Thomas Hoppe
Bernhard Humm
Anatol Reibold *Editors*

Semantic Applications

Methodology, Technology, Corporate Use

Springer Vieweg

Semantic Applications

Thomas Hoppe • Bernhard Humm
Anatol Reibold
Editors

Semantic Applications

Methodology, Technology, Corporate Use

Springer Vieweg

Editors
Thomas Hoppe
Datenlabor Berlin
Berlin, Germany

Anatol Reibold
Ontoport UG
Sulzbach, Germany

Bernhard Humm
Fachbereich Informatik
Hochschule Darmstadt
Darmstadt, Germany

ISBN 978-3-662-55432-6 ISBN 978-3-662-55433-3 (eBook)
https://doi.org/10.1007/978-3-662-55433-3

Library of Congress Control Number: 2018936133

Printed on acid-free paper

This Springer Vieweg imprint is published by the registered company Springer-Verlag GmbH, DE part of Springer Nature.
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface by the Editors

"Why are there hardly any publications about semantic applications in corporate use?"

We asked ourselves this question a few years ago. Most publications about semantic technologies and the Semantic Web are primarily centered around the technology itself, just illustrating them either by toy examples or isolated applications. By developing semantic applications for corporate use, we gained some expertise ourselves, and we were highly interested in exchanging and sharing this knowledge with peers and learning from each other. Therefore, we established the Corporate Semantic Web community in Germany.

In the years 2014–2017, we organized annual Dagstuhl workshops. Making our learnings available to the wider community was our intention from the beginning, so we published our first book "Corporate Semantic Web – Wie semantische Anwendungen in Unternehmen Nutzen stiften" (in German) in 2015. Subsequently, we published two articles on emerging trends in Corporate Semantic Web in the Informatik Spektrum magazine. Due to the high interest in those publications, we decided on starting a new book project. We have deliberately chosen English as the language for this book in order to share our experience with a worldwide community. We dedicated the 2017 Dagstuhl workshop to this book because we wanted to create more than a loose collection of articles: a coherent work which demonstrates the various aspects of engineering semantic applications.

Semantic applications are slowly but steadily being adopted by corporations and other kinds of organizations. By *semantic applications* we mean software applications which explicitly or implicitly use the semantics (i.e., the meaning) of a domain in order to improve usability, correctness, and completeness. We would like to show how to develop semantic applications in a broad range of business sectors. This book is a collection of articles describing proven methodologies for developing semantic applications including technological and architectural best practices. It is written by practitioners for practitioners. Our target audience includes software engineers and knowledge engineers, but also managers, lecturers and students. All of our co-authors are experts from industry and academia with experience in developing semantic applications. One result of the intense community effort over the years is an increasingly aligned understanding of do's and don'ts in developing such applications.

Schloss Dagstuhl, “where Computer Science meets”, has been a wonderful host and is a true incubator for developing and sharing new insights. Therefore, our first thanks go to the friendly and competent staff members of Schloss Dagstuhl.

The most important contributors to this book are, of course, our co-authors: 45 experts in their fields and also in writing high-quality texts. Thank you very much for the great collaboration and for meeting (most) deadlines! We would like to thank in particular Wolfram Bartussek, Hermann Bense, Ulrich Schade, Melanie Siegel, and Paul Walsh, who supported us as extended editorial board. Especially, we would like to thank Timothy Manning for proofreading the entire book and literally suggesting hundreds of improvements!

Finally, we would like to thank the team at Springer, particularly Hermann Engesser and Dorothea Glaunsinger for supporting us over the last years and Sabine Kathke for guiding us in this book project. The collaboration with you has truly been friendly, constructive, and smooth.

Darmstadt and Berlin, Germany, December 2017
Thomas Hoppe, Bernhard Humm, and Anatol Reibold

Contents

1 Introduction to Semantic Applications 1
Wolfram Bartussek, Hermann Bense, Thomas Hoppe, Bernhard G. Humm, Anatol Reibold, Ulrich Schade, Melanie Siegel, and Paul Walsh

2 Guide for Pragmatical Modelling of Ontologies in Corporate Settings 13
Thomas Hoppe and Robert Tolksdorf

3 Compliance Using Metadata 31
Rigo Wenning and Sabrina Kirrane

4 Variety Management for Big Data 47
Wolfgang Mayer, Georg Grossmann, Matt Selway, Jan Stanek, and Markus Stumptner

5 Text Mining in Economics 63
Melanie Siegel

6 Generating Natural Language Texts 75
Hermann Bense, Ulrich Schade, and Michael Dembach

7 The Role of Ontologies in Sentiment Analysis 87
Melanie Siegel

8 Building Concise Text Corpora from Web Contents 95
Wolfram Bartussek

9 Ontology-Based Modelling of Web Content: Example Leipzig Health Atlas 111
Alexandr Uciteli, Christoph Beger, Katja Rillich, Frank A. Meineke, Markus Loeffler, and Heinrich Herre

10 Personalised Clinical Decision Support for Cancer Care 125
Bernhard G. Humm and Paul Walsh

11 Applications of Temporal Conceptual Semantic Systems 145
Karl Erich Wolff

12 Context-Aware Documentation in the Smart Factory 163
Ulrich Beez, Lukas Kaupp, Tilman Deuschel, Bernhard G. Humm, Fabienne Schumann, Jürgen Bock, and Jens Hülsmann

13 Knowledge-Based Production Planning for Industry 4.0 181
Benjamin Gernhardt, Tobias Vogel, and Matthias Hemmje

14 Automated Rights Clearance Using Semantic Web Technologies: The DALICC Framework 203
Tassilo Pellegrini, Victor Mireles, Simon Steyskal, Oleksandra Panasiuk, Anna Fensel, and Sabrina Kirrane

