Michael Schmidt

Curriculum Vitae
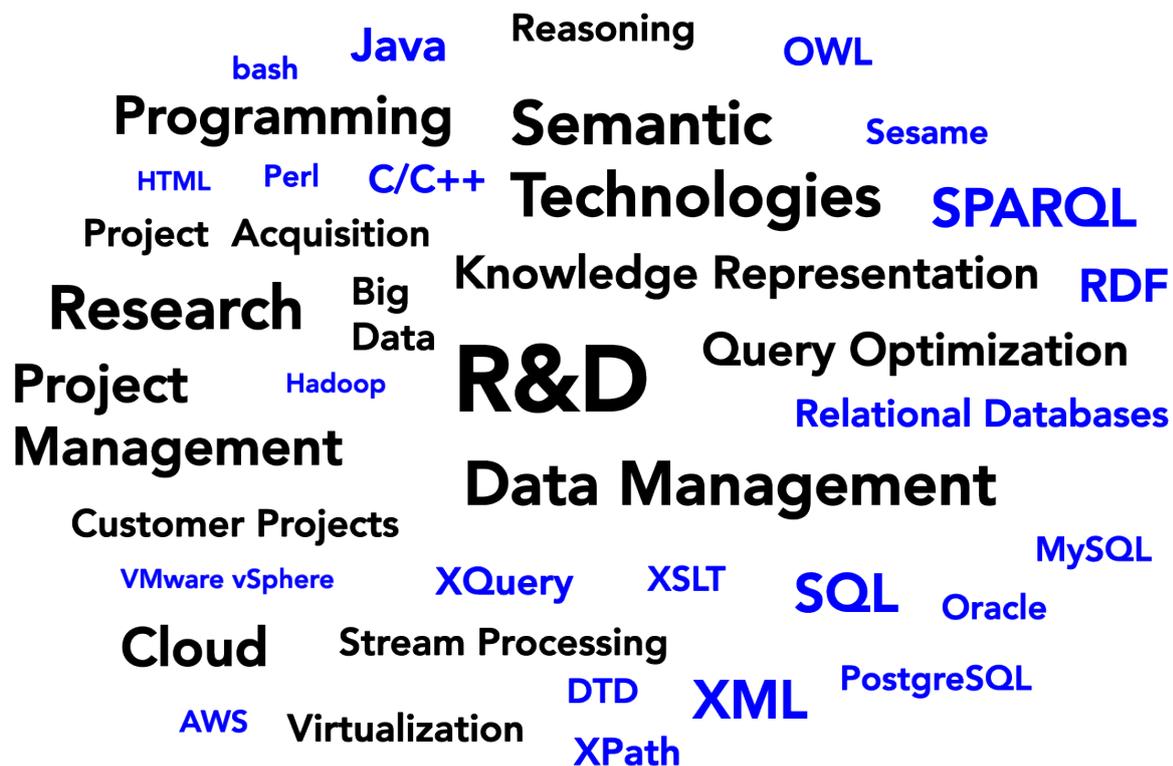
66133 Saarbrücken
✉ m.schmidt00@gmail.com

Career Objective

With enthusiasm for the conceptual simplicity that lies behind complex systems and algorithms, I enjoy working on innovative topics from both a theoretical and practical perspective. Using and systematically expanding my expertise in the areas of data modeling and processing, data-centric application development, and cloud management, I am passionate about pushing technology to its limits and transforming innovative concepts into industrial-strength systems.

Beauty is our business. – Edsger W. Dijkstra

Me at a Glance

bash **Java** Reasoning OWL

Programming **Semantic** Sesame

HTML Perl C/C++ **Technologies** **SPARQL**

Project Acquisition **Knowledge Representation** RDF

Research Big Data **R&D** **Query Optimization**

Project Hadoop Relational Databases

Management **Data Management**

Customer Projects MySQL

VMware vSphere XQuery XSLT **SQL** Oracle

Cloud Stream Processing PostgreSQL

AWS Virtualization DTD **XML**

XPath

Professional Experience

since 08/2013   **information multimedia communication AG**, *Research Manager*

Acquisition and management of research projects, As a Research Manager, I am responsible for the acquisition and coordination of national and international research projects within IMC's innovation department. The main focus of my activities lies on (i) the integration of social interaction and social network features into the IMC Learning Portal, a user friendly E-Learning portal for delivering Massive Open Online Courses (MOOCs) to learners, developed within the projects *Lernnavigator* (EFI) and *KOMMIT* (BMBF), (ii) the didactical design and technical implementation of an online learning course for carers of people burdened with Young Onset Dementia in the project *RHAPSODY* (EU/JPND), and (iii) the design and implementation of context-sensitive assistance systems supporting "learning on the job" in the construction industry (BMWi project *APPsist*, BMBF project *GLASSROOM*) and within the Industry 4.0 context (BMBF project *SmartWerk*).

