

第13回 Webインテリジェンスとインタラクション研究会  
ウェブインテリジェンスを支える XML (再) 入門

## XMLデータベース(再)入門

天笠俊之(筑波大学大学院システム情報工学研究科)

1

### チュートリアルの内容

---

- ▶ XMLデータベース
  - ▶ XMLデータの表現
  - ▶ ネイティブ vs リレーショナル
- ▶ 問合せ言語
  - ▶ XQuery 1.0
  - ▶ SQL/XML
- ▶ XMLデータベースアクセスAPI
  - ▶ XML:DB
- ▶ まとめ

---

▶ 2

## XMLデータベース

3

### XMLデータベースとは？

- ▶ XMLデータ(文書)を扱う機能を持つデータベース
  - ▶ ネイティブXMLデータベース
    - ▶ XMLのための専用の格納エンジンを持つ
    - ▶ 狭義のXMLデータベース
  - ▶ XML対応関係データベース (XML-enabled RDBMS)
    - ▶ 多くのシステムが、関係データに加えてXMLデータをサポート
      - Oracle, IBM DB2, MS SQL Server, PostgreSQL. ...
    - ▶ 広義のXMLデータベース

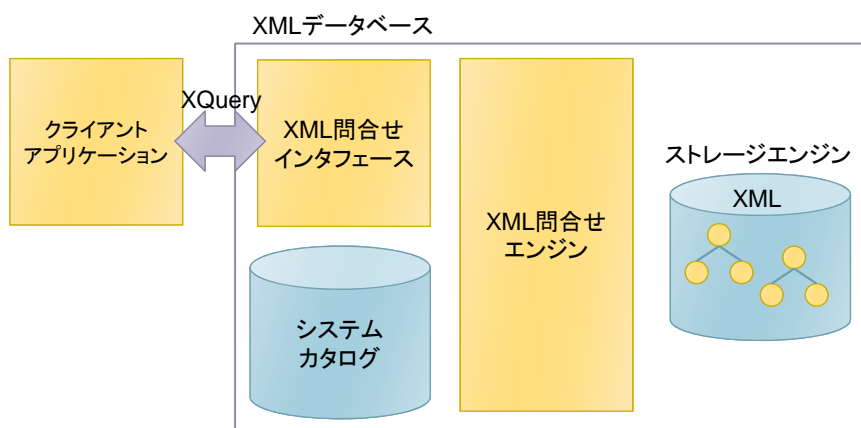
▶ 4

## なぜXMLデータベースが利用されているか？

- ▶ 高い柔軟性
  - ▶ 関係データベース
    - ▶ 運用中のデータ構造の変更が困難
    - ▶ 例外への対応が困難
  - ▶ XMLデータベース
    - ▶ 高い拡張性
      - 必ずしもスキーマを要求しない(wellformedness)
      - データ構造の変更を前提としたデータベース構築
- ▶ 過去10年間でのXMLデータの普及
  - ▶ 多くの応用
  - ▶ 多くのデータがXML形式で生成・流通

