Report on the Third International Workshop on XQuery Implementation, Experience and Perspectives (XIME-P 2006)

Michael Carey

BEA Systems Inc., San Jose, CA, USA
mcarey@bea.com

Torsten Grust

Technische Universität München, Germany
grust@in.tum.de

Summary

On June 30, 2006, *XIME-P 2006*, the International Workshop on *XQuery Implementation, Experience and Perspectives* was held. This workshop marks the third event in a workshop series whose primary aim is to shed light on XQuery systems, specification aspects, foundations of the language, and the many perceivable shapes it may take on in the future. Like the two previous workshops of 2004 (Paris) and 2005 (Baltimore), XIME-P 2006 was held as a co-located workshop of the ACM SIGMOD Conference, which this year was held in Chicago, IL, USA. The 2006 edition was co-chaired by Michael Carey and Torsten Grust. The workshop web site can be found at www.ximep-2006.org.

A Landmark Year for XQuery

After eight years of hard work, in 2006 the XQuery family of standards has entered W3C Candidate Recommendation status and odds are that XQuery will become a Recommendation this very year. XIME-P once more provided *the* forum for XQuery designers, practitioners, and system builders to reflect on the past design process, the current state of the specification, and the XQuery processors implementing the language. Further, the workshop had a distinctive "Now what?" atmosphere to it: several of the talks as well as the panel were addressing language extensions and even discussed efforts to embrace programming models other than the functional approach that XQuery has adopted from the start (see comments on the workshop panel below).

In this landmark year for XQuery, XIME-P 2006 was happy and proud to welcome Don Chamberlin (IBM Almaden, co-editor of the W3C XQuery specifications and co-designer of the original SQL language)

as this year's keynote speaker. Don, who clearly seems to devote a significant share of his lifetime to language design, gave a one-hour keynote entitled "XQuery—Where Do We Go From Here?" which started out with an assessment of what went right and wrong with the design of XQuery. Don placed the serious integration with existing XML standards, declarativity and independence of a specific persistence model, and completeness in the presence of an easy-to-learn language core on the "right" side of the fence. On the other side live a partly fragile syntax, side-effecting node constructors, and a number of important omissions from version 1.0 of the standard (*e.g.*, updates, text search, and explicit support for grouping). Next, since the first meeting of the XQuery Working Group took place back in November 1999, Don tried to respond to the oft-heard critical question on why it took so (too?) long to get to where the language is today. One piece of Don's answer related to the sometimes overwhelming number of public comments on draft XQuery standards. Processing these comments thoroughly took time. According to Don, "XQuery is like SQL was in 1986", which, since XML is probably not going away, answers the question of whether XQuery has a future or not. Apart from the version 1.0 omissions mentioned above, Don closed his talk by underlining the need for language extensions that turn XQuery into a stand-alone programming language, a thesis that received both applause and debate over the course of the workshop day.

The slides of the keynote talk are available on the web (see Electronic Proceedings below).

Workshop Sessions

A total of seven presentations promoted different possible extensions to XQuery, dug deep into innova-

tive implementations details of XQuery processors, and also tried to embrace non-(or multi-)hierarchical XML data instances. Michael Kay presented several use cases for a *positional grouping* extension which can turn a flat sequence of items into a hierarchy based on their position in the sequence. XQueryP, a proposal that was co-authored by a number of W3C XQuery Working Group members, adds defined evaluation order and destructive variable assignment (in case `declare execution sequential` is given in the query preamble). XQueryP ultimately aims to evolve XQuery in the direction of an XML application development language.

The XQuery processing model lends itself to a diversity of implementation approaches. A group of three presentations described (1) a pull-based *streaming* processor (XQPull) for XQuery that can cope with recursive queries and backward axis steps, (2) a *native* XPath processor (Natix) that takes advantage of recurring structures in incoming XML messages by shifting parts of the query evaluation effort from run time to compile time, and (3) a university course (taught at U Saarland, Germany) that teaches students how to use *relational* query compilation and processing techniques to process a subset of XQuery.

Two talks convincingly showed how real applications naturally lead to overlapping, non-tree-shaped data instances, which the markup community refers to as *multi-hierarchical* or *standoff-annotated* XML. The talks proposed language extensions designed to access such overlapping data instances and also described how an efficient implementation of "stand-off " XPath axes found their way into the MonetDB/XQuery processor.

Caffeine and Code

For the first time, the 2006 edition of XIME-P explicitly called for XQuery system and application demonstrations. In retrospect, this can only be considered a success. The workshop's 90-minute "Caffeine and Code" session featured 10 exhibits, ranging from XQuery development and debugging tools to XQuery processors based on different implementation paradigms (streaming, native, relational) to innovative applications with an XQuery core (search, content assembly and publishing). Additionally, Erik Meijer (Microsoft) presented XLinq, a tight integration of an XML-aware in-memory query language—based on the monad comprehension calculus—with .NET-based programming languages. Different from the first edition of XIME-P, the XQuery community now has clearly reached a stage where there is no lack of mature implementations anymore. This provides the much needed playground to test drive the core language as well as its many perceivable extensions. It is also one of the key prerequisites for XQuery to finally obtain W3C Recommendation status.

Panel "Programming for XML"

The workshop day drew to a close with a panel reviving the aforementioned question of whether (and if so, how) XQuery shall evolve into a full-fledged programming language. Led by Dana Florescu (Oracle), the group of panelists was split in two "opposing" camps. Michael Carey (BEA), co-author of the XQueryP proposal, and Erik Meijer, proposing XLinq embedded into Visual Basic, each argued for a procedural-style approach to XML application development. Michael Kay (Saxonica), on the other hand, voiced his fear that imperative constructs in XQuery would be abused by users unfamiliar with the declarative style of querying: instead, stateful logic should only be used in the glue that couples (almost purely) functional XQuery building blocks. Complex applications *can* be built today using the current specification of XQuery, insisted Ron Avnur (Mark Logic). As if to prove his point in advance, Ron undertook some on-the-fly XQuery coding during the earlier Caffeine and Code session. The notion of *monads*, which came up in the XLinq presentation and which has already helped more than once to bring impure constructs to the world of functional programming languages, received some interest among the workshop participants.

Raw Numbers

XIME-P 2006 received 15 paper submissions plus approximately the same number of proposals for system demonstrations. Each submission was reviewed by at least six eyes from the workshop program committee. The PC had 16 members, five of which are also part of the W3C XQuery standardization effort. The workshop room (hard to find, but worth it) in the Renaissance Chicago Hotel hosted a lively group of 40 attendees, with academia and industry each contributing approximately 50% of the participants.

Electronic Proceedings

All accepted papers have been published in the XIME-P 2006 Electronic Proceedings, ISBN 1-5953-465-0, as part of the ACM Digital Library. Pa-

pers will appear in the SIGMOD Digital Symposium Collection (DiSC). The Electronic Proceedings may also be found on the workshop web site at `www.ximep-2006.org`. This site additionally hosts the slides of Don Chamberlin's keynote talk as well as entries for all demonstrations featured in the Caffeine and Code session.

Acknowledgments

The advent of XQuery has led to interesting interconnections between the database, markup, and programming language research and development communities. This year's workshop co-chairs hope that a fourth edition (and further editions) of XIME-P will continue to tighten this bond in 2007 and beyond.