

Querying XML Documents

Querying XML Documents

Paul Cotton, Microsoft Canada
Brisbane, Australia
August, 2003

Paul Cotton, Microsoft

1

Brisbane, August 2003

Organization of Presentation

- ◆ XML query history
- ◆ XML Query WG history, goals and status
- ◆ XML Query working drafts
- ◆ XQuery overview
- ◆ XQuery status and outstanding issues
- ◆ Questions

Paul Cotton, Microsoft

2

Brisbane, August 2003

XML query history

- ◆ Early queries facilities for SGML
- ◆ 1998: “roll your own query language”
- ◆ Feb 1998: XQL proposal
 - <http://metalab.unc.edu/xql>
- ◆ Aug 1998: XML-QL submission
 - <http://www.w3.org/TR/NOTE-xml-ql/>
- ◆ Dec 1998: W3C QL'98 Workshop
 - <http://www.w3.org/TandS/QL/QL98>
- ◆ Nov 1999: XPath Recommendation
 - <http://www.w3.org/TR/xpath>

Paul Cotton, Microsoft

3

Brisbane, August 2003

W3C XML Query WG - History

- ◆ July 1999: Working Group proposed as part of XML Activity Phase 3 rechartering
- ◆ Sept 1999: WG chartered and first F2F
- ◆ July 2002: Re-chartered with XML Activity
- ◆ Currently 26 W3C member companies
23 F2F meetings and 150+ telcons so far
- ◆ Public WDs every three months
- ◆ Two Last Call WDs in May 2003
- ◆ Seven documents on Recommendation track

Paul Cotton, Microsoft

4

Brisbane, August 2003

W3C XML Query WG - Goals

- ◆ “The goal of the XML Query WG is to produce a data model for XML documents, a set of query operators on that data model, and a query language based on these query operators.”

Paul Cotton, Microsoft

5

Brisbane, August 2003

XML Query Requirements WD

- ◆ General Requirements
 - Non-procedural query language
 - XML syntax for query language but also a readable syntax
 - Protocol independent
 - Standard error conditions
 - Future support for updates
- ◆ XML Query Data Model Requirements
 - Built on XML Infoset and PSVI
 - Namespace aware
 - Support for XML Schema data types
 - Support for inter- and intra- document references (NOT in V1)

Paul Cotton, Microsoft

6

Brisbane, August 2003

Querying XML Documents

XML Query Requirements WD

- ◆ XML Query Functionality
 - Operators on all data types
 - Text operators across element boundaries
 - Support for hierarchy and sequence
 - Ability to combine data from different locs
 - Aggregation and sorting
 - Combination of operators including queries as operands
 - Support for NULL/empty values
 - Structural preservations
 - Identity preservation
 - Operations on names
 - Operations on "schemas" (NOT in V1)
 - Extensibility
 - Closure

Paul Cotton, Microsoft

7

Brisbane, August 2003

XML Query Use Cases WD

- ◆ Use Case Organization
 - Description, DTD/Schema, Input Data, Queries and Results
- ◆ Current Use Cases
 - "XMP": Experiences and Exemplars
 - "TREE": Queries that preserve hierarchy
 - "SEQ" - Queries based on Sequence
 - "R" - Access to Relational Data
 - "SGML" - Queries from SGML Open 1992
 - "STRING": String Search
 - "NS" - Queries Using Namespaces
 - "PARTS" - Recursive Parts Explosion
 - "STRONG" - Queries that exploit strongly typed data

Paul Cotton, Microsoft

8

Brisbane, August 2003

XML Query 1.0 Data Model WD

- ◆ Defines what information is available to an XML Query 1.0, XSLT 2.0 or XPath 2.0 processor
- ◆ Published jointly with XSL Working Group
- ◆ Infoset plus the following:
 - Support for XML Schema data types (PSVI)
 - Support for document collections
- ◆ Node-labelled tree constructor model with node identity
- ◆ Mapping from Infoset to Query Data Model uses Infoset terminology and is shown by example

Paul Cotton, Microsoft

9

Brisbane, August 2003

XML Query 1.0 Formal Semantics WD

- ◆ XML Query Formal Semantics is used:
 - to define XQuery and XPath static typing
 - to define XQuery and XPath runtime semantics
- ◆ FS defines both static and dynamic semantics
 - static semantics are presented as type inference rules, which relate XQuery/FS expressions to types
 - dynamic, or operational, semantics are presented as value inference rules, which relate XQuery/FS expressions to values in the XML Query Data Model