▶ 5

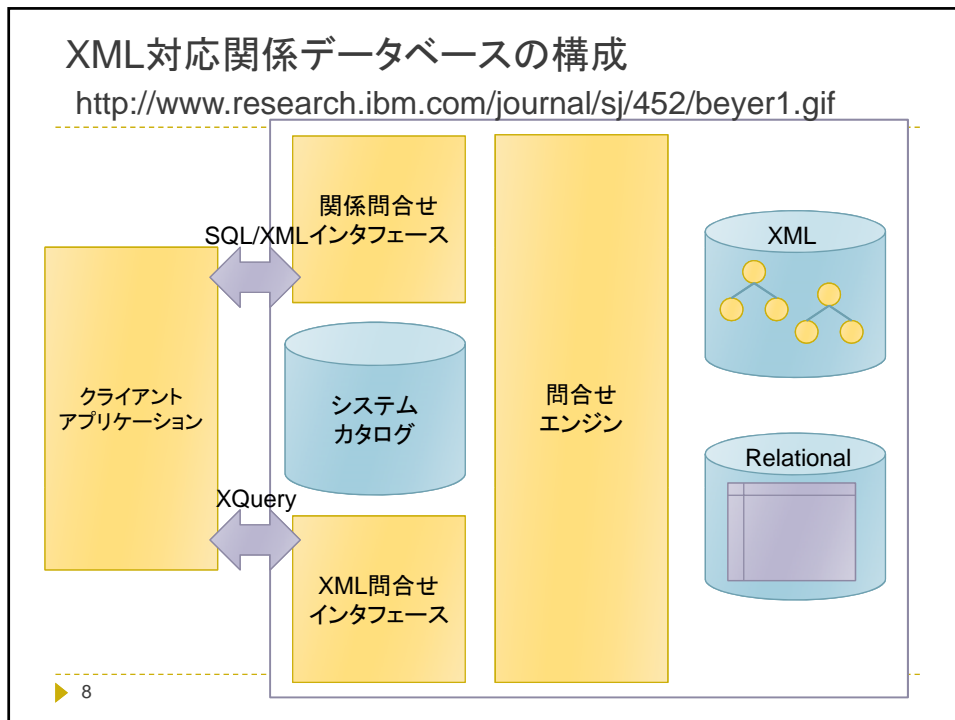
## XMLデータベースの構成



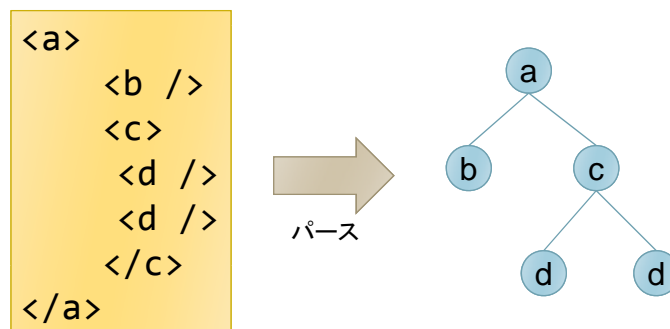
▶ 7

## XML対応関係データベースの構成

<http://www.research.ibm.com/journal/sj/452/beyer1.gif>

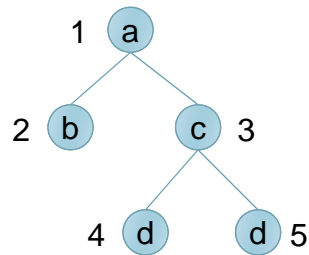


## XMLデータの格納方式 木構造の表現



▶ 9

## 木構造の表現 エッジアプローチ [FK99]



ノード

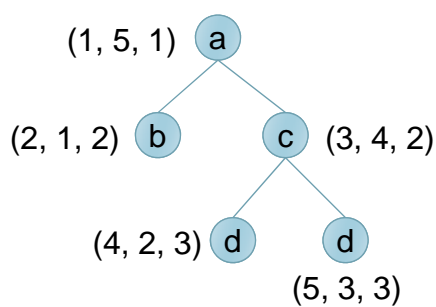
ID	Name
1	a
2	b
3	c
4	d
5	d

エッジ

From	To
1	2
1	3
3	4
3	5

▶ 10

## 木構造の表現 ノードラベリング(範囲ラベリング) [HLM03]



ノード

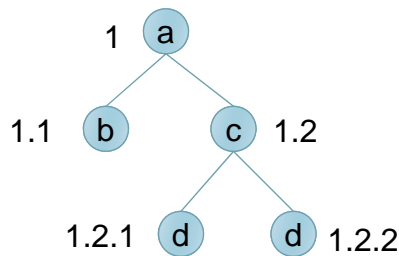
ID	Name	Label
1	a	(1, 5, 1)
2	b	(2, 1, 2)
3	c	(3, 4, 2)
4	d	(4, 2, 3)
5	d	(5, 3, 3)

- ノードラベル  
(preorder, postorder, level)
- 先祖子孫関係(A-D)  
 $A.pre < D.pre$  AND  $D.post < A.post$
- 親子関係(P-C)  
 $P.pre < C.pre$  AND  $C.post < P.post$  AND  $P.level = C.level - 1$

