

DEWS
2002

XMLパーサを考慮した応用向き関係データベース構成法



東京都立大学 大学院

工学研究科 電気工学専攻

横山昌平 太田 学

片山 薫 石川 博

発表内容

- 研究の背景
- 提案システムについて
 - データベースの構成法
 - データの更新
 - SAXによるアクセス
- まとめと今後の課題

DEWS
2002

データベース

■利用している関係データベースの機能

■更新

- SQLのINSERT、UPDATE文

- ODBC/JDBC

- トランザクションで各シーケンスを分離

■呼び出し

- SQLのSELECT文

- ODBC/JDBC

XMLをRDBに格納することで、RDBが既に実装している機能を容易にXMLに適用できる。

関連研究

参考文献: 赤堀正剛, 有澤達也, 遠山元道:
SuperSQLによる関係データベースとXML
データの統合利用, 情報処理学会論文誌:
データベース, Vol.42, No. SIG 8(TOD-10),
pp.66-95 (2001).



参考文献: 吉川 正俊, 志村 壮是, 植村 俊亮: オブジェクト関係データ
ベースを用いたXML文書の格納と検索, 情報処理学会論文誌: デ
ータベース第40巻, 第SIG6(TOD3)号, pp. 115-131 (1999)

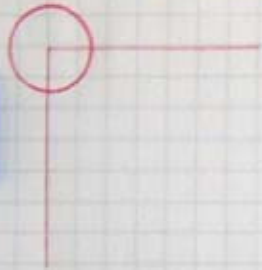
● 提案システム

- 多種複数のXML文書をRDBに格納
 - RDBの機能をXMLに適用
- 更新
 - 差分データの抽出とそれを用いた更新
- 呼び出し
 - diff/patchの更新手法を継承
 - SAXを用いた優位性を考慮
 - 差分情報のみを利用する事による更新用メモリ削減
 - SAXを用いたプログラムの移植が容易
 - 関係データベースの知識を必要としない

● システム構成

- Saxophone
 - XMLをRDBに格納する
- xoDiff
 - XMLファイル同士の差分をXML形式で取得
 - XMLの構造的差異に基づく差分データ
- xPatch
 - XoDiffの出力でRDBのデータ更新
- Saxophone SAX パーサ
 - SAXパーサ(ユーザからRDBを隠蔽する)

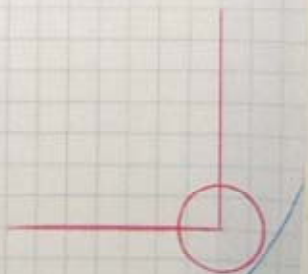
DEWS
2002



[格納]

[更新]

[呼び出し]