Architecture and development lead, Acting at the interface between research and industrial-strength software development, I am responsible for the technical design of Web-based prototype implementations carried out in the research projects. As such, I am coordinating the project specific development activities, constantly pushing innovation transfer into IMC's established product portfolio.

Team lead, In addition to the technical responsibilities, I am disciplinary responsible for a team of five persons working in IMC's research department, including both full time employees and working students.

05/2010–07/2013   **fluid Operations AG**, *Senior Architect R&D*

***Optique* Big Data platform**, *Optique* is a large-scale European FP7 IP with ten partners from academia and industry, pursuing the goal to build a system for efficient end-user access to Big Data based on ontology-based data access and cloud technologies. As the technical leader of the project's central integration work package, I am responsible for the technical architecture, setup of a shared development environment, as well as the coordination of all development activities, in particular the extension and integration of partner components into the fluid Operations platform, which serves as a basis for the *Optique* system.

R&D *Information Workbench*, The *Information Workbench* is a generic, Java-based platform for the rapid development of data-centric applications based on Semantic Technologies. As one of the first members of the development team, I have been proposing, designing, and implementing various system-critical modules, focusing on topics related to efficient semantic data management, data synchronization, data integration, and data visualization. Over the years, I have been actively involved in roadmap and strategy discussions and development planning.

Product Manager *eCloudManager Intelligence Edition*, Settled at the interface between private clouds and semantic data management, the *eCloudManager Intelligence Edition* supports data center operations by providing a unified console that allows data center administrators and C-level users to explore relationships between data center resources, to document and annotate them in a central place, analyze performance indicators, and identify risks, problems, and optimization opportunities. As the technical Product Manager, I was responsible for the system architecture, definition and implementation of development processes and QA, the design, prioritization, and integration of new feature into the product, as well as interfacing to Marketing and Sales teams.

Lead customer projects, I have been successfully leading several customer projects, managing teams of up to five persons. In this role, I was responsible for project management (SRS, SOW, resource and time planning, controlling, approval), internal and external reporting, as well as design and implementation of customer requirements in cooperation with internal development teams. The projects had a strong focus on the integration of data from heterogeneous data sources using the *Information Workbench* as a semantic integration platform, in order to enable a unified view on and ad hoc analysis of the integrated data corpus.

Training & Knowledge Transfer, I have been involved in planning, creating, organizing, and executing product and technology trainings for internal departments, external partners, and customers.

09/2003–07/2006   **hSp GmbH**, *Ottweiler*, Side job during studies

Project *doc.mobile*, *doc.mobile* is a software application that gives general practitioners access to patient data during home visits via mobile devices. It comes with a Windows server, synchronizing data with the mobile device, and a PocketPC GUI application. My responsibilities were the internal project coordination within a team of three persons, external project coordination with distribution partners, and the implementation of the Windows server (using the Qt 4.1 library). I have written 12k lines of C++ source code, including interfaces to the patient information system, data synchronization (via ActiveSync), and user management components. Further, I contributed to the design and implementation of the mobile component written using eMbedded Visual C++.

Project *MARIS*, HL7 is a standard protocol for the exchange of medical information such as patient data. I developed a server application for reading, processing and writing HL7 segments, which communicates with an underlying PostgreSQL database. The application comprises about 20k lines of Java code and is embedded into a system called *MARIS*, giving medical doctors mobile access to patient data via PDAs.