▶ 11

## 木構造の表現

### ノードラベリング (Deweyオーダ) [BTS02]



ノード

ID	Name	Label
1	a	1
2	b	1.1
3	c	1.2
4	d	1.2.1
5	d	1.2.2

- ノードラベル (再帰的な定義)

- 根ノードのラベル: 1

- あるノードのラベル: 親のラベル + 兄弟ノードにおける順序

▶ 12

## 関係XMLデータベースと ネイティブXMLデータベースの違い

- ▶ データの格納先が関係表であるかどうか
  - ▶ 木構造の表現に用いられるテクニックは共通
- ▶ 関係ストレージとXMLストレージの差
  - ▶ ディスクページの構成およびレイアウト
    - ▶ XMLストレージ: XMLの構造やアクセスパターンを重視した設計
    - ▶ 関係ストレージ: 関係スキーマに応じて決定
  - ▶ ノード間順序
    - ▶ XMLストレージ: ノード間の順序を保証
    - ▶ 関係ストレージ: タプル間の順序は保証されない
      - 関係表の走査性能を重視した実装 (カラムストア MonetDB/XQuery)
  - ▶ 索引
    - ▶ XMLストレージ: XMLを意識した設計
    - ▶ 関係ストレージ: 既存の索引 (B木, ハッシュ) を利用

▶ 13

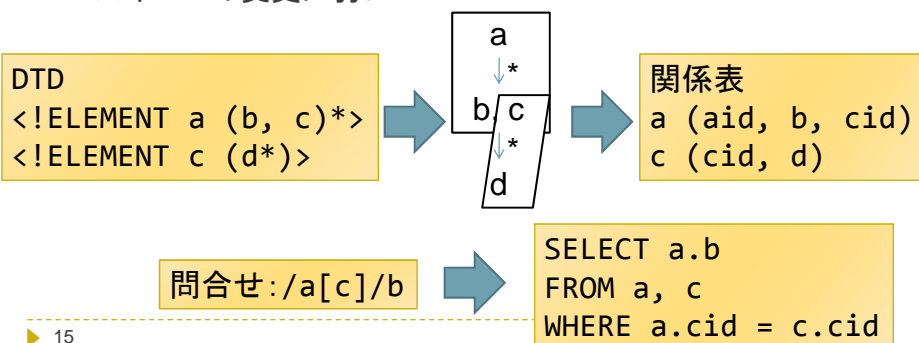
## XMLデータの格納方式 [YASU01]

- ▶ 関係XMLデータベース
  - ▶ 構造写像
    - ▶ XMLスキーマを関係スキーマに変換
  - ▶ モデル写像
    - ▶ 木構造を格納できるような関係スキーマを設計
- ▶ ネイティブXMLデータベース
  - ▶ (関係データベースを利用しない)XML専用の格納方式全て

▶ 14

## 関係XMLデータベース 構造写像 [BFH+02]

- ▶ DTD / XMLスキーマを関係表に写像
  - ▶ DTD(XMLスキーマ) → DTDグラフ → 関係表
- ▶ スキーマ情報が必須
  - ▶ スキーマの異なるXMLデータは格納不可
  - ▶ スキーマの変更に弱い



▶ 15

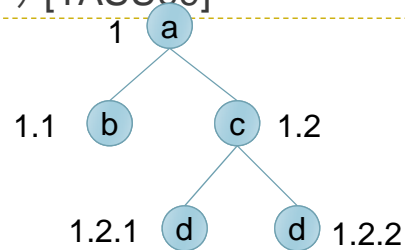
## 関係XMLデータベース モデル写像(経路アプローチ) [YASU99]

- ▶ 木構造の表現
  - ▶ 範囲ラベル, Deweyオーダ
- ▶ 検索高速化の工夫
  - ▶ パス情報
  - ▶ 親, 子等々へのリンク

問合せ: /a[c]/b

```
SELECT nb.*
FROM Node na, Node nb, Node nc
WHERE na.path = '/a'
AND nb.path = '/a/b'
AND nc.path = '/a/c'
AND nb.label contains na.label
AND nc.label contains na.label
```