15 Managing Cultural Assets: Challenges for Implementing Typical Cultural Heritage Archive's Usage Scenarios 219
Kerstin Diwisch, Felix Engel, Jason Watkins, and Matthias Hemmje

16 The Semantic Process Filter Bubble 231
Christian Fillies, Frauke Weichhardt, and Henrik Strauß

17 Domain-Specific Semantic Search Applications: Example SoftwareFinder 243
Bernhard G. Humm and Hesam Ossanloo

Index 259

List of Contributors

Wolfram Bartussek OntoPort UG, Sulzbach, Germany

Ulrich Beez Hochschule Darmstadt, Darmstadt, Germany

Christoph Beger University of Leipzig, Leipzig, Germany

Hermann Bense textOmatic AG, Dortmund, Germany

Jürgen Bock KUKA Roboter GmbH, Augsburg, Germany

Michael Dembach Fraunhofer-Institut FKIE, Wachtberg, Germany

Tilman Deuschel Hochschule Darmstadt, Darmstadt, Germany

Kerstin Diwisch Intelligent Views GmbH, Darmstadt, Germany

Felix Engel University of Hagen, Hagen, Germany

Anna Fensel University of Innsbruck, Innsbruck, Austria

Christian Fillies Semtation GmbH, Potsdam, Germany

Benjamin Gernhardt University of Hagen, Hagen, Germany

Georg Grossmann University of South Australia, Mawson Lakes, Australia

Matthias Hemmje University of Hagen, Hagen, Germany

Heinrich Herre University of Leipzig, Leipzig, Germany

Thomas Hoppe Datenlabor Berlin, Berlin, Germany

Jens Hülsmann ISRA Surface Vision GmbH, Herten, Germany

Bernhard G. Humm Hochschule Darmstadt, Darmstadt, Germany

Lukas Kaupp Hochschule Darmstadt, Darmstadt, Germany

Sabrina Kirrane Vienna University of Economics and Business, Wien, Germany

Markus Loeffler University of Leipzig, Leipzig, Germany

Wolfgang Mayer University of South Australia, Mawson Lakes, Australia

Frank A. Meineke University of Leipzig, Leipzig, Germany

Victor Mireles Semantic Web Company GmbH, Wien, Austria

Hesam Ossanloo Braunschweig, Germany

Oleksandra Panasiuk University of Innsbruck, Innsbruck, Austria

Tassilo Pellegrini University of Applied Sciences St. Pölten, St. Pölten, Austria

Anatol Reibold OntoPort UG, Sulzbach, Germany

Katja Rillich University of Leipzig, Leipzig, Germany

Ulrich Schade Fraunhofer-Institut FKIE, Wachtberg, Germany

Fabienne Schumann dictaJet Ingenieurgesellschaft mbH, Wiesbaden, Germany

Matt Selway University of South Australia, Mawson Lakes, Australia

Melanie Siegel Hochschule Darmstadt, Darmstadt, Germany

Jan Stanek University of South Australia, Mawson Lakes, Australia

Simon Steyskal Siemens AG Österreich, Wien, Austria

Henrik Strauß Semtation GmbH, Potsdam, Germany

Markus Stumptner University of South Australia, Mawson Lakes, Australia

Robert Tolksdorf Berlin, Germany

Alexandr Uciteli University of Leipzig, Leipzig, Germany

Tobias Vogel University of Hagen, Hagen, Germany

Paul Walsh NSilico Life Science, Dublin, Ireland

Jason Watkins University of Hagen, Hagen, Germany

Frauke Weichhardt Semtation GmbH, Potsdam, Germany

Rigo Wenning European Research Consortium for Informatics and Mathematics (GEIE ERCIM), Sophia Antipolis, France

Karl Erich Wolff Hochschule Darmstadt, Darmstadt, Germany

About the Authors

Wolfram Bartussek Dipl.-Inform. Wolfram Bartussek (CTO, OntoPort) studied electrical engineering with an emphasis on control theory at the Technical University of Darmstadt and was one of the first students in Computer Science in Germany at the University of Karlsruhe. He then worked for 3 years as a scientific assistant under Prof. D. L. Parnas focusing on the design and verification of operating systems kernels and on software engineering. After 25 years as CEO of a software house, he returned to research. About 2 years ago he established OntoPort with two other seasoned consultants, now focusing on language technologies and the application of ontologies. Wolfram Bartussek continued lecturing since his time as scientific assistant mainly on software engineering, programming, and software quality management at several German Universities and currently, since 1999, at the University of Applied Sciences at Darmstadt on information systems engineering, knowledge management and language technologies.

Ulrich Beez Ulrich Beez is a Research Associate at the Computer Science Department of Hochschule Darmstadt – University of Applied Sciences, Germany. At Hochschule Darmstadt in 2015, he received his master's degree while working on Terminology-Based Retrieval of Medical Publications. At Frankfurt University of Applied Sciences in 2013, he received his bachelor's degree while working on Continuous Integration. Prior to his academic career he was working for a large IT company.

Christoph Beger Christoph Beger works as a scientist, software developer and system administrator. He received the degree of Master of Computer Science with focus on Medical Informatics from the University of Leipzig, Germany. Since 2013 he administers the Growth Network CrescNet in Leipzig, which supports early detection of growth disorders of German children. Since 2016 he participates in the BMBF funded project Leipzig Health Atlas.

