

# XQuery: Where Do We Go From Here?

Don Chamberlin

IBM Almaden Research Center

June 30, 2006

# Mindshare

---

- Google hits:
  - WQuery: 615
  - XQuery: 14,700,000
  - YQuery: 3,920
  - ZQuery: 686
- [www.w3.org/XML/Query](http://www.w3.org/XML/Query)  
lists 47 XQuery implementations
- [www.sqlsummit.com/XQueryProv.htm](http://www.sqlsummit.com/XQueryProv.htm)  
lists 53 XQuery implementations
- Amazon.com lists 8 books on XQuery

# Some university courses that cover XQuery

---

- Arizona State (CSE 494)
- Cornell (CS 433)
- ETH Zurich (ISK2004)
- Harvard (CSCIE 259)
- Nat'l Tsing Hua Univ., Taiwan (CS 6710)
- Stanford (CS 145)
- SUNY (CSE 636)
- Tech. Univ. of Denmark (02170)
- Technion, Israel (236363)
- U. of Hannover
- U. of Illinois, Urbana-Champaign
- U. of Penn. (CIS 550)
- U. of Pittsburgh (CS 1655)
- U. of Southern Calif. (CSCI 585)
- U. of Texas (CS 6360)
- U. of Utah (CS 5530)

# Resources for XQuery developers

---

- Websites maintaining wikis, tutorials, discussion groups, and shared code for XQuery developers
  - [xquery.com](http://xquery.com) (by Jason Hunter)
  - [xqzone.com](http://xqzone.com) (by Mark Logic)
  - [www.xqdev.com](http://www.xqdev.com) (by Ryan Grimm)
- xqDoc is a commenting convention (similar to JavaDoc) and a set of open-source tools for generating hyperlinked documentation from XQuery modules (by Darin McBeath) (see [xqDoc.org](http://xqDoc.org))
- xqLib is an open-source library of vendor-independent XQuery modules (example: for formatting dates) (see [xqdoc.org/xqlib.html](http://xqdoc.org/xqlib.html))

# XQuery resources (cont'd.)

---

- xq2xsl is an open-source tool that transforms XQuery into XSLT (by David Carlisle)  
(see [monet.nag.co.uk/xq2xml](http://monet.nag.co.uk/xq2xml))
- BumbleBee is a test harness for testing and comparing XQuery implementations (by Jason Hunter)  
(see [www.xquery.com/bumblebee](http://www.xquery.com/bumblebee))
- Stefano Ceri at Politecnico di Milano is working on "XQuery By Example" (see [dbgroup.elet.polimi.it/xquery/papers/XQBE-www03.pdf](http://dbgroup.elet.polimi.it/xquery/papers/XQBE-www03.pdf))
- Several XQuery-related blogs  
(Frank Cohen, Jeff Dexter, Jonathan Robie, . . . )

# XQuery development tools

---

- Stylus Studio (by Data Direct)
  - Syntax-directed editing and debugging (breakpoints, etc.)
  - "Drag and drop" from various data sources
  - Generates Java programs using XQJ
  - See [www.stylusstudio.com](http://www.stylusstudio.com)
- oXygen provides an XQuery plug-in for Eclipse
  - Syntax-directed editing, debugging, execution profiler
  - Currently uses Saxon and DataDirect XQuery engines
  - See [www.oxygenxml.com/xquery\\_debugger.html](http://www.oxygenxml.com/xquery_debugger.html)
- xqIde is an open-source XQuery Eclipse plug-in
  - See [xqdoc.org/xqide.html](http://xqdoc.org/xqide.html)
- BEA has a graphical XQuery editor
  - See [edocs.bea.com/aldsp/docs21/datasrvc/xqueryeditor.html](http://edocs.bea.com/aldsp/docs21/datasrvc/xqueryeditor.html)

# Cool XQuery application: SafariU

The screenshot shows a web browser window with the address bar displaying "safariu :: index - Mozilla Firefox". The browser's menu bar includes "File", "Edit", "View", "Go", "Bookmarks", "Tools", and "Help". The website header is yellow and features the "safariU v2" logo on the left and the tagline "Your Course-In Print, On The Web-Your Way." on the right. Below the header is a navigation menu with links for "BUILD", "BROWSE CONTENT", "MY PROJECTS", "ADOPT A BOOK", and "HELP". On the right side of the header, there are links for "Sign Out" and "My Account", and a personalized greeting "Welcome, stephen budton".

The main content area features a yellow-bordered box with the text: "Fast, elegant, and content-rich – SafariU v2 is here. Try the new version of SafariU. It's up to 10 times faster and the interface is more intuitive and flexible. We've added almost 600 new titles to the library, and there is much more – get the full details [here](#)."