▶ 16



ノード

ID	Name	Label	Path
1	a	1	/a
2	b	1.1	/a/b
3	c	1.2	/a/c
4	d	1.2.1	/a/c/d
5	d	1.2.2	/a/c/d

## ネイティブXMLデータベース 既存のシステム(一部)

- ▶ 商用
  - ▶ Tamino (Software AG)
  - ▶ NeoCore XMS (Xpiori)
  - ▶ TX1 (東芝)
  - ▶ Shunsaku (富士通)
  - ▶ Cyber Luxeon (サイバーテック)
- ▶ オープンソース
  - ▶ Apache Xindice <http://xml.apache.org/xindice/>
  - ▶ eXist <http://exist.sourceforge.net/>
  - ▶ BaseX <http://basex.org/>
  - ▶ MonetDB/XQuery <http://monetdb.cwi.nl/XQuery/>
  - ▶ Sedna <http://modis.ispras.ru/sedna/>
  - ▶ Xbird

▶ 17



## XML対応関係データベース XML-enabled Relational Database Systems

- ▶ ストレージ
  - ▶ 関係ストレージ+XML専用ストレージ
  - ▶ 関係データとXMLデータの混在
    - ▶ XMLスキーマから関係表への写像
    - ▶ ネイティブXMLデータ型のサポート
- ▶ 問合せ言語
  - ▶ XPath / XQuery
  - ▶ SQLへのXPath/XQueryの埋め込み (SQL/XML)
- ▶ 両者の垣根があいまいに...

▶ 18

## XMLデータベース問合せ言語 XQuery 1.0: XML Query Language

19

## XQuery 1.0: XML Query Language

- ▶ XMLのための問合せ言語
  - ▶ 静的型付き関数型言語
- ▶ XPath 2.0の上位言語
  - ▶ XPath 2.0: ノードの位置を特定
  - ▶ XQuery 1.0: XML情報源のための照会言語

### 注意点

XPath 2.0とXPath 1.0との基本的な相違

XPath 1.0の問合せ結果: 重複のないノード集合

XPath 2.0の問合せ結果: 順序付きのノードシーケンス(重複あり)

▶ 20

## XQuery vs SQL

- ▶ XQueryは集合操作言語
  - ▶ 集合に項目を割り当てる
  - ▶ 集合から項目を取り出す
  - ▶ SQLと類似性

	XQuery 1.0	SQL
構文	FLWOR (For, Let, Where, Order by, Return)	SELECT-FROM-WHERE
理論的背景	形式的意味論 (formal semantics)	関係代数・関係論理 (一階述語論理)

▶ 21

## サンプルXML

```
<?xml version="1.0"?>
<bib>
  <book year="1994">
    <title>TCP/IP Illustrated</title>
    <author>Stevens</author>
    <publisher>Addison-Wesley</publisher>
  </book>
  <book year="1992">
    <title>Advanced Unix Programming</title>
    <author>Stevens</author>
    <publisher>Addison-Wesley</publisher>
  </book>
  <book year="2000">
    <title>Data on the Web</title>
    <author>Abiteboul</author>
    <author>Buneman</author>
    <author>Suciu</author>
  </book>
</bib>
```

▶ 22

## FLWOR構文

### ▶ 例

```
for $b in doc("books.xml")//book
where $b/@year < "2000"
order by $b/title
return $b/title
```

- ▶ for句: タプルストリームを生成し、各タプルを変数に割り当てる
- ▶ let句: タプルストリーム全体を、変数に割り当てる
- ▶ where句: 条件式によるタプルのフィルタリング
- ▶ order by句: タプルストリームの要素を整列
- ▶ return句: 各タプルに対し、FLWOR式の評価結果を生成

▶ 23

## 要素構築子

- ▶ XML中にXQueryを埋め込む

```
<bib>
{
  for $b in doc("book.xml")//book
  where $b/@year < 2000
  return <t1>{$b/title/text()}</t1>
}
</bib>
```

