

文書を部品化し再利用する仕組み

SUGI-J2001

July 27 (v1.1)

福島 竜 (lyu@mamezou.com)

株式会社 豆蔵



概要

- 大量の文書にたちむかう
- 大量の文書はこう管理したい
- 再利用をめざす仕組み
 - UMLについて
 - XMLについて
 - SVGについて
- SASによる統計情報の利用
 - SAS XML LIBNAME Engine

株式会社 豆蔵



大量の文書にたちむかう

なにが問題か？

大量の文書を

- どう作るか
- どうメンテナンスするか

再利用はできるのか

- どのレベルまで利用できるのか
- 計算データ、計算結果



株式会社 豆蔵



大量の文書はこう管理したい

ステップ1 :

電子的につくること

- 保管しやすい(CDROM)
- 複製をつくりやすい
- フルテキストサーチできる
- カット&ペーストできる
- ...

株式会社 豆蔵



さらなる要求

シリーズものの文書をつくるとき...

- 同じところは手作業でなく自動的にしたい
- 変わるところだけ変えたい
- 計算データ、結果は変わる
- 計算処理、計算データと結果の表現(図、表)は変わらない



株式会社 豆蔵



大量の文書はこう管理したい(2)

ステップ2: 構造をつくること

- 章立て
- 情報とビューを分ける
- 文字、イメージ(写真、図)、表、音
- 変わるところと変わらないところを分ける
- 情報と処理を分ける

(... XMLが得意、まとめて持てる)

再利用するところと、再利用はできないところをわかる

株式会社 豆蔵



構造化は有効か？

つくりたい文書(情報)を分析してみて
再利用するところがはっきりし
その量が多いときは

構造化は有効



株式会社 豆蔵



実は、

企業などで作成する文書はこのタイプが多い
(繰り返し作られているので)

- 企画書
- 製品仕様書
- 製品マニュアル
- 議事録
- 設計書

...

株式会社 豆蔵



再利用をめざす仕組み

- 部品化の検討: OMI (UML)
- 部品化: XML
- 情報の生成: SAS
- 文書情報の格納: DB + XML
- 文書情報の検索: DB + XML(XPath)
- 文書の作成: SAS + XML + (Java..)
- 統計情報の参照: SVG



株式会社 豆蔵



OIMについて

Open Information Model

- メタデータの仕様
- UMLでモデリングする



株式会社 豆蔵



UMLについて

Unified Modeling Language

- システム開発に使われる
- 10種類のダイアグラムを駆使
- モデルを作るための記法
- Javaなどで開発するときに効果発揮
 - オブジェクト指向



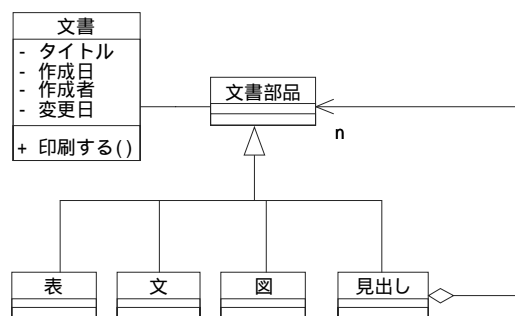
株式会社 豆蔵



UML - クラス図の例

クラス図は、最も良く使われるダイアグラム

文書の構造
概念モデル



株式会社 豆蔵



XMLについて

eXtensible Markup Language

- HTMLの仲間
- 構造と内容をあわせもつ
- 内容は電子的なものなら、なんでもうけいれる
(プログラムも)

株式会社 豆蔵



XMLの構造 – 約束はシンプル

<タグ名>内容</タグ名>

例: “お昼のあいさつは、こんにちは(英語でhello)で夜の
あいさつは、こんばんは(英語でgood evening)”

CSV+IDLだと

IDL

タイトル:文字

*(昼:文字,英語:文字,夜:文字,英語:
文字)

CSV

あいさつ,こんにちは,hello,
こんばんは,good evening

XMLだと

<あいさつ>

<昼>こんにちは

<英語>hello</英語>

</昼>

<夜>こんばんは

<英語>good evening</英語>

>

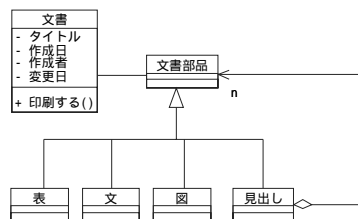
</夜>

</あいさつ>

株式会社 豆蔵



XML と クラス図(再見)



<文書>
<見出し>
 <文></文>
</見出し>
<見出し>
 <図></図>
</見出し>
...
</文書>

株式会社 豆蔵



XMLの利用されている分野

- 文書管理管理をはじめ、様々なシステムの分野で利用されている
- EAI(Enterprise Application Integration): 企業内・間のさまざまな業務の相互連携を実現する
- ワークフローエンジン
- ...



あまりにありすぎてどれからあげればいいのか...

株式会社 豆蔵



XMLとデータベースについて

- データベースとの親和性について

シンプルなアプローチ

- タグをカラムとする

XML専用のデータベース製品もある

- eXcelon

株式会社 豆蔵



SVGについて

Scalable Vector Graphics

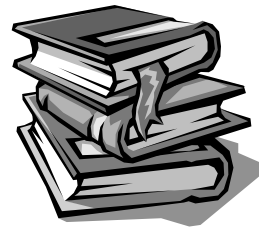
- XMLを利用したグラフィックフォーム
- データ駆動(Data-driven graphics)
- メンテナンス性が高い
- 作成コストも低い
- いろいろなところで利用できる
- グラフィックデータと処理をあわせもっている

株式会社 豆蔵



文書情報をXMLで表現する

- 構造を決める
- 内容を検討する(変わるか、変わらないか)
 - 再利用性が高まる
- 管理はDBと組み合わせる



株式会社 豆蔵



SASによる統計情報の利用

- データと処理結果をXMLに出力する
- SAS XML LIBNAME Engineの利用

LIBNAME librefXML 'external-file' <XML-engine-options>

XMLTYPE=GENERIC | ORACLE | OIMDBM | EXPORT | HTML

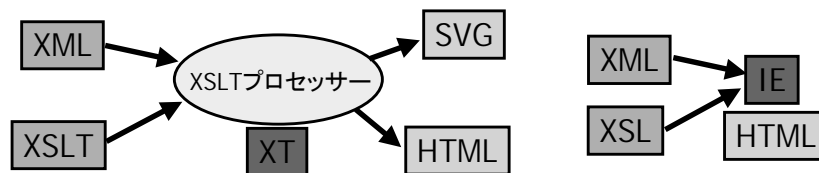


株式会社 豆蔵



LIBNAMEによる処理 (demo)

- SAS XML LIBNAME EngineでXMLを作成
- XMLをXSLTプロセッサでSVGなどに変換



株式会社 豆蔵



XML出力をDBで管理する

- 検索した結果を文書のかたちでみる
- 検索した結果から新しい文書をつくる
- 図表はSVGで出力(表示)する

株式会社 豆蔵



最後に

この方法の利点

- 文書部品の生成が自動化できる
 - 再計算した内容を反映
- 大量の文書を扱うことができる
 - 電子化
 - 構造化
 - DBに構造化して保存・管理(DB+XML)

株式会社 豆蔵



End of ppt

株式会社 豆蔵