Hermann Bense Hermann Bense studied Computer Science at the Technical University of Dortmund and received his diploma in 1980. He founded the internet agency bense.com in 1999 and the text generation company textOmatic AG in 2015. In both companies, he works as CEO and is responsible for software development and research in the fields of ontologies, Artificial Intelligence, databases, content managements systems, search engine optimisation (SEO) and natural language generation (NLG). In the Google DNI sponsored project 3dna. news, Hermann Bense works on the hyper-personalisation of multi-language news based on Big Data from the fields of weather, finance, sports, events and traffic.

Jürgen Bock Dr. Jürgen Bock graduated in 2006 from Ulm University, Germany, as "Diplom-Informatiker" and from Griffith University, Brisbane, Australia, as Bachelor of Information Technology with Honours. Intensifying his research in the area of formal knowledge representation and Semantic Web technologies at the FZI Research Center for Information Technology in Karlsruhe, Germany, he received his PhD in 2012 from the Karlsruhe Institute of Technology (KIT). In 2015 he joined the Corporate Research department of KUKA Roboter GmbH, Augsburg, Germany, where he is coordinating a team of researchers in the area of Smart Data and Infrastructure.

Michael Dembach Michael Dembach was born on April 21st, 1987 in Brampton, Canada. He studied German and English language at the University Bonn and received his Master of Arts in 2013. He works for Fraunhofer FKIE since November 2013 and is mainly responsible for ontology engineering and natural language processing.

Tilman Deuschel Tilman Deuschel is a PhD student at the Cork Institute of Technology, Ireland. He is exploring how to improve the user experience of adaptive user interfaces, a type of graphical user interface that automatically change their appearance during runtime. He is also co-founder of a software agency that provides tools to support psychotherapy patients and therapists. He is dedicated to user centered design, requirement engineering, usability engineering and user experience design of interactive media.

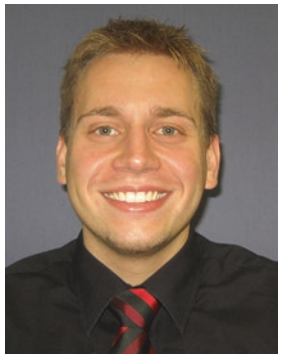
Kerstin Diwisch Kerstin Diwisch received the M.Sc. degree in Information Science from the University of Applied Sciences Darmstadt, Germany, in 2013. Currently, she is a Ph.D. candidate of Matthias Hemmje at the University of Hagen, Germany. She is working on ways to support archives in the Cultural Heritage domain to digitize and link their collections. Aside from her studies, she works as a Software Engineer at Intelligent Views GmbH, mainly in the field of semantic technologies.

Felix Engel Dr. Ing. Felix Engel studied applied computer science at the university in Duisburg-Essen, Germany where he received his Diploma in 2009 from the Department of Information Science. Since 2009 he is working for the FernUniversität in Hagen at the chair for Multimedia and Internet Applications. At this chair he successfully defended his PhD thesis in 2015 and has since been engaged as a post-doctoral researcher. Within this engagement, he is involved in different teaching activities and has contributed to various national and international projects in the area of digital preservation and knowledge management. In this context, Felix

Engel has authored and co authored various scientific publications at national and international conferences. Furthermore, he has worked as program committee member, reviewer and organizer of conferences/workshops. His research interests cover knowledge management, digital preservation, semantic technologies and information retrieval.

Anna Fensel Dr. Anna Fensel is Senior Assistant Professor at the University of Innsbruck, Austria. Anna has been a doctoral student in Informatics at the University of Innsbruck, where she has defended her PhD thesis in 2006. Prior to that, she has received a diploma in Mathematics and Computer Science equivalent to the master's degree in 2003 from Novosibirsk State University, Russia. Anna has been extensively involved in European and national projects related to Semantic technologies and their applications (in areas such as energy efficiency, smart homes, tourism and telecommunications), as an overall coordinator, a local project manager and a technical contributor. She has been a co-organizer or a Program Committee member of more than 100 scientific events, a reviewer for numerous journals, and a project proposals evaluator for the European Commission, as well as national research funding agencies. She is a (co-)author of ca. 100 refereed scientific publications.

Christian Fillies Christian Fillies studied Information Science at the Friedrich Alexander University in Erlangen-Nürnberg and at Technical University of Berlin, graduating from there with a diploma. Early on, he specialised in artificial intelligence, natural language processing and object-oriented systems with applications in projects for simulation of office processes at GMD and FAW Ulm. Following this, he worked for several years in developing tools for business process representation and modelling in Berlin, Germany and in California, USA. In 2001 Christian Fillies was one of the founders of Semtation GmbH, with its product SemTalk$^{(R)}$ focusing on the development of semantic modelling tools on the Microsoft platform and has been the engineering lead and Managing Director since then.



Benjamin Gernhardt Benjamin Gernhardt received the M.Sc. degree in Computer Science from the University of Applied Sciences Dresden, Germany, in 2013. Currently, he is a Ph.D. candidate of Matthias Hemmje at the University of Hagen, Germany. I addition to this, he is working as a IT-Manager of a Software Company in Munich, Germany.

Georg Grossmann Georg Grossmann is a Senior Lecturer at the University of South Australia. He is working on the integration of business processes and complex data structures for systems interoperability and has applied his knowledge successfully in industry projects. Current research interests include integration of service processes, ontology-driven integration and distributed event-based systems. He is currently Co-Chief Investigator in the Data to Decisions Cooperative Research Centre, Steering Committee Chair of the IEEE Conference on Enterprise Computing (EDOC) and Secretary of the IFIP Working Group 5.8 on Enterprise Interoperability.