- ▶ 以下のように書いてもよい

```
element bib {
  for $b in doc("book.xml")//book
  where $b/@year < 2000
  return element t1 {
    $b/title/text()
  }
}
```

▶ 24

## ジョイン

- ▶ 複数のXMLデータを結合
  - ▶ 入力XMLデータを変数に割り当て (for, let)
  - ▶ 結合の条件をwhere句で与える
- ▶ 例:

```
for $b1 in doc("book1.xml")//book,
    $b2 in doc("book2.xml")//book
where $b1/author = $b2/author
and $b1/title != $b2/title
return <pair>{$b1/title, $b2/title}</pair>
```

▶ 25

## 入れ子問合せ

- ▶ XQuery式はタプルシーケンスを返す
- ▶ そのまま別のXQuery式の入力に使える
- ▶ 例:

```
for $a in let $x := distinct-values(doc("book.xml")//author)
         return $x
return <author name="{ $a }">{
  for $b in doc("book.xml")//book
  where $b/author = $a
  return $b/title
}</author>
```

- ▶ ご利用はほどほどに...

▶ 26

## XMLデータの更新 XQuery Update Facility 1.0

- ▶ XMLデータ更新のための言語
  - ▶ W3C Candidate Recommendation 1 August 2008
- ▶ 拡張構文
  - ▶ Insert, Delete, Replace, Rename, Transform
- ▶ 例:
 

```
insert node <year>2005</year> after
fn:doc("bib.xml")/books/book[1]/publisher
```
- ▶ XUpdate (XML Update Language)
  - ▶ XML:DB Initiativeによる開発
  - ▶ いくつかの実装
    - ▶ Sedna

▶ 27

## XMLの全文検索 XQuery and XPath Full Text 1.0

- ▶ 部分文字列検索との違い
  - ▶ トークンやフレーズを検索
  - ▶ 言語検索
    - ▶ “mouse” に対して, “mouse” や “mice” が検索される
  - ▶ ランキング
- ▶ 例:
  - ▶ `for $b in /books/book where $b/title  
ftcontains ("dog" with stemming) ftand "cat"  
return $b/author`

▶ 28

## XQuery 1.1

- ▶ XQuery 1.0 との後方互換性
- ▶ 拡張
  - ▶ 値に基づくグルーピング
  - ▶ シーケンス中の位置に基づくグルーピング
  - ▶ シーケンス上のウィンドウイング
  - ▶ エラー処理とリカバリ
  - ▶ 数値, 日付情報のフォーマッティング
  - ▶ 名前空間の動的割り当て
  - ▶ 外部結合 (outer for)
  - ▶ 高階関数のサポート
  - ▶ ...

▶ 29

## SQL/XMLによるデータベースアクセス

30

### SQL/XMLについて

---

- ▶ SQL:2003におけるXML拡張
  - ▶ SQLでのXMLデータの取り扱いを可能に
  - ▶ SQL中へのXML (XPath, XQuery) の埋め込み
  - ▶ XML上の関係ビュー
- ▶ 拡張の内容
  - ▶ XMLデータ型
  - ▶ 関数
  - ▶ XMLデータ型からSQLデータ型へのマッピング
  - ▶ XMLを扱うための専用構文
- ▶ 参考
  - ▶ <http://sqlx.org>

---

▶ 31

## SQL/XMLの利用 問合せ

### ▶ 関係表の作成

```
CREATE TABLE book (bookid INTEGER, content XML)
```

### ▶ データの登録

```
INSERT INTO book VALUES (1, '<book>...</book>')
```

### ▶ データの検索

```
SELECT bookid, XMLQuery(
  'for $b in /bib/book
   where $b/@year = 2000
   return $b/title'
  PASSING BY VALUE content
  RETURNING CONTENT) XMLData
FROM book;
```