データベース格納

1. XMLファイルをSAXで読み込む
2. イベントに基づき二次元表形式に変換

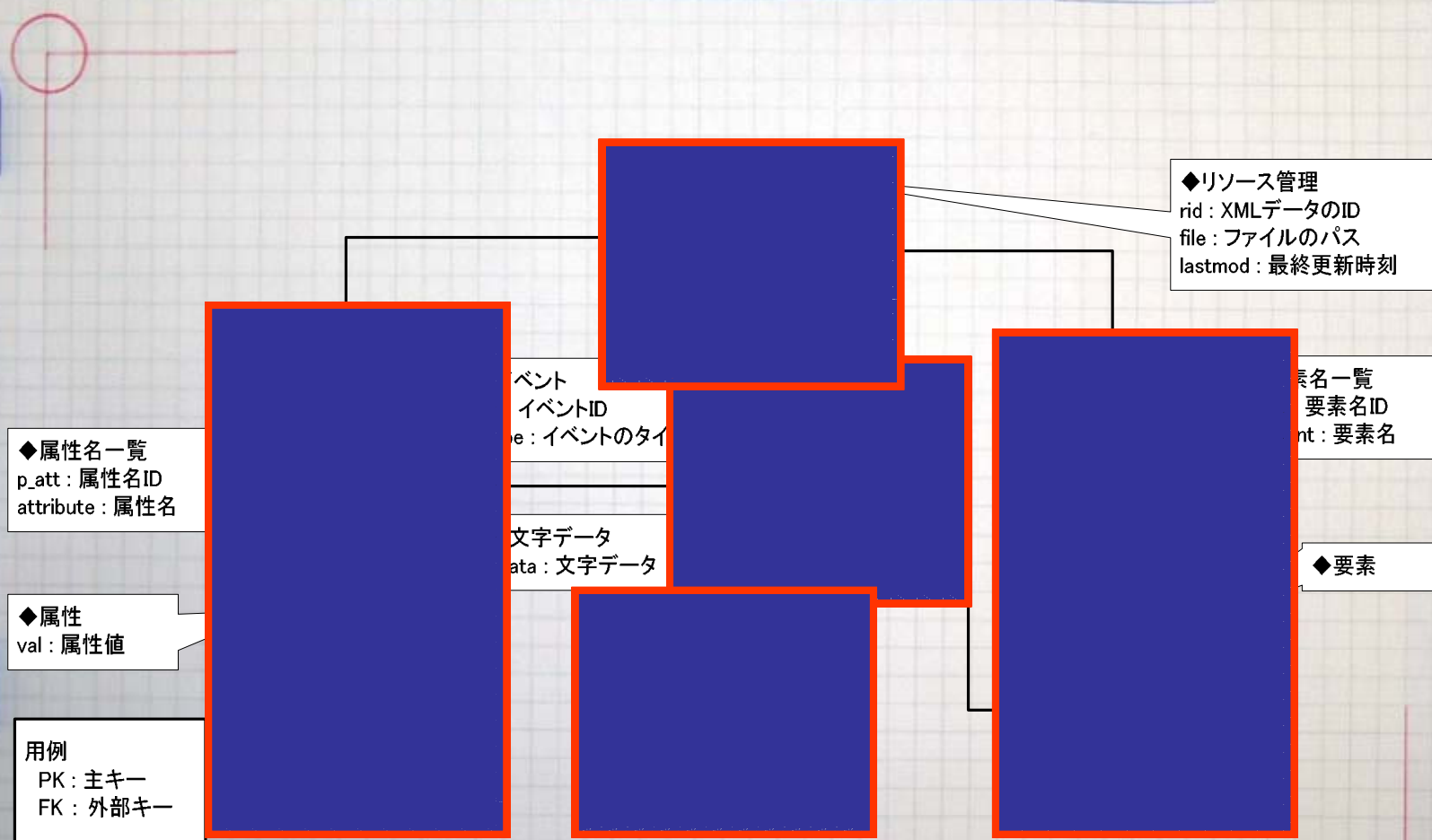
	イベント種類	プロパティ
1. 要素開始: theCUP	要素開始	theCUP year="2002">
2. 属性出現: {year,2002}	属性出現	{year,2002}
3. 文字データ: Japan	文字データ	Japan
4. 要素終了: theCUP	要素終了	theCUP

Saxophoneデータモデル(SaxDM)

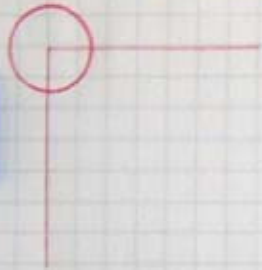
3. 関係データベースに収納

DEWS
2002

データベース構成



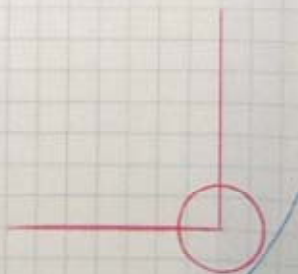
DEWS
2002



[格納]

[更新]

[呼び出し]



バージョン間差分取得

- テキストファイルの差分取得: diff

- SaxDMのXML表現に変換

```

XML1: <theCUP year="2002" >Japan</theCUP>
XML2: <theCUP year="2002" >Korea Japan</theCUP>
1行目(0) 1<startElement name="theCUP" type="text" />
2行目(1) 1<startElement name="theCUP" type="text" />
3行目(2) 1<startElement name="theCUP" type="text" />
4行目(3) 1<startElement name="theCUP" type="text" />
5行目(4) 1<startElement name="theCUP" type="text" />
6行目(5) 4<startElement name="theCUP" type="text" />
<sE name=" theCUP" />
<eventAttribute name="year" value="2002"/>
-<eventCdata>Japan</eventCdata>
+<eventCdata> Korea Japan</eventCdata>
<endElement name="theCUP"/>
-0.31229900 1014739106.hiken1.eei.metro-u.ac.jp </saxophone>
+0.31653400 1014739106.hiken1.eei.metro-u.ac.jp </saxophone>