Matthias Hemmje Matthias Hemmje received the Ph.D. degree in Computer Science from the University of Darmstadt, Germany in 1999. He is Full Professor of Computer Science at the University of Hagen. His teaching and research areas include information retrieval, multimedia databases, virtual environments, information visualization, human-computer interaction and multimedia. He holds the Chair of Multimedia and Internet Applications, at the University of Hagen.

Heinrich Herre Heinrich Herre is head of the research group "Ontologies in Medicine in Life Sciences" at the Institute for Medical Informatics, Statistics and Epidemiology (IMISE) and emeritus professor at the Institute of Computer Science of the University of Leipzig. His research interests cover topics in formal logic, applied ontology, and artificial intelligence. He is the founder of the group "Ontologies in Biomedicine and Life Sciences" of the German Society for Computer Science, and board member of several Journals and Series (Applied Ontology, Axiomathes, Categories: de Gruyter). H. Herre teaches Applied Ontology at the University of Leipzig.

Thomas Hoppe Thomas Hoppe works as a Data Scientist and Knowledge Engineer. He is a lecturer for information systems, content management, data- and text mining and search technology at the Computer Science Department of the Hochschule für Technik und Wirtschaft Berlin – University of Applied Sciences, Germany. He received his doctorate from the University of Dortmund (now Technical University Dortmund) and the degree Dipl.-Inform. from the Technical University Berlin, Germany. In 2008, together with three associates, he founded Ontonym GmbH for the development and commercialization of semantic applications, which he led as CEO until 2015. In 2014 he founded the Datenlabor Berlin (www.datenlabor.berlin) for the purpose of developing customer specific data products for SMEs and for quality assurance and certification of semantic applications and predictive models.

Jens Hülsmann Jens Hülsmann is a Senior Research Scientist at ISRA Surface Vision GmbH, a manufacturer of machine vision solutions from Herten, Germany. He joined the company in 2014 and is in charge of the development infrastructure software for optical inspection systems and image processing algorithms. Prior to this, he worked as a Research Associate for the Smart Embedded Systems Group at the University of Osnabrück. His research was focused on the application of robust classification in industrial environments and on uncertainty modelling. In 2009 he graduated with a Master of Science in physics and computer science.

Bernhard G. Humm Bernhard G. Humm is a professor at the Computer Science Department of Hochschule Darmstadt – University of Applied Sciences, Germany. He coordinates the Ph.D. programme and is managing director of the institute of applied informatics, Darmstadt (aiDa). He is running several national and international research projects in cooperation with industry and research organizations and publishes his results regularly. Before re-entering university, he worked in the IT industry for 11 years as a software architect, chief consultant, IT manager and head of the research department of a large software company in Germany.

Lukas Kaupp Lukas Kaupp is a Research Associate at the Computer Science Department of Hochschule Darmstadt – University of Applied Sciences, Germany. At Hochschule Darmstadt in 2016, he received his master's degree while working on conception, development and evaluation of a container-based cloud platform with focus on user collaboration. In 2014, he received his bachelor's degree while working on a framework to reverse engineer java-based components from source code using a genetic algorithm. Prior to his academic career, he was working for a large IT consulting company and a web-focused startup.

Sabrina Kirrane Dr. Sabrina Kirrane is a postdoctoral researcher at the Vienna University of Economics and Business (WU) and co-director of WU's Privacy and Sustainable Computing Lab. Prior to taking up the position at WU, she was a researcher at the Insight Centre for Data Analytics, Ireland. Sabrina's current research focuses on privacy implications associated with the publishing, interlinking and reasoning over personal information. She is particularly interested in the multidisciplinary nature of privacy preservation that spans the humanities, social sciences, information technology, security and legal domains. Moreover, Sabrina is the technical coordinator of an EU project around building technical solutions to support the upcoming European General Data Protection Regulation (GDPR).

Markus Loeffler Prof. Markus Loeffler is full Professor and Head of the Institute for Medical Informatics, Statistics and Epidemiology (IMISE) at the University of Leipzig. His scientific research covers a wide range of subjects in biometry, biomathematics, clinical trial research, systems biology and computational biology. He is Scientific Director of the Center for Clinical Trials (ZKS), Scientific Director of the Interdisciplinary Center for Bioinformatics (IZBI), Director of the LIFE Research Center for Civilisation Diseases, and PI of the consortium SMITH in the BMBF Medical Informatics Initiative.

Wolfgang Mayer Wolfgang Mayer is a Senior Lecturer at the University of South Australia. His research interests include artificial intelligence and knowledge representation methods and their applications in industrial contexts. He has made contributions to information modelling and natural language knowledge extraction methods, scalable data architectures, causal inference from data, technologies for software systems interoperability, expert systems for process and product customization, and fault diagnosis in software systems.

Frank A. Meineke Frank A. Meineke works as Data Scientist and Project Manager at the Institute of Medical Informatics, Statistics and Epidemiology in Leipzig. He studied applied computer science with a focus on computer linguistics in Koblenz-Landau and graduated in 1994 with a degree (Dipl Inform) in computer science. He worked at the Universities of Cologne and Leipzig in the field of systemic medicine and software development (oncologic treatment planning) and received his doctorate (Dr. rer. med.) in 2007 from the medical faculty of Leipzig for his work on stem cell development. He took over a leading position in an IT-department of a local hospital and returned to the University of Leipzig in 2010, where he now works as IT-Coordinator of the Institute and heads up IT groups in various clinical research projects (IFB Obesity, Leipzig Health Atlas, SMITH).

