Constraint Checking using DB2 pureXML and DataPower
An Evaluation based on the Healthcare Environment

Christian Pichler

Motivation

Constraint Classification

- Type I: Structural Requirements
- Type II: Reference data within XML document instances
- Type III: Reference data across XML document instances
- Type IV: Reference data between XML document instance and non-XML data source

Constraint Notations and Technology

- Notations
  - XML Schema: defines the structure of XML document instances
  - XQuery: queries the content of XML document instances
  - Schematron: defines assertions on the content of XML document instances

- Technology
  - DB2 pureXML: database system managing relational and XML data natively
  - Data Web Services: allow to expose database operations as Web Services
  - WebSphere DataPower SOA Appliance: hardware to process XML

Problems, Goals and Requirements

- Problems:
  - Healthcare information represented electronically may contain erroneous information
  - Humans can't evaluate masses of healthcare information

- Goal:
  - Identify sources where erroneous information may occur
  - Identify technology, including hardware and software, that can be used to validate healthcare information to avoid erroneous information

- Requirements:
  - Constraint definitions should be separated from application and data
  - Implementation should be in an SOA Environment
  - Quality of Error messages that result from validation failures should be improved

Evaluation

Architecture of Approach 1 and Approach 2 utilizing DB2 pureXML

- Client
  - SOAP/REST
  - Application Server
  - Data Web Services

Constraint Validation and Data Store

DB2 pureXML and XQuery / Schematron

Architecture of Approach 3 utilizing WebSphere DataPower SOA Appliance

- Client
  - SOAP/REST
  - Application Server
  - Data Web Services

WebSphere DataPower SOA Appliance

Constraint Validation

DB2 pureXML

Comparison

<table>
<thead>
<tr>
<th>Constraint Definition</th>
<th>DB2 pureXML (Approach 1)</th>
<th>DB2 pureXML (Approach 2)</th>
<th>WebSphere DataPower SOA Appliance (Approach 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraint Types Covered</td>
<td>XML Schema and XQuery</td>
<td>XML Schema and Schematron</td>
<td>XML Schema and Schematron</td>
</tr>
<tr>
<td>Type I</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Type II</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Type III</td>
<td>✔</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Type IV</td>
<td>✔</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Constraint Mechanism</td>
<td>implementation of diverse stored procedures</td>
<td>implementation of diverse stored procedures</td>
<td>configuration only</td>
</tr>
<tr>
<td>Constraint Validation</td>
<td>parallel</td>
<td>sequential</td>
<td>sequential</td>
</tr>
<tr>
<td>Constraint Violations</td>
<td>store valid as well as invalid information</td>
<td>store valid as well as invalid information</td>
<td>store valid as well as invalid information</td>
</tr>
<tr>
<td>Error Messages</td>
<td>customizable</td>
<td>customizable</td>
<td>customizable</td>
</tr>
</tbody>
</table>

Outlook

- Other industries next to the healthcare environment also define specific XML based formats for information exchange (e.g. finance, government, business)
- Trend to store information in the same format as it is exchanged (“what you exchange is what you store”)
- Recommendation to pursue the development of constraint notations that may be applied to various industry formats
- Development of Schematron extension that allows to define Type III and Type IV constraints

Literature