Querying XML

Paul Cotton

SC32 Plenary

Santa Fe, Jan 24, 2000

Biography

◆ IBM representative to CAC SC32

◆ ISO SQL/MM Full-Text Editor

◆ DB2 architect, IBM Toronto Lab

◆ XML Activity participant since 1998

◆ Chairman of W3C XML Query WG

Organization of Presentation

- ◆ Example usage scenarios
- ◆ XML query history
- ◆ Query Languages '98 workshop
- ◆ Candidate requirements
- ◆ W3C XML Query working group
- ◆ Questions

Example usage scenarios

- XML document management
- XML processing of DBMS data
- Filtering of XML messages
- Integration of multiple XML sources
- Existing W3C query facilities
- Other scenarios

XML document management

- ◆ Documents structured using XML

- ◆ Collections of documents
  - – Technical manuals

- ◆ Query facilities
  - – Retrieve individual documents or fragments
  - – Generate new XML output
  - – Text search operations including ranking

XML processing of DBMS data

- ◆ DBMS data
  - – SQL databases
  - – Object-oriented databases
  - – Native XML repositories
- ◆ XML can be physical or virtual
- ◆ Queries facilities
  - – Operations on text and non-text data types
  - – Operations on object types
  - – Aggregations

Filtering of XML messages

◆ Messages formatted using XML

◆ Streams with multiple sources/sinks

◆ Software-based message routing

◆ Query facilities

   – Message selection and filtering

   – Message transformation

   – Text and non-text data type operations

Integration of multiple XML sources

◆ Integration of document-oriented and data-oriented data sources

◆ More than single document processing

◆ Queries facilities
  – Data integration
  – Joining of data in multiple sources

Existing W3C query facilities

◆ DOM Iterators or TreeWalkers

◆ XPath queries as used by XPointer and XSLT

◆ IETF DAV Searching and Location (DASL)

Other scenarios

◆ Your favorite XML search/query scenario

XML query history

- ◆ Early 1998: "roll your own query language"
- ◆ XSL Working Group
  - – XSLT needed syntax to select nodes
- ◆ XML Linking Working Group
  - – XPointer needed syntax to select a location
- ◆ February 1999 joint meeting
  - – Rapprochement on 90% of requirements
- ◆ XPath
  - – W3C recommendation with XSLT

XML query history - 2

- ◆ Early queries facilities for SGML
- ◆ Academic research into semi-structured data and its operations
- ◆ XQL: See http://metalab.unc.edu/xql
- ◆ XML-QL, August, 1998
  - – http://www.w3.org/TR/NOTE-xml-ql/

Query Languages Workshop '98

◆ W3C sponsored workshop

◆ Boston (USA), December 2-3, 1998

◆ 98 participants: W3C members, database vendors, invited experts,

◆ 66 position papers

◆ See: http://www.w3.org/TandS/QL/QL98

Candidate requirements

◆ **QL'98 workshop summary**

– Candidate Requirements for XML Query, Paul Cotton and Ashok Malhotra, IBM

– http://www.w3.org/TandS/QL/QL98/pp/queryreq.html

◆ **See also:**

– Database Desiderata for an XML Query Language, David Maier, Oregon Graduate Institute

– http://www.w3.org/TandS/QL/QL98/pp/maier.html

Query language and structure

◆ Non-procedural query language

◆ XML syntax for query language

◆ Build on syntax used by other XML standards

◆ Ability to transmit a query in a URL

◆ Queries should be XPointer/XLink aware

◆ Uniform support for elements and attributes

Query language facilities

◆ Support for query operations
- selection of a document or element based on content, structure or attribute values
- extraction of particular elements
- reduction: removing sub-elements
- restructuring: construct new elements
- combination: merging of elements
- joins across data sources

◆ Support for insert, update and delete

Query language facilities - 2

◆ Support for nested queries and closure

◆ Support for full-text queries

◆ Facilities to construct XML documents

Query data sources

◆ Ability to query multiple documents

◆ Ability to query distributed data stored in a variety of formats such as SQL and OO databases. XML query must be translatable to query facility for underlying data.

◆ Ability to create XML schemas from non-XML data sources

◆ Support for "live" data.

Using the XML query language

- ◆ Query should be usable on a document without a schema

- ◆ Presence of a schema should permit query validation

- ◆ Support for local environment variables

- ◆ Possible to run queries in several different environments/contexts

XML query language semantics

- ◆ Precise semantics
  - – Query equivalence
  - – Query containment

- ◆ Compositional semantics
  - – Referential transparency

- ◆ Provided by data model of XML, set of query operators and then a query syntax

W3C XML Query WG - History

- July 1999 -Working Group proposed as part of XML Activity Phase 3 rechartering

- September - Call for WG participation

- September - WG constituted and chartered

- More than 30 W3C member companies

- September 15-16 - Initial F2F meeting

- December 2-3 - Second F2F meeting

W3C XML Query WG - Status

- ◆ Requirements document to be submitted as a W3C Working Draft

- ◆ Regular F2F meetings

- ◆ Weekly teleconference calls

- ◆ Public working drafts every three months

- ◆ Proposed recommendation(s)

XML Query WG relationships

- ◆ Existing W3C recommendations
  - – XPath and XSLT
  - – DOM
  - – Internationalization
- ◆ Emerging W3C recommendations
  - – Infoset
  - – XML Schema
  - – XPointer and XLink
- ◆ IETF DASL

Questions

◆ Today

◆ Public email list: www-ql@w3.org

◆ Later - cotton@ca.ibm.com