Victor Mireles Victor Mireles is researcher at Semantic Web Company GmbH, Vienna, Austria. He holds B.Sc. and M.Sc. degrees in computer science from the National Autonomous University of Mexico (UNAM). He has conducted research in Mexico, Germany and Austria, in the fields of Computational Biology, Natural Language Processing, and Matrix Decompositions. His current research focuses on the interface between semantics, machine learning, and automatic reasoning.

Hesam Ossanloo Hesam was born in Iran in 1983. His passion for discovering the unknown ignited the will to get out of Iran to learn more about other cultures, science and humanity. He received his M.Sc. degree in computer science from the Hochschule Darmstadt – University of Applied Sciences where he is also doing his Ph.D. in the field of semantic web. In the morning, he is a senior consultant and at night, a scientist doing research on how to improve the search experience for the end-users. In his Ph.D. he is part of a team, which is developing a semantic application for finding software components called "SoftwareFinder" (www.softwarefinder.org). SoftwareFinder uses a domain-specific ontology for offering semantic functionalities to make the user's life a bit easier.

He is also passionate about working on the frontier of science, which is why he is part of the project "Mission To the Moon" at the PTScientists GmbH (http://ptscientists.com), working on landing a rover on the moon.

Oleksandra Panasiuk Oleksandra Panasiuk is a Ph.D. student at the Semantic Technology Institute Innsbruck, Department of Computer Science, University of Innsbruck, Austria. She holds B.Sc. and M.Sc. degrees in Informatics from Taras Shevchenko National University of Kyiv, Ukraine. Her current research focuses on structure data modelling and semantic annotations. She is a co-chair of the Schema Tourism Working Group by STI International.

Tassilo Pellegrini Tassilo Pellegrini is a professor at the Department of Media Economics of the University of Applied Sciences in St. Pölten, Austria. His research interest lies at the intersection of media convergence, technology diffusion and ICT policy. Besides his focus on media economics, he is doing research on the adoption of semantic technologies and the Semantic Web among the media industry. He is member of the International Network for Information Ethics (INIE), the African Network of Information Ethics (ANIE) and the Deutsche Gesellschaft für Publizistik und Kommunikationswissenschaft (DGPUK). He is co-founder of the Semantic Web Company in Vienna, editor and author of several publications in his research areas and conference chair of the annual I-SEMANTICS conference series founded in 2005.

Anatol Reibold Anatol Reibold studied mathematics and mechanical engineering at the University of Novosibirsk. Being a passionate mathematician, his main interests focus on semantic technologies, data analysis, data science, optimization, transport logistics, coloured Petri nets, GABEK, DRAKON, and polyductive theory. Currently he works as a supply chain analyst and business mathematician at Raiffeisen Waren-Zentrale Rhein-Main eG. He is also one of the co-founders of OntoPort and their chief data scientist.

Katja Rillich Katja Rillich finished her biology studies at the University of Leipzig with the diploma in 2003. In 2008 she finished her doctoral thesis on the subject of light-induced calcium increases in glial cells of the guinea pig retina and continued working as a research associate at the Paul-Flechsig-Institute for brain research (University of Leipzig) until 2010. Since July 2010, she works at the Institute for Medical Informatics, Statistics and Epidemiology of the University of Leipzig and has been involved in several groups and topics.

Ulrich Schade Ulrich Schade received the Ph.D. degree in Linguistics from Bielefeld University, Germany, in 1990. He is a senior research scientist with Fraunhofer FKIE where he leads a research unit on information analysis. Additionally, he is associate professor to the Department of English, American, and Celtic Studies, at Rheinische Friedrich-Wilhelms-Universität Bonn.

Fabienne Schumann Fabienne Schumann is an Information Manager at dictaJet Ingenieurgesellschaft mbH, a service provider for technical documentation and process optimisation in the field of information management, located in Wiesbaden, Germany. At dictaJet, Fabienne is in charge of the ProDok 4.0 project coordination. In 1994 she graduated from the Hochschule Darmstadt, University of Applied Sciences, Germany, with a diploma degree in Documentation and Information Sciences. Prior to this, she graduated from the Sorbonne University in Paris, France, with a Maîtrise degree in Applied Foreign Languages.

Matt Selway Matt Selway is a Research Fellow at the University of South Australia. He is working on the extraction of process models from text and their transformation into executable models for a variety of simulation environments, as well as contributing to interoperability projects in industrial contexts. His research interests include Natural Language Understanding, Knowledge Engineering and integration, and software systems interoperability.

Melanie Siegel Melanie Siegel is a Professor for Information Science at Hochschule Darmstadt, with a strong background in language technology. She holds a Ph.D. in linguistics and a Habilitation in linguistics and computational linguistics from Bielefeld University. From 1993 to 2000, she worked in the Verbmobil project on machine translation and was responsible for the semantic analysis of Japanese. Her Ph.D. thesis, which was finished in 1996, concerns mismatches in Japanese to German machine translation. The habilitation thesis, which was finished in 2006, is titled “JACY – A Grammar for Annotating Syntax, Semantics and Pragmatics of Written and

Spoken Japanese for NLP Application Purposes". She participated in research projects at the German Research Center for Artificial Intelligence (DFKI) and Saarland University between 1995 and 2006. From 2006 to 2012, she worked as a computational linguist and head of research and innovation at Acrolinx GmbH in Berlin, in the area of automatic consistency checking of technical documentation. www.melaniesiegel.de

Jan Stanek Jan Stanek is a lecturer at the University of South Australia (in Business intelligence and Analytics and Health informatics). He is interested in internal medicine, medical clinical databases, medical process analytics and modelling, interoperability in health (especially in pathology and in general practice) and clinical decision support systems. He led projects on the design and development of clinical databases for genetic data collection, and was a member of the team developing a standard for clinical databases with the Royal College of Pathologists of Australasia. Currently, he is working on projects in pathology process analytics and modelling and on general practice data collection and analytics.

