

Rich Clients need Rich Interfaces

Query Languages for XML and RDF Access on the Web

Proposal for 3h Tutorial at the “XML-Tage” 2006 in Berlin, Germany

James Bailey*, François Bry†, Tim Furche†, Benedikt Linse†, Paula-Lavinia Pătrânjan†, and Sebastian Schaffert‡

*Department of Computer Science and Software Engineering, University of Melbourne,

111 Barry Street, Carlton, Victoria 3010, Australia

<http://www.cs.mu.oz.au/~jbailey/>

†Institute for Informatics, University of Munich,

Oettingenstraße 67, 80538 München, Germany

<http://pms.ifi.lmu.de/>

‡Salzburg Research Forschungsgesellschaft,

Jakob Haringer Str. 5/II, 5020 Salzburg, Austria

<http://www.salzburgresearch.at/>

Abstract— Access to Web data has become an integral part of many applications and services. In the past, such data has usually been accessed through human-tailored HTML interfaces. Nowadays, rich client interfaces in desktop applications or, increasingly, in browser-based clients ease data access and allow more complex client processing based on XML or RDF data retrieved through Web service interfaces. Convenient specifications of the data processing on the client and flexible, expressive service interfaces for data access become essential in this context. Web query languages such as XQuery, XSLT, SPARQL, or Xcerpt have been tailored specifically for such a setting: declarative and efficient access and processing of Web data. This tutorial introduces, compares, and classifies the most relevant exemplars of Web query languages for XML, RDF, and/or TopicMaps data. Interesting features as well as differences in expressiveness and adequacy are digested along practical and concrete use cases. Emphasis is placed on recent W3C standardization activities, contrasted with alternative approaches from industry and academia.

I. TUTORIAL DETAILS

Length of Tutorial:

3h (three hours). Giving the attendants an impression of current Web query languages and a set of guidelines for their use in developing and deploying Web applications requires a certain breadth for the tutorial. Despite a careful selection of highly relevant exemplars from the large number of languages considered in the underlying survey, a three hour time frame is necessary to cover the material in sufficient depth.

Intended Audience:

The topic is highly relevant for both practitioners and managers involved in the creation of modern Web applications that use Web service interfaces for data access. In this context, the tutorial serves as a decision help on how to provide flexible access to large data bases through query interfaces. In particular, the impact of query language concepts and features on the intended service are discussed. The tutorial is also relevant for experts in XML or Semantic Web technology as it provides a novel perspective over areas of research usually considered separate and introduces into many languages that are still largely unknown. However, though both the underlying survey and the tutorial itself provide ample pointers to more in-depth coverage of the languages considered, it is not intended as a comprehensive

introduction into any of the languages discussed to the depth needed to be able to use them immediately.

Required Knowledge:

A basic understanding of Web technologies such as XML, HTTP, and Web Services is needed. Some knowledge of RDF and/or similar Semantic Web technologies is advantageous though not strictly necessary. Prior knowledge on query languages or Web query interfaces is *not* required though it certainly helps in some of the tutorial parts.

(Short) Tutorial Description:

Convenient specifications of Web data processing on the client and flexible, expressive service interfaces for data access on the server become essential in the context of Web services commoditizing structured information on the Web. Web query languages such as XQuery, XSLT, SPARQL, or Xcerpt have been tailored specifically for such a setting: declarative and efficient access and processing of Web data. This tutorial introduces, compares, and classifies the most relevant exemplars of Web query languages for XML, RDF, and/or TopicMaps data. The selection of these languages is based upon a recent survey of Web query languages [1] conducted by the authors. Interesting features as well as differences in expressiveness and adequacy are digested along practical and concrete use cases. Emphasis is placed on recent W3C standardization activities, contrasted with alternative approaches from industry and academia. The tutorial concludes with a discussion of advantages and challenges for deploying Web query languages today.

Overview/Outline of the Tutorial:

See following section.

Prior Presentations of this Tutorial:

An extended version of the tutorial (with intensive exercises) has been conducted (by James Bailey, François Bry, and Tim Furche) at the “Reasoning Web” 2005 Summer School, Mdina, Malta. A slightly shorter version has been presented (by Sebastian Schaffert) at the Franco-Mexican (Summer) School on Distributed Systems, Grenoble, France. The tutorial is based on an extensive survey of Web and Semantic Web query languages published as a chapter [1] in the Springer tutorial volume “Reasoning Web”.