```

XML表現に変換

1行目(0) 1<startElement name="theCUP" type="text" />

2行目(1) 1<startElement name="theCUP" type="text" />

3行目(2) 1<startElement name="theCUP" type="text" />

4行目(3) 1<startElement name="theCUP" type="text" />

5行目(4) 1<startElement name="theCUP" type="text" />

6行目(5) 4<startElement name="theCUP" type="text" />

<sE name=" theCUP" />

<eventAttribute name="year" value="2002"/>

-<eventCdata>Japan</eventCdata>

+<eventCdata> Korea Japan</eventCdata>

<endElement name="theCUP"/>

-0.31229900 1014739106.hiken1.eei.metro-u.ac.jp </saxophone>

+0.31653400 1014739106.hiken1.eei.metro-u.ac.jp </saxophone>

diffへ入力

更新 (イベント挿入)

<theCUP year="2002">Korea Japan</theCUP>

<theCUP winner="Japan" year="2002">Korea Japan</theCUP>

1.要素開始: theCUP

2.属性出現: {year, 2002}

3.文字データ: Korea Japan

4.要素終了: theCUP

[挿入]

1.要素開始: theCUP

1.1.属性出現: {winner, Japan}

2.属性出現: {year, 2002}

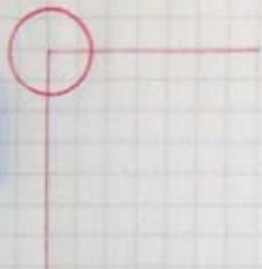
3.文字データ: Korea Japan

4.要素終了: theCUP

問題点: イベントIDの付け替えが大量に発生してしまう。
解決策: eidフィールドに小数值を用いる

eidを昇順でソートした順序 = イベントIDを昇順でソートした順序

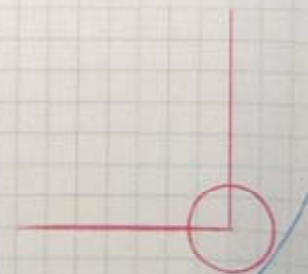
DEWS
2002



[格納]

[更新]

[呼び出し]



SAXパーサ

1. XMLファイルが更新されているか？

- 関係データベースの隠蔽
 - XMLファイルの呼び出しと同じ手順
- XMLファイルとデータベースを同期
 - xPatch以外のツール(例:XSLT)で更新できる
- 関係データベースの機能が利用出来る
 - ODBC、JDBC経由のアクセス
 - ストレージサーバを物理的に分離できる
 - 三層サーバ・クライアントシステム

3.文字データ: 杯 盃 口 卒 `</theCUP>`

3.文字データ: Japan

4.要素終了: theCUP

● まとめ

- SAXを考慮したデータベース構成法を提案
- 格納
 - 更新やSAXでの呼び出しを考慮
- 更新
 - XMLの構造を理解した差分データの取得
 - 差分データを用いた更新
 - イベントIDに少数値を用いることによる最適化
- 呼び出し
 - SAXパーサを用いたアクセス
 - 関係データベースは完全に隠蔽されている

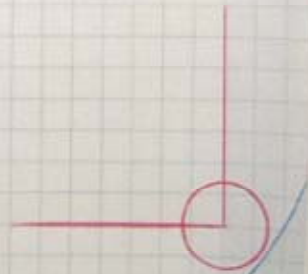
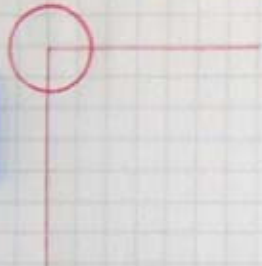
● 今後の課題

- 定量的評価
 - SAXと動作速度を比較
- 最適化
 - データベースの最適化
 - 更新・呼び出しの最適化
- 他のXML処理系と連携
 - SAXパーサを用いた処理系の移植は容易
 - XML問い合わせ言語等

DEWS
2002



ご清聴を感謝します。
ありがとうございました。



DEWS
2002

● システムの概要



DEWS
2002

応用例

(提案システムを用いた 3層アーキテクチャ)