Simon Steyskal Simon Steyskal holds master's degrees in Software Engineering & Internet Computing as well as Information & Knowledge Management, both from the Vienna University of Technology (TU) and joined Vienna University of Business and Economics in January 2014 as a Ph.D. candidate under the supervision of Prof. Axel Polleres. He is working as a research scientist for Siemens AG Austria and was involved in a joint research project between Siemens AG and WU. He was a member of the W3C RDF Data Shapes working group and involved in the standardization process of a W3C permissions and obligations expression language carried out by the W3C Permissions and Obligations Expression working group.

Henrik Strauß Henrik Strauß holds a master's degree in Business Information Science from the Technische Hochschule Brandenburg, University of Applied Sciences. He is working on projects concerning process modelling tools and the use of ontologies on the Microsoft platform at Semtation GmbH in Potsdam, Germany.

Markus Stumptner Markus Stumptner is Professor for Computer Science at the University of South Australia (UniSA) and head of the Knowledge & Software Engineering research group. He is the Director of the Advanced Computing Research Centre (ACRC) at UniSA, and is Research Program Lead for Data Management in the Data to Decisions Collaborative Research Centre (D2D CRC). He has worked on knowledge representation and reasoning in a variety of problem areas such as configuration, design, diagnosis, interoperability, service composition and automated model creation from natural language.

Robert Tolksdorf Robert Tolksdorf was a Professor for Computer Science at Freie Universität Berlin. His group, Networked Information Systems, was involved in numerous projects on applications of the Semantic Web from 2002 until 2016. He holds a Dr.-Ing. from Technische Universität where he also received his Diplom-Informatiker after studying Computer Science. He was one of the co-founders of Ontonym GmbH, Berlin. www.robert-tolksdorf.de

Alexandr Uciteli Alexandr Uciteli received the degree of Master of Computer Science from the University of Leipzig, Germany in 2008. After that, he worked as a software developer at the Max Planck Institute for Evolutionary Anthropology in Leipzig. Since August 2008, he works as Research Associate at the Institute for Medical Informatics, Statistics and Epidemiology (IMISE) of the University of Leipzig and belongs to the research group “Ontologies in Medicine and Life Sciences”. His research is focused, inter alia, on modelling clinical metadata, ontology-based information retrieval and ontology-driven software development.

Tobias Vogel Tobias Vogel received the Ph.D. degree in Computer Science from the University of Hagen, Germany in 2012. Currently, he is a post-doctoral researcher at the University of Hagen in the research team of Matthias Hemmje. In addition to this, he is working as Managing Director of a Technology Company in Munich, Germany.

Paul Walsh Paul Walsh is Professor of Computer Science in the Cork Institute of Technology (CIT) and a Senior Visiting Research Fellow at the University of Edinburgh where he manages research in medical informatics and bioinformatics. He holds a Ph.D., M.Sc. and B.Sc. Hons in Computer Science from the National University of Ireland and has a long list of publications including a number with outstanding paper awards. He has consulted on a wide spectrum of projects, ranging from startup technology companies to managing projects for global corporations. He is a Science Foundation Ireland Research Fellow and is funded under national and international research programmes, such as the Enterprise Ireland and Horizon 2020, where he oversees research in data analytics, machine learning and high performance computing. He is founder of NSilico Life Science, a software company that provides machine learning driven solutions for global clients.

Jason Watkins Jason Watkins received his B.Sc. degree in Automation Technology from the University of Applied Sciences Mannheim, Germany, in 2011. At his first job at Bihl+Wiedemann GmbH, he was designing firmware for fieldbus interfaces of industrial devices as an embedded software developer. In 2014, he joined John Deere as a design engineer developing software for frontloader applications, electro-hydraulic front axle suspensions and electro-hydraulic cabin suspensions. Alongside his professional activities, he started studying a master's program in Electrical Engineering at the University of Hagen. Currently, he is working on his master's thesis in the field of Ontology Matching under supervision of Kerstin Diwisch and Matthias Hemmje.

Frauke Weichhardt Frauke Weichhardt holds a Ph.D. in Business Informatics after graduating in Business and Mechanical Engineering from the Technical University of Berlin. She was a lecturer for Business Informatics at TU Berlin and then worked as a consultant in business process design and knowledge management projects before, in 2001, she became one of the founders of Semtation GmbH with its product SemTalk(R), focusing on the development of semantic modelling tools on the Microsoft platform and is working as Managing Director for the company.

Rigo Wenning Rigo Wenning is the legal counsel of the European Research Consortium for Informatics and Mathematics (GEIE ERCIM), a joint venture of CWI, Fraunhofer, IIT-CNR, INRIA and ICS-FORTH. GEIE ERCIM is the European host of the World Wide Web Consortium (W3C) where Rigo also acts as legal counsel. He is an attorney at law registered with the Munich Bar Association and affiliated with the law firm Frösner & Stadler in Freising. He studied law in Saarbrücken and Nancy. Apart from legal advice, Rigo does research in the area of privacy and security. He has been involved in the PRIME project, the Primelife project and was technical coordinator of the STREWS project on IT-Security. He is now involved with the SPECIAL project that creates a scalable policy-aware linked data architecture for privacy transparency and compliance for big data. Rigo Wenning is a member of the board of Deutscher EDV-Gerichtstag and also a member of the scientific council of the Leibniz Information Centre for Science and Technology University Library (TIB).