Technical Requirements:

The only technical requirements of the tutorial are common presentation equipment for the lecturers including an HTML browser for the slides. There are no technical requirements for the participants of the tutorial.

Note on Presenters:

This tutorial is based on work from all of the authors. At least two of the tutorial authors will present it, if accepted, at the “XML-Tage” 2006. The survey chapter [1] forms the basis of this tutorial but the material has been extended and updated continuously. This tutorial particularly considers also two recent developments: First, the ongoing activities on rule languages and rule interchange at the W3C and their relevance for Web query languages; second, the challenges for Web query languages arising from novel Web applications such as Google Maps, Flickr, or Zimbra, where the communication between Web browser and Web server moves more and more from “plain” HTML to (syntactically and semantically) “rich” XML and RDF data that can be processed by the client in more interesting ways than plain HTML.

Tutorial Language:

Based on the preferences of the conference organizers, the tutorial can be held either in English or in German language.

II. OUTLINE OF THE TUTORIAL

In this tutorial, a large number of Semantic Web query languages are considered with focus on those (a) already seeing wide-spread adoption, (b) under consideration for standardization, or (c) providing novel and influential perspectives on Semantic Web querying.

The languages considered can roughly be divided in three groups based on the format of the data queried (XML, RDF, or TopicMaps). Additionally, the tutorial also gives a brief outlook into two very recent research directions: rules and rule interchange in Web query languages and versatile Semantic Web query languages that allow intertwined access to data in different representation formalisms, e.g., to RDF and Topic Maps data. The actual discussion of the query languages will therefore be oriented on the following structure:

- 1) **Introduction, Sample Data, and Query Scenario** (15 minutes)
- 2) **XML Query Languages** (60 minutes)
 - (a) “*W3C’s Query Languages: The Navigational Approach*” introduces the mainstream XML query languages XPath, XQuery, XSLT. Emphasis is placed on foundational principles and recent developments.
 - (b) “*The Positional Approach to XML Querying: An Better Way?*” surveys a group of alternative query languages from academia that suggest a different approach for XML querying similar to the “Query-by-Example” [23] paradigm for relational data.
- 3) **RDF Query Languages** (60 minutes):
 - (a) “*The SPARQL family*” presents the family of languages originating from the proposal of SquishQL [13]. The focus lies on RDQL [20], [20], a widely adopted RDF query language implemented, e.g., in HP’s Jena Toolkit [10], and on SPARQL [17], the RDF query language currently under standardization at the W3C.
 - (b) “*The RQL family*” discusses the RQL [11] language, a very different perspective on RDF querying than SPARQL focused on a strong type system and rich, but complex language

constructs. Several proposals for simplifying RQL, e.g., SeRQL and eRQL, highlight strength and weaknesses of this approach. (c) “*Navigational Access to RDF: Versa*” presents the only exemplar of an RDF query language that uses the navigational access dominating the mainstream XML query languages. (d) “*Reactive Rules in RDF Query Languages*” discusses reactive rules and their use in some prototypical RDF query languages, e.g., Algae [16], the query language of the W3C Annotea project, iTQL [22] and Nokia’s WQL [12], two industry proposals for RDF query languages.

(f) “*Deductive Query Languages.*” A final pitch on RDF query languages is provided by query languages in logic-programming style. TRIPLE [21] is an early proposal based on frame logic. Xcerpt [18], [19] is a query language developed by some of the authors of this tutorial that integrates RDF querying with access to other data formats, especially XML, and provides Prolog-style reasoning.

- 4) **Topic Maps Query Languages** (15 minutes):

In contrast to RDF, the number of query languages for Topic Maps is still rather small. Therefore this part of the tutorial only briefly skims over the most important developments including the current standardization activities at ISO that are expected to result in a standard Topic Maps Query Language.
- 5) **Outlook on Rules and Web Querying** (10 minutes):

Many of the above mentioned languages have rules or similar mechanisms, e.g., for integrating Web data, basic reasoning, or reactivity. The W3C is currently starting an activity to standardize rule languages and rule interchange on the Web. Goals and charter of this activity, as well as its relevance for the larger topic of this tutorial, are briefly summarized.
- 6) **Outlook on Versatile Web Query Languages** (10 minutes):

The vision of versatile Web query languages will be introduced and exemplified using Xcerpt [18], [19].
- 7) **Summary and Conclusion** (10 minutes):

The discussion of the query languages closes with a summary and comparison of central language features.