▶ 32

## SQL/XMLの利用 XMLTable

### ▶ XQuery式の評価結果を関係表に見せる

```
<dept bldg="114">
  <employee id="903">
    <name>
      <first>Mary</first>
      <last>Jones</last>
    </name>
    <office>415</office>
    <phone>905-402-6112</phone>
    <phone>SELECT X.*
    <salary>FROM emp,
  </employee>XMLTABLE ('$d/dept/employee' passing doc as "d"
  </dept>COLUMNS
  empID INTEGER PATH '@id',
  firstname VARCHAR(20) PATH 'name/first',
  lastname VARCHAR(25) PATH 'name/last') AS X
```

empID	firstname	lastname
901	John	Doe
902	Peter	Pan
903	Mary	Jones

▶ 33 <http://www.ibm.com/developerworks/jp/xml/library/techarticle/dm-0708nicola/>



## XMLデータベースアクセスインタフェース XML:DB

34

### XMLデータベースへのアクセス XML:DB

- ▶ XML:DB Initiativeにより策定
- ▶ JavaからXMLデータベースにアクセスするための標準APIを提供
  - ▶ 関係データベースにおけるJDBCに相当
- ▶ 多くのXMLデータベースシステムがサポート

XML Database	License	Support	Notes
<a href="#">Apache Xindice</a>	Open source, free	Yes	Xindice Version 1.1 released 9th May 2007
<a href="#">BaseX</a>	Open source, free	Yes	Supported since version 4.2
<a href="#">Gemfire Enterprise</a>	Commercial	Yes	
<a href="#">DOMSafeXML</a>	Commercial	Yes	
<a href="#">eXist</a>	Open source, free	Yes	TransactionService (ACID Transactions) not supported
<a href="#">MonetDB/XQuery</a>	Open source, free	Yes	
<a href="#">myXMLDB</a>	Open source, free	Yes	Works on top of MySQL; seems to have been discontinued long ago.
<a href="#">OZONE</a>	Open source, free	Yes	100% support including TransactionService (ACID).
<a href="#">Safina</a>	Open source, free	Yes	100% support, including TransactionService (ACID).
<a href="#">Software AG's Tamino</a>	Commercial	Partial	Lacks update support, XUpdateQueryService, or any other Update language not implemented

▶ 35 [http://en.wikipedia.org/wiki/XML\\_database](http://en.wikipedia.org/wiki/XML_database)

## XML:DBによるデータベースへのアクセス手順

- ▶ Database実装クラスの登録
- ▶ コレクションの取得
- ▶ XPathQueryServiceの取得
- ▶ ネームスペース設定
- ▶ 検索を実行
- ▶ ResourceSetから検索結果を取り出す
- ▶ コレクションのクローズ

### ▶ 参考

<http://www.atmarkit.co.jp/fxml/tanpatsu/18xindice/xindice03.html>

▶ 36

おわりに

37

## おわりに

- ▶ 実用的なXMLデータベースの出現
  - ▶ 商用・オープンソース
- ▶ 主要関係データベースシステムのXMLネイティブサポート
  - ▶ 商用・オープンソース
  
- ▶ ウェブインテリジェンスでのXMLの利活用
  - ➔XMLデータベースを活用して, 研究活動を効率化!!

▶ 38

## 参考文献

- ▶ [FK99] D. Florescu, and D. Kossman, A performance evaluation of alternative mapping schemes for storing XML data in a relational database, Rapport de Recherche No. 3680 INRIA, Rocquencourt, France, May 1999.
- ▶ [HLM03] P. J. Harding, Q. Li, B. Moon. "XISS/R: XML Indexing and Storage System using RDBMS," Proc. VLDB, 2003.
- ▶ [BTS02] K. Beyer, I. Tatarinov, J. Shanmugasundaram, "Storing and Querying Ordered XML Using a Relational Database System," Proc. ACM SIGMOD, 2002.
- ▶ [YASU01] M. Yoshikawa, T. Amagasa, T. Shimura, S. Uemura, "XRel: a path-based approach to storage and retrieval of XML documents using relational databases," ACM TODS, 1(1):110-141, 2001.
- ▶ [BFH+02] P. Bohannon, J. Freirey, J. R. Haritsa, M. Ramanath, P. Royy, Jerome Simeon, "LegoDB: Customizing Relational Storage for XML Documents," Proc. VLDB, 2002.

▶ 39