Paul Cotton, Microsoft

10

Brisbane, August 2003

XQuery: A Query Language for XML

- ◆ XQuery is a functional language in which a query is represented as an expression
- ◆ XQuery expressions can be nested with full generality
- ◆ The input and output of an XQuery are instances of the XML Query Data Model
- ◆ Based on OQL, SQL, XML-QL, XPath
- ◆ Readable vs. XML syntax

Paul Cotton, Microsoft

11

Brisbane, August 2003

XQueryX

- ◆ XQueryX is an XML representation of an XQuery
- ◆ It was created by mapping the productions of the XQuery BNF directly into XML productions
- ◆ XQueryX useful to enable:
 - Parser reuse
 - Queries on queries
 - Generation of queries
 - Embedding of queries in XML documents
- ◆ XQueryX status - Issue 152

Paul Cotton, Microsoft

12

Brisbane, August 2003

Querying XML Documents

XQuery Function and Operators WD

- ◆ XML Query Functions and Operators is used:
 - to defines basic operators and functions on the datatypes defined in XML Schema Part 2: Datatypes for use in XQuery, XPath, XSLT and other related XML standards
 - to define operators and functions on nodes and node sequences as defined in the XQuery 1.0 and XPath 2.0 Data Model for use in XQuery, XPath, XSLT and other related XML standards
- ◆ Includes several new data types derived from XML Schema data types

Paul Cotton, Microsoft

13

Brisbane, August 2003

XSLT and XQuery Serialization WD

- ◆ XSLT and XQuery Serialization WD
 - Based on material from XSLT 1.0
 - Now common to XSLT 2.0 and XQuery 1.0
 - Defines how to convert a Data Model instance into a set of octets in XML, HTML, XHTML or text

Paul Cotton, Microsoft

14

Brisbane, August 2003

XQuery Expressions

- ◆ XQuery expressions
 - Path expressions
 - Element constructors
 - FLWOR expressions
 - Expressions involving operators and functions
 - Conditional expressions
 - Quantified expressions
 - List constructors
 - Expressions that test or modify datatypes

Paul Cotton, Microsoft

15

Brisbane, August 2003

XQuery Path Expressions

- ◆ Based on abbreviated syntax of XPath
- ◆ Select the second section of the fifth chapter of the doc /doc/chapter[5]/section[2]

Paul Cotton, Microsoft

16

Brisbane, August 2003

XQuery Element Constructors

- ◆ XQuery element constructor consists of a start tag and an end tag, enclosing an optional list of expressions that provide the content of the element.
- ◆ **Generate an <emp>element that has an “empid” attribute. The value of the attribute and the content of the element are specified by variables that are bound in other parts of the query.**

```
<emp empid = {$id}>
  {$name}
  {$job}
</emp>
```

Paul Cotton, Microsoft

17

Brisbane, August 2003

XQuery FLWOR Expressions

- ◆ A FLWOR expression binds some expressions, applies an optional predicate and ordering, and constructs a new result.



FOR and LET clauses generate a list of tuples of bound expressions, preserving document order.

WHERE clause applies a predicate, eliminating some of the tuples

RETURN clause is executed for each surviving tuple, generating an ordered list of outputs

Paul Cotton, Microsoft

18

Brisbane, August 2003

Querying XML Documents

XQuery FLWOR Expressions

- ◆ List the titles of books published by Morgan Kaufmann in 1998.

```
FOR $b IN doc("bib.xml")//book
WHERE $b/publisher = "Morgan Kaufmann"
  AND $b/year = "1998"
RETURN $b/title
```

- ◆ List each publisher and the average price of its books.

```
FOR $p IN distinct(doc("bib.xml")//publisher)
LET $a := avg(doc("bib.xml")
  /book[publisher = $p]/price)
RETURN
  <publisher>
    <name> {$p/text()} </name>
    <avgprice> {$a} </avgprice>
  </publisher>
```