III. ABOUT THE AUTHORS

James Bailey holds a PhD in Computer Science from the University of Melbourne, Australia (1998), where he is currently employed as a Senior Lecturer. His research interests are in database technology, in particular active databases and database theory, and more recently in XML transformations and queries, data mining, and robust agent systems [2], [8]. He has held numerous University lectures and seminars both on general topics in Computer Science and on Web technologies closely related to the content of this tutorial.

François Bry (PhD in 1981) is currently investigating methods and applications emphasizing XML, RDF, query answering, reactivity and reasoning on the Web [3], [6], [7], [14], [15], [19]. Formerly, he worked on knowledge representation, deductive databases, automated theorem proving, and logic programming. Since 1994, he is a full professor at the Institute for Informatics of the University of Munich, Germany. Formerly, he worked with the industry research center ECRC in Munich. He is regularly giving University lectures on XML, knowledge representation, logic programming, and the Web and has conducted several tutorials at international conferences and workshops. He is scientific coordinator of the European Network of Excellence REVERSE (<http://reverse.net>) on Reasoning in the Semantic Web launched in 2004. He is a member of the Rule

Interchange Format (RIF) Working Group at the W3C (<http://www.w3.org/2005/rules/>).

Tim Furche holds a master's degree in Computer Science from the University of Munich, Germany (2002). He is currently employed as a research assistant at the Institute for Informatics of the University of Munich, assisting in the coordination of the REVERSE working group on "Reasoning-aware Querying". His present research interests are in XML and semistructured data, in particular query evaluation and optimization, and advanced Web systems. He has (co-) authored publications on characterizing, optimizing, and evaluating the W3C query language XPath and, more recently, on the use of XML query languages, in particular Xcerpt, for querying Semantic Web data [5], [14], [15]. He has given several graduate seminars on querying, visualizing, and representing Web and Semantic Web data as well as tutorials on Xcerpt, a Web and Semantic Web query language developed in Munich. He has also developed and conducted a number of industrial training courses on office and data acquisition applications for employees in health care services.

Benedikt Linse holds a master's degree in Computer Science from the University of Munich, Germany (2006), where he is currently employed as a research and teaching assistant. His present research interests include versatile query languages for both the Web and the Semantic Web, automatic translation between XQuery and Xcerpt, reasoning and efficient rule chaining in Xcerpt, and efficient query evaluation for both XML and RDF data. He has held several presentations on various topics including XML query languages at advanced and post-graduate seminars.

Paula-Lavinia Pătrânjan holds a PhD in Computer Science from the University of Munich, Germany (2005). She works as standardization coordinator for the Network of Excellence REVERSE on reasoning on the Web with rules and semantics. She is currently actively involved as a member within the W3C Rule Interchange Format (RIF) Working Group (<http://www.w3.org/2005/rules/>), part of W3C's Semantic Web Activity. Her research interests are in Web and Semantic Web systems with reactive capabilities. She is one of the organizers of the workshop "Reactivity on the Web" to be held as an official satellite event at the Int'l. Conf. on Extending Database Technology (EDBT 2006). She has given several presentations on Xcerpt and XChange [4], a reactive language for the Web developed at the University of Munich, and has contributed to a summer school's tutorial on XChange.

Sebastian Schaffert holds a PhD in Computer Science from the University of Munich, Germany (2004). He is currently employed as a senior researcher at Salzburg Research, Austria. His research interests are in Web and Semantic Web systems, in particular social and educational software, knowledge representation, XML and semistructured data, and reasoning languages for the Web, as well as functional and logic programming. He has contributed to many scientific conferences as author and program committee member and has authored several publications on the rule-based Web query language Xcerpt and related topics [3], [5], [6], [18], [19]. He has given several graduate seminars and exercises on XML and Web technologies, as well as several tutorials on related issues at international summer schools and workshops.

Acknowledgements

We would like to thank Renzo Orsini and Ian Horrocks for contributing to the survey of Web and Semantic Web query languages [9] that has been used as reference material for this tutorial.

This research has been funded by the European Commission and by the Swiss Federal Office for Education and Science within the 6th Framework Programme project REVERSE number 506779 (cf. <http://reverse.net>).