Below this is a "SIGN UP AND GET STARTED." section with the text: "Join SafariU, at no cost to you, and enjoy flexibility and control like never before. Sign up now — create custom books and syllabi for your course use." and a red "REGISTER NOW" button.

To the right is a section titled "Create, Publish, and Share Customized Computer Science and Information Technology Course Materials." with a paragraph: "SafariU is a revolutionary way to teach exactly what you want to teach, how you want to teach it. With SafariU you can give your students more in-depth and focused content, far greater convenience, and unprecedented value. And do it all from your own computer, whenever you're ready." and a red "LAUNCH FLASH TOUR" button.

Below the sign-up section is a "MORE INFORMATION" section with links: "Launch [SafariU Flash tour](#)", "Download the [SafariU brochure](#)", "Select content from over [2800 books](#)", "Email [SafariU@oreilly.com](mailto:SafariU@oreilly.com)", and "Call 1.800.826.1938 or 707.829.5471".

At the bottom left is a "NEWS" section with a link: "[The Future of Textbooks: An Interview with John Preston](#)".

On the right side, there are three more sections, each with an icon and a red button: "CREATE YOUR CUSTOM TEXTBOOK" (with a red icon of an open book), "CREATE YOUR ONLINE SYLLABUS" (with a green icon of a bookshelf), and "EXCHANGE MATERIALS" (with a blue icon of a bookshelf).

www.safariu.com

# SafariU, continued

---

- Build your own textbook
- O'Reilly maintains a "content base" of more than 1.6 million pages of books, articles, images, etc.
- Professors can search and combine fragments from multiple sources, including uploaded material
- The resulting custom book is bound and delivered to the university or bookstore for 16 cents per page
- Profs can also create an on-line syllabus with links to many sources—students can subscribe for \$10/month
- Profs can upload and share learning materials: videos, web pages, Powerpoint slides, etc.
- SafariU is implemented in XQuery (Mark Logic) with a thin Java presentation layer.

# Other cool XQuery applications

---

- "Electronic flight bags"
  - Airlines must maintain about 1M pages of documentation on each individual aircraft
  - Part of this documentation must be onboard the aircraft
  - Thousands of updates per month
  - FAA has recently approved an all-electronic XML solution
  - Jet Blue and Northwest Airlines are using XQuery to maintain aircraft documentation
- [www.navio.com](http://www.navio.com)
  - "Rights-based" commerce platform for digital content
  - Music, games, videos, ring-tones, screensavers
  - Package, catalog and deliver content to any web device
  - Platform is implemented using XQuery (Ipedo)

# More cool XQuery applications

---

- Who earns more, Steve Jobs or Michael Dell?
  - Go to [paycheck.demo.marklogic.com](http://paycheck.demo.marklogic.com)
  - Mashup of information from public online data sources (SEC filings etc.)
  - Look up executives of about 200 public companies
  - Compare their salaries, bonuses, stock options, etc.
  - For each executive, get a job description and links to competitors
- NASA Astronomical Data Center
  - Go to [www.cogneticsystems.com:8080](http://www.cogneticsystems.com:8080)
  - Journal articles, data sets and other resources
  - Hyperlinked and searchable using XQuery

# Status

---

- XQuery 1.0 is a W3C Candidate Recommendation
- The W3C XQuery Test Suite is now available
  - See <http://www.w3.org/XML/Query/test-suite/>
  - 14,500 test cases
  - Final version expected July 7, 2006
  - Test results now available for 4 implementations (two demonstrate >95% conformance)
  - Test results solicited during six-week period
- The XML Query Working Group is still working:
  - Full-text extensions (working draft May 2006)
  - Update Facility (working draft May 2006)
  - "Last Call" for both expected Fall 2006
  - Follow-on query features: (error handling, grouping, . . . )

# What did we do right?

---

- Language operates on XML in its own data model
  - No need to transform XML into something else
  - Much less code than conventional XML apps
  - Rapid prototyping, apps are easy to build and change
- Declarative, functional language
  - Enables optimization
  - Enables function shipping and data shipping
- We took existing standards seriously
  - Superset of XPath
  - Uses type system of XML Schema

# What did we do right? (cont'd.)

---

- Gracefully integrates navigation with construction
- Powerful and complete
- Has an easy-to-learn subset
- Usable in many environments
  - Not bound to a specific persistence model
  - Can be used stand-alone or with various host languages
  - Can be used on either typed or untyped data
  - Can be used on databases, RSS feeds, etc.
  - Can integrate different kinds of data sources

# What did we do wrong?

---

- We took *way* too long
- We took existing standards seriously
  - Inherited all the complexity and foibles of XPath + Schema + Namespaces
  - = is not a transitive operator, so had to invent `eq` etc.
  - "Nillable"
  - `$x[$y]` might be a positional predicate, or might not
- Our syntax is fragile and sometimes ugly
  - No reserved words: `return` is a query
  - What is this? `delete union + 2`
  - Double-token approach: `do delete`

# What did we do wrong? (cont'd.)

---

- XQuery 1.0 is missing some important things:
  - updates
  - text search
  - error handling
  - grouping
  - dynamically computed namespace bindings
  - overloaded functions
- Constructors lack referential transparency
  - A problem for view definitions?