Skills & Qualifications

IT & Development

Programming	Java, C/C++, Perl, bash, HTML, PHP, JavaScript, CSS, UML, LaTeX
Technologies	REST, SOAP, Java Server Faces (JSF), Foundation, EJB
Development	Eclipse, SVN, Ant, Make, Jenkins, JUnit, Selenium
Operating Systems	Use, configuration, and basic administration of Linux (Debian, Ubuntu, Suse), MacOS X, and Windows
IT Administration	Setup and maintenance of IT infrastructure (Apache, SVN, Bugzilla, etc.)
Cloud	AWS, VMware vSphere, private and hybrid cloud management
E-Learning	MOOCs, learning management systems, performance support systems, experience in E-Learning content development

Data Management

Relational Databases	**Technology:** SQL, ERM, Relational Algebra, Algebraic & Semantic Query Optimization, Database Design, Column Stores
	Systems: MySQL, Oracle, PostgreSQL, Oracle, Berkeley DB
XML	**Technology:** XQuery, XPath, XSLT, SAX & DOM, streaming XML
	Systems: Galax, GCX, MonetDB, Saxon, QizX
Semantic Stack	**Technologies:** RDF & RDFS, Linked Data, SPARQL, OWL, R2RML
	Systems: Sesame, OWLIM, Jena, Virtuoso, Systap Bigdata
	Benchmarking: Development of SP^{2}Bench and FedBench, two prominent benchmarks for semantic technologies
NoSQL	MapReduce, Hadoop, Optique

General

Project Management	Management of research and customer projects
Others	Scientific writing
	Acquisition of research projects
	Experience in mobile application development

Education & Academic Projects

10/2009–04/2010   **Database and Information Systems Group**, *Albert-Ludwigs-Universität Freiburg*, Research assistant and PhD student, supervised by Prof. Dr. Georg Lausen, PhD degree (Dr. rer. nat.) obtained in Dec. 2009, title of PhD thesis: "Foundations of SPARQL Query Optimization" Grade: summa cum laude

PhD Thesis, In my PhD thesis I have been working on techniques for the efficient evaluation of SPARQL queries over semantic data. Motivated by the success of Relational Database research, I focused on the applicability of established concepts for algebraic and semantic query optimization in the context of SPARQL and the RDF data format, the study of the complexity of SPARQL query evaluation, and benchmarking techniques covering the new challenges that arise in the context of SPARQL query processing. In this line, I have developed the SP^2Bench SPARQL performance benchmark, a scalable, holistic benchmark suite that has been extensively used in research as well as the industrial space for SPARQL performance evaluation. Further, my results on algebraic rewriting form the basis for the optimizer of the *Meronymy SPARQL Database Server*. Over the years, I have published results in leading international database and Semantic Web conferences (such as ICDE, ICDT, ISWC, PODS, and VLDB), books, and journals.

07/2007–09/2009   **Database and Information Systems Group**, *Albert-Ludwigs-Universität Freiburg*, Scholarship from the DFG-Graduiertenkolleg "Mathematical Logic and Applications"

09/2006–12/2006   **Cornell Database Group**, *Cornell University, Ithaca (NY)*, Visiting Scientist in the Cornell Database Group, lead by Prof. Gehrke

08/2006–06/2007   **Information Systems Group**, *Universität des Saarlandes*, Research assistant and PhD student, supervised by Prof. Dr. Christoph Koch

XML Prefiltering Engine, The idea behind XML prefiltering, a common technique in streaming XQuery evaluation, is to project away irrelevant parts of the XML stream prior to query evaluation, in order to optimize main memory consumption and speed up evaluation by minimizing the amount of data that is loaded to main memory for query evaluation. As a significant improvement over the state-of-the-art, our group developed a prefiltering scheme that surpasses the performance of existing systems by orders of magnitude. The prototype, written in C++ (about 11k lines of code), uses DTD knowledge in combination with established string matching techniques to efficiently locate those parts of the XML document that are relevant for query answering.

Garbage Collected XQuery Engine (GCX), In my diploma thesis, I developed and implemented concepts for a novel buffer management strategy for streaming XQuery evaluation, which combine both a static and dynamic analysis to minimize the amount of data that needs to be buffered at runtime. The GCX engine is written in C++ and comprises about 32k lines of code. The system has been released as an open source project at SourceForge (see `http://gcx-engine.sourceforge.net/`) and was promoted in the *entwickler* magazine in 2007.

10/2001–07/2006	**Studies of *Applied Computer Science***, *Universität des Saarlandes*, Focus on Software Design, Information Systems and Databases, title of Diploma thesis: "Combined Static and Dynamic Buffer Optimization for Streaming XQuery Evaluation" Grade: 1.2

XML Database Engine, In the DBMS course, I have implemented a disk-based database management system for XML databases, supporting an expressive fragment of the XQuery language. The main challenge of this project was XQuery evaluation under strict memory bounds, with a strong focus on efficient data access and algebraic query optimization. The XML database system, which comprises about 17k lines of C/C++ code, was built on top of the Berkeley DB database library.

