を投与してしまうかも
    - 患者はアレルギー症状を起こす可能性がある
    - 最悪死亡するケースも考えられる
- だからやっぱりやらない
  - 患者は一から検査のやり直し（病院は儲かるが）
- 両方の病院ともデジタルシステムを使っているというのにこの有り様

# XMLと医療情報（その4）

- XMLでデータ交換ができれば・・・
  - 他病院の患者管理システムに人間の手を介さずにインプットできるので、データが正確に瞬時に伝わる
  - 少なくとも他病院からの情報伝達ミスにおける医療ミスは根絶される
  - 余計な検査が減り、即治療に入れる
  - 医療費負担が減る

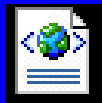
# XMLと医療情報（その5）

- 前提条件として・・・
  - 厚生労働省により、患者情報XMLスキーマの標準化が行われていることが必要
- 世界的な患者情報の標準化を行うことも可能
  - 海外旅行中に事故にあっても適切かつ迅速な治療を受けられる

# XMLが使われているHP

- P P I

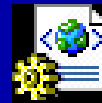
- <http://www.ppi.go.jp/>
- 国土交通省の入札情報検索サービス
- 工事内容がXMLになっている



bukken\_DATA.xml



Opfhykj.dtd



Opfhykj.xsl

# XMLとは自由3rd

- 1970年代：大企業メインフレームの時代 ⇒ 端末を操作する社員には何の自由もない
- 1980年代：パソコンの登場 ⇒ ソフトウェアを選べる自由が与えられた（自由1st）
- 1990年代：インターネット、WEBの登場 ⇒ 世界中に情報を発信できる自由が与えられた（2nd）
- 2000年代：XMLの登場 ⇒ 世界が理解できる情報を創り出す自由が与えられた（3rd）

# XMLは21世紀型ツール

- 人類の意識レベルに合ったツールが与えられる
- そしてそのツールが意識変革を促す
- 21世紀型の意識とは
  - 自己の確立
  - 自立・分散・協調
  - オープン・透明性
- XMLは人類の21世紀型意識変革に貢献する