# Referential transparency

---

- Data: A library of newspaper articles

```
<library>
  <newspaper>
    <name> Chicago Tribune </name>
    <article>
      <date> 2006-06-30 </date>
      <title> Elvis Seen In Chicago </title>
      . . .
    </article>
    . . .
  </newspaper>
  . . .
</library>
```

# Referential transparency (cont'd.)

---

- Define a view:

```
define view $V as
  <elvis-articles>
    { /newspaper/article
      [contains(title, "Elvis")] }
  </elvis-articles>
```

- Query the view:

- List the newspapers that contain articles about Elvis on a given date:

```
$V/article[date = date("2006-06-30")] / ../name
```

- Fails because the view copies the articles
- Copied articles have a new parent
- A possible solution: use references, not copies

# Why did XQuery 1.0 take so long?

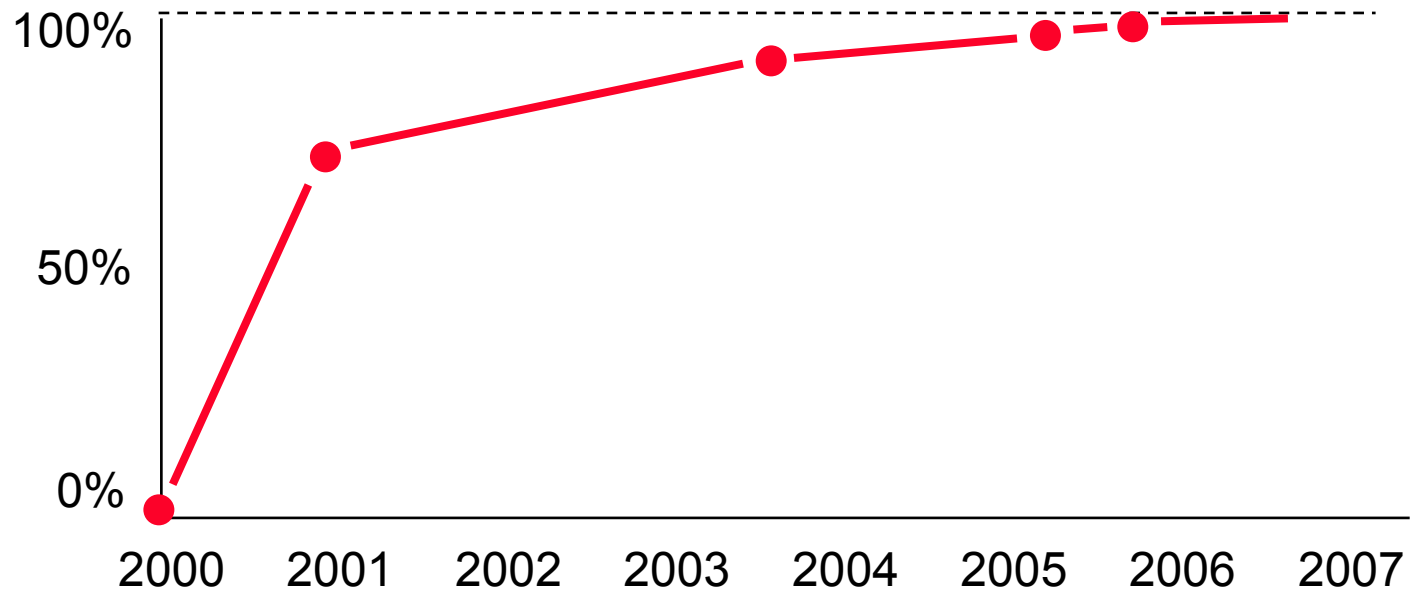
---

- Nov. 1999: First meeting of Query Working Group  
Most of 2000 spent on definition of data model and considering alternative approaches;  
Quilt selected as basis for XQuery in Nov. 2000
- Jan. 2001: First XQuery working draft **75%**  
Most of 2002 and 2003 spent on reconciling XPath with XML Schema and debating the type system.  
Named typing vs. structural typing  
What is this: `<a>12</a>`
- Nov. 2003: XQuery enters "Last Call" **90%**  
>1200 public comments
- April 2005: Second "Last Call" **95%**  
654 public comments

# Why did it take so long?

---

- Nov. 2005: XQuery 1.0 becomes Candidate Rec. **99+%**
  - 212 public comments so far
  - Creation of test suite with 14,500 test cases
- A W3C Recommendation by late 2006?