Paul Cotton, Microsoft

19

Brisbane, August 2003

XQuery Operators and Functions

- ◆ Infix and prefix operators
- ◆ Parenthesized expressions
- ◆ Arithmetic and logical operators
- ◆ Sequence operators UNION, INTERSECT and EXCEPT
- ◆ Functions can be defined in XQuery
- ◆ May 2003 WDs support definition and import of modules of functions

Paul Cotton, Microsoft

20

Brisbane, August 2003

XQuery Operators and Functions

- ◆ Find the maximum depth of the document named "partlist.xml."

```
namespace xsd="http://www.w3.org/2001/XMLSchema-datatypes"

DEFINE FUNCTION depth($e AS ELEMENT) AS xsd:integer
{
  {-- An empty element has depth 1 --}
  {-- Otherwise, add 1 to max depth of children --}
  IF empty($e/*) THEN 1
  ELSE max(FOR $c IN $e/* RETURN depth($c)) + 1
}

depth(doc("partlist.xml"))
```

Paul Cotton, Microsoft

21

Brisbane, August 2003

XQuery Conditional Expressions

- ◆ IF THEN ELSE construct
- ◆ Make a list of holdings. For journals, include the editor, and for all other holdings, include the author.

```
FOR $h IN //holding
RETURN
  <holding>
    {$h/title,
     IF $h/@type = "Journal" THEN
       $h/editor
     ELSE
       $h/author
    }
  </holding>
```

Paul Cotton, Microsoft

22

Brisbane, August 2003

XQuery Quantified Expressions

- ◆ Existential and Universal quantifiers
- ◆ Find titles of books in which both sailing and windsurfing are mentioned in the same paragraph.

```
FOR $b IN //book
WHERE SOME $p IN $b//para SATISFIES
  contains($p, "sailing")
  AND contains($p, "windsurfing")
RETURN $b/title
```

- ◆ Find titles of books in which sailing is mentioned in every paragraph.

```
FOR $b IN //book
WHERE EVERY $p IN $b//para SATISFIES
  contains($p, "sailing")
RETURN $b/title
```

Paul Cotton, Microsoft

23

Brisbane, August 2003

Sequence-related Operators

- ◆ A sequence may be constructed by enclosing zero or more expressions separated by commas.
- ◆ For example: (\$x, \$y, \$z) denotes a sequence containing three members represented by variables
- ◆ >> (precedes) and << (follows) boolean functions
- ◆ () denotes an empty sequence.

Paul Cotton, Microsoft

24

Brisbane, August 2003

Querying XML Documents

XQuery Operators on Data Types

- ◆ INSTANCE OF returns True if its first operand is an instance of the type named in its second operand
- ◆ CAST is used to convert a value from one data type to another
- ◆ TREAT AS causes the query processor to treat an expression as though its data type were a subtype of its static type
- ◆ VALIDATE executes XML Schema validation

Paul Cotton, Microsoft

25

Brisbane, August 2003

W3C XML Query WG - Status

- ◆ Requirements (May/June 2003)
 - XML Query (XQuery) Requirements
<http://www.w3.org/TR/xquery-requirements>
 - XQuery and XPath Full-Text Requirements
<http://www.w3.org/TR/xquery-full-text-requirements/>
- ◆ XQueryX (June 2001)
 - XML Syntax for XQuery 1.0: XQueryX
<http://www.w3.org/TR/xqueryx>

Paul Cotton, Microsoft

26

Brisbane, August 2003

W3C XML Query WG - Status

- ◆ Last Call Working Drafts (May 2 2003)
 - XML Query 1.0 and XPath 2.0 Functions and Operators
<http://www.w3.org/TR/xpath-functions/>
 - XML Query 1.0 and XPath 2.0 Data Model
<http://www.w3.org/TR/xpath-datamodel/>

Paul Cotton, Microsoft

27

Brisbane, August 2003

W3C XML Query WG - Status

- ◆ Working Drafts (May 2 2003)
 - XQuery 1.0: An XML Query Language
<http://www.w3.org/TR/xquery/>
 - XML Query Use Cases
<http://www.w3.org/TR/xmlquery-use-cases/>
 - XML Path Language (XPath) 2.0
<http://www.w3.org/TR/xpath20/>
 - XML Query 1.0 and XPath 2.0 Formal Semantics
<http://www.w3.org/TR/xquery-semantics/>
 - XSLT 2.0 and XQuery 1.0 Serialization
<http://www.w3.org/TR/xslt-xquery-serialization/>
- ◆ Next publication status
- ◆ WG Charter status
<http://www.w3.org/2001/12/xmlbp/xml-query-wg-charter.html>