REFERENCES

- [1] J. Bailey, F. Bry, T. Furche, and S. Schaffert. Web and Semantic Web Query Languages: A Survey. In J. Maluszinsky and N. Eisinger, editors, *Reasoning Web Summer School 2005*, pages 35–133. Springer-Verlag, LNCS 3564, 2005.
- [2] J. Bailey, G. Papamarkos, A. Poulouvasilis, and P. Wood. An Event-Condition-Action Language for XML. In *Web Dynamics: Adapting to Change in Content, Size, Topology and Use*, pages 223–248. Springer-Verlag, 2004.
- [3] S. Berger, F. Bry, S. Schaffert, and C. Wieser. Xcerpt and visXcerpt: From Pattern-Based to Visual Querying of XML and Semistructured Data. In *Proc. 29th Intl. Conference on Very Large Data Bases (VLDB03)*, Berlin, Germany, September 2003.
- [4] F. Bry, M. Eckert, and P.-L. Pătrânjan. Reactivity on the Web: Paradigms and Applications of the Language XChange. *Journal of Web Engineering*, 5(1):3–24, 2006.
- [5] F. Bry, T. Furche, L. Badea, C. Koch, S. Schaffert, and S. Berger. Querying the Web Reconsidered: Design Principles for Versatile Web Query Languages. *Journal of Semantic Web and Information Systems*, 1(2), 2005. 14.
- [6] F. Bry, T. Furche, P.-L. Patranjan, and S. Schaffert. Data Retrieval and Evolution on the (Semantic) Web: A Deductive Approach. In *Proc. Workshop on Principles and Practice of Semantic Web Reasoning*, LNCS 3208. Springer-Verlag, 2004.
- [7] F. Bry and P. Kröger. A Computational Biology Database Digest: Data, Data Analysis, and Data Management. *Distributed and Parallel Databases*, 13(1), January 2003.
- [8] C. Dong and J. Bailey. Optimization of XML Transformations Using Template Specialization. In *Proc. 5th International Conference on Web Information Systems Engineering (WISE 2004)*, 2004.
- [9] T. Furche, F. Bry, S. Schaffert, R. Orsini, I. Horrocks, and O. Bolzer. Survey over Existing Query and Transformation Languages. Deliverable I4-D1, REVERSE, 2004.
- [10] H. L. S. W. R. Group. Jena – A Semantic Web Framework for Java. Online only, 2004.
- [11] G. Karvounarakis, S. Alexaki, V. Christophides, D. Plexousakis, and M. Scholl. RQL: A Declarative Query Language for RDF. In *Proc. International World Wide Web Conference*, May 2002.
- [12] O. Lassila. Wilbur Query Language Comparison. Online only, 2004.
- [13] L. Miller, A. Seaborne, and A. Reggiori. Three Implementations of SquishQL, a Simple RDF Query Language. In *Proc. International Semantic Web Conference*, June 2002.
- [14] D. Olteanu, T. Furche, and F. Bry. An Efficient Single-Pass Query Evaluator for XML Data Streams. In *Proc. ACM Symposium on Applied Computing*, 2004.
- [15] D. Olteanu, H. Meuss, T. Furche, and F. Bry. XPath: Looking Forward. In *Proc. EDBT Workshop on XML Data Management*, LNCS 2490. Springer Verlag, 2002.
- [16] E. Prud'hommeaux. Algae RDF Query Language. Online only, 2004.
- [17] E. Prud'hommeaux and A. Seaborne. SPARQL Query Language for RDF. Working draft, W3C, 2005.
- [18] S. Schaffert. *Xcerpt: A Rule-Based Query and Transformation Language for the Web*. Dissertation/Ph.D. thesis, University of Munich, 2004.
- [19] S. Schaffert and F. Bry. Querying the Web Reconsidered: A Practical Introduction to Xcerpt. In *Proc. Extreme Markup Languages*, August 2004.
- [20] A. Seaborne. RDQL – A Query Language for RDF. Online only, January 2004.
- [21] M. Sintek and S. Decker. TRIPLE—A Query, Inference, and Transformation Language for the Semantic Web. In *Proc. International Semantic Web Conference*, June 2002.
- [22] Tucana Technologies. iTQL Commands. Online only, 2004.
- [23] M. Zoof. Query By Example: A Data Base Language. *IBM Systems Journal*, 16(4):324–343, 1977.