Advanced Software Lab, I have designed and implemented a mobile zoo audio guide. The system consists of a PDA (PocketPC), a stereo headset and a connected GPS system, including an electronic compass. The user is navigated to its destination via stereo audio output, where multiple wave files are mixed together at runtime with different volumes on both channels, simulating the distance and direction of the user's objectives. The system was implemented using eMbedded Visual C++.

08/2001–09/2001	**hSp GmbH**, *Ottweiler*, Internship

Selected Publications (8 out of 30+)

2007	**M. Schmidt, S. Scherzinger, C. Koch**, *Combined Static and Dynamic Analysis for Effective Buffer Minimization in Streaming XQuery Evaluation*, 23rd Intl. Conf. on Data Engineering (ICDE), Istanbul.
2008	**C. Koch, S. Scherzinger, M. Schmidt**, *XML Prefiltering as a String Matching Problem*, 24th Intl. Conf. on Data Eng. (ICDE), Cancún.
2009	**M. Schmidt, T. Hornung, G. Lausen, C. Pinkel**, *SP²Bench: A SPARQL Performance Benchmark*, 25th Intl. Conf. on Data Engineering (ICDE), Shanghai.
	M. Meier, M. Schmidt, G. Lausen, *On Chase Termination Beyond Stratification*, 35th Intl. Conf. on Very Large Databases (VLDB), Lyon.
2010	**M. Meier, M. Schmidt, F. Wei, G. Lausen**, *Semantic Query Optimization in the Presence of Types*, 29th Symposium on Principles of Database Systems (PODS), Indiana.
	M. Schmidt, M. Meier, G. Lausen, *Foundations of SPARQL Query Optimization*, 13th Intl. Conf. on Database Theory (ICDT), Lausanne.
	P. Haase, T. Mathäß, M. Schmidt, A. Eberhart, U. Walther, *Semantic Technologies for Enterprise Cloud Management*, 9th Intl. Semantic Web Conference (ISWC), Industrial Track, Shanghai.
2011	**A. Schwarte, P. Haase, K. Hose, R. Schenkel, M. Schmidt**, *FedX: Optimization Techniques for Federated Query Processing on Linked Data*, 10th Intl. Semantic Web Conference (ISWC), Boston.

Selected Research Projects

Durchblick **ZIM (BMWi)**, *Acquisition and Project Management*, development of a context-sensitive conference assistance system for HMDs, since 2013

Optique **FP7 IP (EU)**, *Acquisition and Work Package Lead*, development of a platform for scalable end-user access to Big Data, since 2012

CollabCloud **KMU (BMBF)**, *Participation*, semantic, cloud-based collaboration platform for scalable analysis of distributed information, 2010–2011

Community Services

PC Membership **AAAI** (2013), **ISWC** (2009, 2013), **KICSS** (2010), **PAKDD** (2009) **EON Workshop** (2010), **SSW Workshop** (2011, 2012, 2013)

Journal Reviews **ACM Comp. Surveys**, **DKE**, **IS**, **JWS**, **TKDE**, **VLDB Journal**

External Reviews **EDBT**, **ER**, **ESWC**, **ISWC**, **ODBASE**, **SLE**, **VLDB**

Continuing Education & Conferences

2010 **International Semantic Web Conference (ISWC)**, *Shanghai*
VMworld, *San Francisco*

2011 **International Semantic Web Conference (ISWC)**, *Bonn*
Semantic Technology Conference (SemTech Biz), *San Francisco*
VMworld, *Las Vegas*

2012 **International Semantic Web Conference (ISWC)**, *Boston*
Semantic Technology Conference (SemTech Biz), *San Francisco*
Semantic Technology Conference (SemTech Biz), *London*

2013 **Java concurrency training, 1 day**, *Walldorf*

2014 **Project managament training (PMI), 5 days**, *Saarbrücken*

Languages

German **native**

English **fluent** *school, 6 years; university; job*

French **conversant** *school, 9 years*

Spanish **basic knowledge** *school, 3 years*

Hobbies

Chess **two-time Saarland youth champion**

Badminton **long-standing voluntary service as youth trainer**

Music **playing guitar**

References

References are available upon request.

Saarbrücken, December 1, 2014