Paul Cotton, Microsoft

28

Brisbane, August 2003

Current XQuery Issues

- ◆ Need to close 5 issues to get XQuery/XPath to Last Call
 - Issue 152. What is best XML syntax for XQuery?
 - Issue 307: Schema Types from input documents?
 - Issue 546. What conformance levels?
 - Issue 554. What is normative text?
 - Issue 564: Loading same schema-component twice
- ◆ Need to close a further 6 issues to get Formal Semantics to Last Call
 - Issue 481 Semantics of Schema Context
 - Issue 496 Support for lax and strict wildcards
 - Issue 555 Formal Semantics of Module Import
 - Issue 556 Formal Semantics of Variable Definitions
 - Issue 557 Formal semantics of Validation Declaration
 - Issue 559 New Sequence Type needs to be fully implemented in FS

Paul Cotton, Microsoft

29

Brisbane, August 2003

Current XQuery Issues

- ◆ Solutions in Aug 2003 WDs
 - Issue 319. Namespace definitions and in-scope namespaces
 - Issue 340. How should errors be represented?
 - Issue 343. Do functions in the null namespace clash with functions in the default namespace?
 - Issue 475 FS-Issue-0132: Typing for descendant
 - Issue 508. Namespaces in element constructors
 - Issue 527 Static typing of XPath index expressions
 - Issues 547, 548, and 550. Grammar issues
 - Issue 560 Exactness of Type Inference
- ◆ Future work
 - XQuery test suite
 - Support for an update language
 - Support for full-text retrieval

Paul Cotton, Microsoft

30

Brisbane, August 2003

Querying XML Documents

Full-Text Support in XQuery

- ◆ Full-Text issues
 - history within XML Query WG
 - Library of Congress Use Case
http://lcweb.loc.gov/crsinfo/xml/lc_usecases.html
 - related to I18N issues
 - Is cross-language definition of characters, words, sentences or paragraphs feasible?
 - XQuery and XPath Full-Text Requirements
<http://www.w3.org/TR/xquery-full-text-requirements>
 - XQuery and XPath Full-Text Use Cases
<http://www.w3.org/TR/xmlquery-full-text-use-cases/>
 - Target is XML Query 1.1

Paul Cotton, Microsoft

31

Brisbane, August 2003

Early XQuery implementations

- ◆ CL-XML http://homepage.mac.com/james_anderson/XML/index.html
- ◆ Enosys Markets <http://www.enosysmarkets.com/products/xq.html>
- ◆ Fatdog <http://www.fatdog.com/>
- ◆ Kawa-Query <http://www.gnu.org/software/kawa/xquery/>
- ◆ IPSI-XQ <http://xml.ipsi.fhg.de/xquerydemo>
- ◆ Lucent <http://db.bell-labs.com/galax/>
- ◆ Kweelt <http://db.cis.upenn.edu/Kweelt/>
- ◆ Microsoft <http://xqueryservices.com>
- ◆ Software AG
<http://www.softwareag.com/developer/downloads/default.htm>
- ◆ SourceForge <http://sourceforge.net/projects/xquench/>
- ◆ X-Hive <http://www.x-hive.com/xquery>
- ◆ XML Global <http://www.xmlglobal.com>
- ◆ and more ... (see <http://www.w3.org/XML/Query#products>)

Paul Cotton, Microsoft

32

Brisbane, August 2003

XQuery/XPath Grammars

- ◆ W3C's Grammar Test Pages
 - XQuery 1.0 Grammar Test Page
<http://www.w3.org/2003/05/applets/xqueryApplet.html>
 - XPath 2.0 Grammar Test Page
<http://www.w3.org/2003/05/applets/xpathApplet.html>

Paul Cotton, Microsoft

33

Brisbane, August 2003

Questions

- ◆ Today
- ◆ Later:
pcotton@microsoft.com
- ◆ Feedback email list:
public-qt-comments@w3.org
(archived at <http://lists.w3.org/Archives/Public/public-qt-comments/>).
- ◆ Public discussion email list:
www-ql@w3.org

Paul Cotton, Microsoft

34

Brisbane, August 2003