# What kinds of apps need XQuery?

---

- Querying and updating XML repositories
- Data integration
  - Merging heterogeneous data sources and services
  - Practically anything can be viewed as XML
  - Mashups
- Web services
  - XML in, XML out
  - Streams of messages, data transform pipelines
  - Merging and routing
  - Scripting language for business processes
- Web user interfaces (REST/AJAX environment)
  - Dynamically generate XHTML from XML databases
  - Like PHP + MySQL, without the impedance mismatch

# XQuery Implementations

---

- XML databases
  - Store XML with updates, indexing and search
  - Transaction semantics, recovery, replication, etc.
  - Some extend XQuery as a scripting language
  - Some include app servers: process incoming HTTP messages using XQuery apps, return results in XHTML
  - Examples: Mark Logic, X-Hive, eXist-db, Berkeley DB XML
- Hybrid relational/XML databases
  - Store XML in tables together with relational data
  - Call XQuery from SQL/XML to access XML data
  - Usually rely on host languages for business logic
  - Examples: IBM DB2 Viper, MS SQL Server, Oracle 10gR2

# XQuery Implementations (cont'd.)

---

- Data Integrators
  - Provide XML "wrappers" for many kinds of information
  - Use XQuery for merging and transformation
  - Support Service-Oriented Architectures
  - Often operate on streaming data
  - May use either XQuery or host language for business logic
  - Currently use many different API's (XQJ, JDBC, etc.)
  - Examples: BEA AquaLogic, Data Direct XQuery, Ipedo Enterprise Info. Integration, TigerLogic, Abacus
- Open-source XQuery language implementations
  - Examples: Saxon, Galax, Qexo, etc.

# Connecting XQuery to the world

---

- Many ways to connect XQuery to host languages
  - XQJ, JDBC, SQL/XML, REST, etc.
  - This is like the early days of SQL
- All these approaches suffer from "impedance mismatch"
  - Declarative vs. procedural; different type systems
  - Crossing interfaces adds complexity, hurts performance, prevents global optimization
- How the relational world dealt with this problem:
  - Programming extensions to SQL (PL/SQL, PSM, etc.)
  - Enabled many commercial apps to be written directly in SQL
- Could XQuery be extended as an app language?
  - YES. (See XQueryP paper in this conference)
  - Compared to PSM, benefits are greater and cost is smaller

# Example Query

---

- Data: Many books in "books.xml"

```
<book>  
  <title>Huckleberry Finn</title>  
  <author>Mark Twain</author>  
</book>
```

- Query: Find titles of books by a given author

```
doc("books.xml")/book[author = $x]/title
```

# Storing book data in a table

---

BOOK\_TABLE

BOOK_COLUMN
<book>
<book>
<book>

# SQL/XML, continued

---

```
conn = getConnection(db, user, pwd);
Qstring = "
    select xmlquery('
        $B/book/title
        ' passing books_column as B)
    from books_table
    where xmlexists('
        $B/book[author = $PARAM]
        ' passing books_column as B,
        ? as PARAM)
";
stmt = conn.prepareStatement(Qstring);
stmt.setString(1, "Mark Twain");
result = stmt.executeQuery();
while (result.next())
    System.out.println(result.getString(1));
```

# Does XQuery have a future?

---

- XQuery is like SQL was in 1986
- XML is probably not going away
- Business apps need storage, logic, and presentation
  - XML is growing quickly in storage and presentation
  - XQuery makes logic consistent with the rest of the app
  - XQuery operates on XML declaratively, in its native model
- Alternative approaches to processing XML:
  - Host languages with DOM, SAX, and STAX
  - SQL/XML
  - XSLT
  - SparQL
  - Direct language integration: XLinQ etc.

# What does XQuery need?

---

- Updates
- Search
- Error handling (try/catch)
- Dynamic namespace bindings
- Group By
- Usable host-language API's
- Extensions for stand-alone programming
- Better development tools
- Good educational materials

# Thanks to the XQuery Pioneers!

---

- Ron Avnur, Mark Logic
- Kevin Beyer, IBM
- Per Bothner, Consultant
- Ronald Bourret, Consultant
- Mike Carey, BEA
- David Carlisle, Numerical Algorithms Group (UK)
- Frank Cohen, Raining Data
- Jeff Dexter, Raining Data
- Frans English, Telia.com
- Mary Fernandez, AT&T
- Dana Florescu, Oracle
- Ryan Grimm, O'Reilly
- Mary Holstege, Mark Logic
- Jason Hunter, Mark Logic
- Michael Kay, Saxonica
- Darin McBeath, Elsevier Publishers
- Srinivas Pandrangi, Ipedo
- Martin Probst, X-Hive
- Jonathan Robie, Data Direct
- Michael Rys, Microsoft
- Jerome Simeon, IBM
- Andy Townsend, John Wiley Publishers