Karl Erich Wolff Karl Erich Wolff studied mathematics and physics, received the Ph.D. in 1973 and the habilitation in 1978 in mathematics from Gießen University, Germany, where he lectured from 1974 until 1980 at the Mathematical Institute. From 1980 until his retirement in 2011, he taught mathematics at Darmstadt University of Applied Sciences. Since 1983, he is also a member of the research group "Concept Analysis" in the Department of Mathematics at the university of technology TU Darmstadt. Since 2011, he is president of the Ernst-Schröder-Center for Conceptual Knowledge Processing. He has extended Formal Concept Analysis to Temporal Concept Analysis.

Introduction to Semantic Applications

1

Wolfram Bartussek, Hermann Bense, Thomas Hoppe, Bernhard G. Humm, Anatol Reibold, Ulrich Schade, Melanie Siegel, and Paul Walsh

Key Statements

1. Semantic applications today provide benefits to numerous organisations in business sectors such as health care, finance, industry, and the public sector.
2. Semantic applications use the semantics of a domain in order to improve usability, correctness, and completeness.

W. Bartussek (✉) · A. Reibold
OntoPort UG, Sulzbach, Germany
e-mail: w.bartussek@ontoport.de; a.reibold@ontoport.de

H. Bense
textOmatic AG, Dortmund, Germany
e-mail: hermann.bense@textomatic.ag

T. Hoppe
Datenlabor Berlin, Berlin, Germany
e-mail: thomas.hoppe@datenlabor.berlin

B. G. Humm
Hochschule Darmstadt, Darmstadt, Germany
e-mail: bernhard.humm@h-da.de

U. Schade
Fraunhofer-Institut FKIE, Wachtberg, Germany
e-mail: ulrich.schade@fkie.fraunhofer.de

M. Siegel
Hochschule Darmstadt, Darmstadt, Germany
e-mail: melanie.siegel@h-da.de

P. Walsh
NSilico Life Science, Dublin, Ireland
e-mail: paul.walsh@nsilico.com

T. Hoppe et al. (eds.), *Semantic Applications*,
https://doi.org/10.1007/978-3-662-55433-3_1

3. Developing semantic applications requires methodological skills, e.g., ontology engineering, quality assurance for ontologies, and licence management.
4. Various technologies are available for implementing semantic applications, e.g., data integration, semantic search, machine learning, and complex event processing.

1.1 Introduction

Semantic applications today provide benefits to numerous corporations and other kinds of organisations worldwide. This book describes proven methodologies for developing semantic applications including technological and architectural best practices: from data to applications. The methodologies are backed up by a large number of applications that are in corporate use today.

Figure 1.1 gives overview of the book chapters and which methodologies, technologies, and applications in corporate use they cover.

In this chapter, we give an introduction to semantic applications and provide an overview of the most prominent methodologies, technologies, and applications in corporate use.

1.2 Foundations

Since not everybody is acquainted with the terminology in the field of semantic applications, we provide definitions of the most important terms used in this book. These definitions are neither intended to be complete nor perfectly consistent with scientific definitions. They are intended to show the intuition behind some of the major terms.

Semantics tries to capture and normalise the relationships between words (respectively terms, phrases, symbols etc.) and their meaning. For example, the word "cancer" can have the meaning of a disease or a zodiac sign. The concrete meaning of a term in a formalisation is usually determined by its context, i.e. the other terms used for its definition and the terms related to it. Such formalisations are often called ontologies.

In the context of semantic applications, an *ontology* is an explicit representation of the semantics of the used terms, usually restricted to a specific application domain [1]. The term *ontology* has been defined as a "formal, explicit specification of a conceptualisation" [2], emphasising that the terms are explicitly chosen on a particular level of granularity, or as a "formal, explicit specification of a shared conceptualisation" [3] additionally emphasising that its purpose is to share the meaning of terms between different stakeholders. For example, an ontology for medicine may define melanoma as a disease, Warfarin as a medication, and the relationship between both indicating that Warfarin can be used for treating melanoma. In fact, Fig. 1.1 shows a simple ontology, specifying applications, business sectors for corporate use, methodologies and technologies, and indicating relationships to chapters of this book.

