

Web Queries: From a Web of Data to a Semantic Web^{*}

François Bry, Tim Furche, Klara Weiland

Institute for Informatics, University of Munich
Oettingenstr. 67, 80538 München, Germany
<http://pms.ifi.lmu.de>

1 Abstract

One significant effort towards combining the virtues of Web search, viz. being accessible to untrained users and able to cope with vastly heterogeneous data, with those of database-style Web queries is *keyword-based Web query languages*. These languages operate essentially in the same setting as XQuery or SPARQL but with an interface for untrained users.

Keyword-based query languages trade some of the precision that languages like XQuery enable by allowing to formulate exactly what data to select and how to process it, for an easier interface accessible to untrained users. The yardstick for these languages becomes an easily accessible interface that does not sacrifice the essential premise of database-style Web queries, that selection and construction are precise enough to *fully automate* data processing tasks.

To ground the discussion of keyword-based query languages, we give a summary of what we perceive as the main contributions of research and development on Web query languages in the past decade. This summary focuses specifically on what sets Web query languages apart from their predecessors for databases.

Further, this tutorial (1) gives an overview over keyword-based query languages for XML and RDF (2) discusses where the existing approaches succeed and what, in our opinion, are the most glaring open issues, and (3) where, beyond keyword-based query languages, we see the need, the challenges, and the opportunities for combining the ease of use of Web search with the virtues of Web queries.

2 Biographies

Prof. Dr. François Bry is a full professor at the Institute for Informatics of the Ludwig-Maximilian University of Munich, Germany, heading the research group for programming and modeling languages. He is currently investigating methods and applications related to querying answering and reasoning on the Web and social semantic Software and Media. In particular his research focuses on query and rule languages for Web data formats such as XML and RDF, complex events

^{*} This tutorial is based on [21]

and social media. François Bry has a research record of over 140 peer-reviewed scientific publications. He has supervised over 15 doctoral projects. Among his former doctoral students are Slim Abdennadher, dean of Computer Science at the German University in Cairo, Egypt, Sebastian Schaffert, project leader with Salzburg Research, Dan Olteanu, lecturer at Oxford University, and Tim Furche, postdoc at the University of Munich. François Bry regularly contributes to scientific conferences and journals, especially in the areas Web and Semantic Web as an author, reviewer or program committee member. Before joining University of Munich in 1994, he worked in industry in France and Germany, in particular with the research center ECRC.

Dr. Tim Furche is a postdoctoral researcher at the Institute for Informatics of the Ludwig-Maximilian University of Munich, Germany, in the research group for programming and modeling languages. His research interests are XML and semi-structured data, in particular query evaluation and optimization, and advanced Web systems. His main contributions are on XPath optimization (XPath: Looking Forward) and evaluation and on linear time and space querying of large graphs. He has authored over 40 peer-reviewed scientific publications, some of them cited over 150 times. Tim Furche regularly contributes to scientific conferences and journals, especially in the areas Web and Semantic Web as an author, reviewer or program committee member. Before joining University of Munich in 1994, he worked in industry in France and Germany, in particular with the research center ECRC. From 2004-2008 he was co-coordinating the REVERSE working group on Reasoning-aware Querying.

Klara Weiland is a doctoral student in the KiWi - Knowledge in a Wiki project, working on the development of KWQL, a versatile and powerful but at the same time user-friendly query language. She graduated with a Bachelor's degree in Cognitive Science from the University of Osnabrück in 2005. She went on to pursue a Master's degree in Artificial Intelligence at the University of Amsterdam. During this time, she focussed on speech and language processing and machine learning while working as a research assistant in computational phonetics and authoring publications in formal semantics, phonetics and forensic artificial intelligence.

References

1. J. Bailey, F. Bry, T. Furche, and S. Schaffert. Web and Semantic Web query languages: A survey. In *Reasoning Web: Int. Summer School*, volume 5224 of *LNCS*, 2005.
2. A. Berglund, S. Boag, D. Chamberlin, M. Fernandez, M. Kay, J. Robie, and J. Simeon. XML Path Language (XPath) 2.0. Recommendation, W3C, 2005.
3. F. Bry, T. Furche, B. Linse, and A. Pohl. XcerptRDF: A pattern-based answer to the versatile web challenge. In *Workshop on (Constraint) Logic Programming*, 2008.

4. F. Bry, J. Kotowski, and K. Weiland. Querying and reasoning for social semantic software. In *European Semantic Web Conf. (Poster session)*, 2009.
5. D. Chamberlin. XQuery: A query language for XML. In *SIGMOD: Int. Conf. on Management of Data*, 2003.
6. D. Chamberlin, J. Robie, and D. Florescu. Quilt: An XML Query Language for Heterogeneous Data Sources. In *Workshop on Web and Databases*, 2000.
7. S. Cohen, J. Mamou, Y. Kanza, and Y. Sagiv. XSearch: A semantic search engine for XML. In *Int. Conf. on Very Large Databases*, 2003.
8. J. Davies and R. Weeks. QuizRDF: Search technology for the Semantic Web. In *Int. Conf. on System Sciences*, 2004.
9. D. Florescu, D. Kossmann, and I. Manolescu. Integrating keyword search into XML query processing. In *Int. WWW Conf. on Computer networks*, 2000.
10. T. Furche, B. Linse, F. Bry, D. Plexousakis, and G. Gottlob. RDF querying: Language constructs and evaluation methods compared. In *Reasoning Web, Int. Summer School*, volume 4126 of *LNCS*, 2006.
11. L. Guo, F. Shao, C. Botev, and J. Shanmugasundaram. XRANK: Ranked keyword search over XML documents. In *SIGMOD: Int. Conf. on Management of Data*, 2003.
12. V. Hristidis, Y. Papakonstantinou, and A. Balmin. Keyword proximity search on XML graphs. In *Int. Conf. on Data Engineering*, 2003.
13. G. Karvounarakis, S. Alexaki, V. Christophides, D. Plexousakis, and M. Scholl. RQL: A Declarative Query Language for RDF. In *Int. WWW Conf.*, 2002.
14. Y. Li, C. Yu, and H. V. Jagadish. Schema-free XQuery. In *Int. Conf. on Very Large Data Bases*, 2004.
15. Z. Liu, J. Walker, and Y. Chen. XSeek: A semantic XML search engine using keywords. *Int. Conf. on Very Large Data Bases*, 2007.
16. D. Olteanu, H. Meuss, T. Furche, and F. Bry. XPath: Looking Forward. In *EDBT Workshop on XML-Based Data Management*, volume 2490 of *Lecture Notes in Computer Science*. Springer Verlag, 2002.
17. E. Prud'hommeaux and A. Seaborne. SPARQL Query Language for RDF. Recommendation, W3C, 2007.
18. S. Schaffert and F. Bry. Querying the web reconsidered: A practical introduction to Xcerpt. In *Extreme Markup Languages Conf.*, 2004.
19. A. Schmidt, M. L. Kersten, and M. Windhouwer. Querying XML documents made easy: Nearest concept queries. In *Int. Conf. on Data Engineering*, 2001.
20. H. Wang, K. Zhang, Q. Liu, D. T. Tran, and Y. Yu. Q2semantic: A lightweight keyword interface to semantic search. In *Int. Semantic Web Conf.*, 2008.
21. K. Weiland, T. Furche, and F. Bry. Quo Vadis, Web Queries? In *Int. Workshop on Semantic Web Technology (Web4Web) 2008*, 2008.
22. Y. Xu and Y. Papakonstantinou. Efficient keyword search for smallest LCAs in XML databases. In *SIGMOD: Int. Conf. on Management of Data*, 2005.