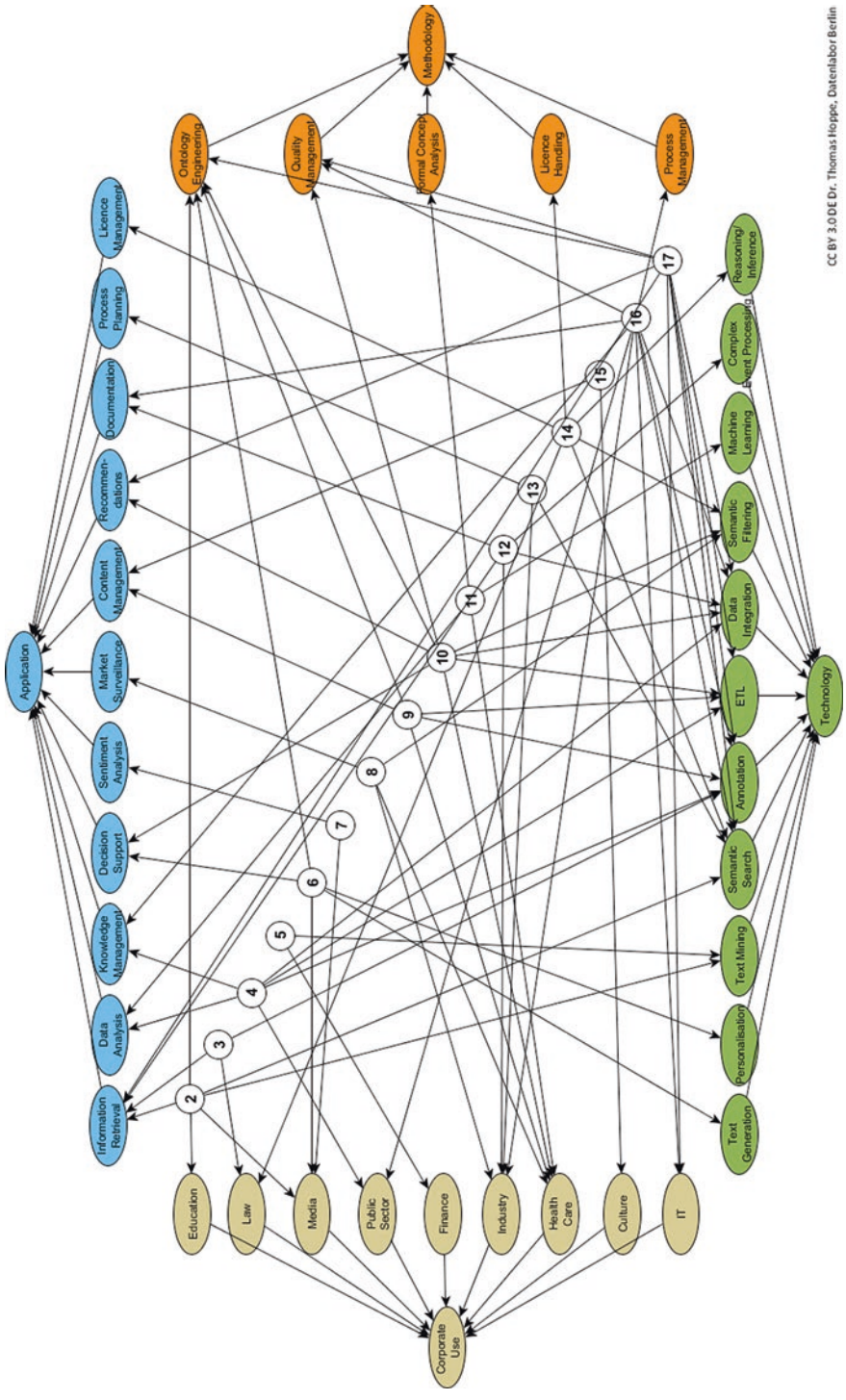


Fig. 1.1 Overview of the book chapters (White circles indicate chapter numbers. The arrows exiting the white circles indicate the methodologies, technologies, applications and corporate domains covered by each chapter.)

Various forms of ontologies are used under different terms, depending on the complexity of relationships provided. Lassila and McGuiness [4] differentiate between "lightweight ontologies" and "heavyweight ontologies". Lightweight ontologies in particular include controlled vocabularies, thesauri and informal taxonomies. *Controlled vocabularies* are, in their simplest form, just a list of prominent terms used in an application domain, e.g., "melanoma" in the medical domain. *Taxonomies* add hierarchies of broader/narrower terms to controlled vocabularies, e.g., melanoma is a type of cancer, which is a disease. *Thesauri* add additional information to the terms in taxonomies, including prefered names (e.g., "melanoma"), synonyms ("malignant melanoma", "cutaneous melanoma"), and relations to other terms (e.g. "see also skin cancer"). Heavyweight ontologies extend thesauri by giving the informal hierarchical broader/narrower term relation (i.e. is_a relation) a formal foundation and extending the expressiveness by additional fine grained relations (e.g., gene CRYBG1 is associated with melanoma), definitions (e.g., "melanoma is a malignant neoplasm comprised of melanocytes typically arising in the skin"; Source: National Cancer Institute Thesaurus), properties, and metadata. The focus of ontologies is not only the terminology of a domain, but also the inherent ontological structure, i.e. which objects exist in the application domain, how they can be organised into classes, called *concepts*, and how these classes are defined and related.

Ontology engineering is the discipline of building ontologies. It includes methodologies and best practices, e.g., incremental ontology development in tight collaboration with domain experts, and ranges from text analysis of available documents and information sources, over the extraction of information from various data sources, to the modelling of an ontology. *Ontology modelling* covers either the adaptation of existing ontologies, the merging and aligning of several ontologies covering different domain aspects, or modelling the needed ontology from scratch. As a key methodology, ontology engineering is covered in further detail in Sect. 1.4.1.

A *semantic application* is a software application which explicitly or implicitly uses the semantics of a domain. This is to improve usability, validity, and completeness. An example is semantic search, where synonyms and related terms are used for enriching the results of a simple text-based search. Ontologies are the centerpiece of semantic applications.

Information retrieval usually subsumes different approaches for obtaining information based on a specific information need from a collection of information sources. Besides pure textual information, it also usually covers image, speech and video retrieval. Nowadays, the most prominent examples of information retrieval applications are general-purpose search engines, such as Google, Yahoo, and Bing. Today, such search engines include semantic search, making them semantic applications. For example, entering "When was JFK born?" in the Google search will result in an information box containing "John F. Kennedy/Date of birth: May 29, 1917". In contrast to general-purpose search engines, domain-specific search applications have a narrower focus but provide more semantic depth. Examples are hotel and travel portals, partner search, job portals, used vehicles websites, etc.