Torsten Eymann
Franziska Klügl
Winfried Lamersdorf
Matthias Klusch
Michael N. Huhns (Eds.)

LNAI 3550

Multiagent System Technologies

Third German Conference, MATES 2005
Koblenz, Germany, September 2005
Proceedings

Springer

Torsten Eymann Franziska Klügl
Winfried Lamersdorf Matthias Klusch
Michael N. Huhns (Eds.)

Multiagent System Technologies

Third German Conference, MATES 2005
Koblenz, Germany, September 11-13, 2005
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Torsten Eymann
Universität Bayreuth, Wirtschaftsinformatik (BWL VII)
Universitätsstrasse 30, 95440, Bayreuth, Germany
E-mail: eymann@uni-bayreuth.de

Franziska Klügl
Universität Würzburg
Department of Artificial Intelligence and Applied Computer Science
Am Hubland, 97074 Würzburg, Germany
E-mail: kluegl@informatik.uni-wuerzburg.de

Winfried Lamersdorf
Universität Hamburg
Department of Informatics, Distributed and Information Systems (VSIS)
Vogt-Kölln-Str. 30, 22527 Hamburg, Germany
E-mail: lamersd@informatik.uni-hamburg.de

Matthias Klusch
DFKI, German Research Center for Artificial Intelligence
Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany
E-mail: klusch@dfki.de

Michael N. Huhns
University of South Carolina, Department of Computer Science and Engineering
Columbia, SC 29208, USA
E-mail: huhns@sc.edu

Library of Congress Control Number: 2005931591
CR Subject Classification (1998): I.2.11, I.2, C.2.4, D.2.12, D.1.3, J.1

ISSN	0302-9743
ISBN-10	3-540-28740-X Springer Berlin Heidelberg New York
ISBN-13	978-3-540-28740-7 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11550648 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 3550
Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Preface

After two successful MATES conferences in Erfurt 2003 and 2004, the 3rd German conference on Multi-agent System Technologies (MATES 2005) took place in Koblenz, Germany, in September 2005, and was co-located with the 28th German Conference on Artificial Intelligence (KI 2005).

Building on other agent-related events in Germany in the past, and organized by the GI German Special Interest Group on Distributed Artificial Intelligence, the MATES conference series aims at promoting the theory and applications of agents and multiagent systems. Incorporating the 9th International Workshop on Cooperative Information Agents (CIA 2005), the topics of interest for MATES 2005 also covered the fields of intelligent information agents and systems for the Internet and the (Semantic) Web.

As in recent years, MATES 2005 provided a distinguished, lively and interdisciplinary forum for researchers, users, and developers of agent technology, to present and discuss the latest advances of research and development in the area of autonomous agents and multiagent systems. Accordingly, the topics of MATES 2005 covered the whole range from the theory to applications of agent- and multiagent technology. The technical program included a total of 24 scientific talks, and demonstrations of selected running agent systems, and both the MATES 2005 Best Paper and the CIA 2005 System Innovation awards.

The international Program Committee for MATES 2005 selected carefully 14 out of 54 submissions from all over the world to be accepted as full papers, and an additional 5 short papers as well as 5 posters to be presented. The program also included four distinguished invited speakers: Karl Aberer (EPF Lausanne, Switzerland), John-Jules C. Meyer (Utrecht University, The Netherlands), Steffen Staab (Universität Koblenz, Germany), and jointly with KI 2005, Luc Steels (SONY Computer Science Lab Paris and Free University of Brussels), as well as a doctoral colloquium and a mentoring program.

Finally, as general co-chairs and PC co-chairs, and in the name of all members of the Steering Committee, we would like to thank all authors of submitted papers and all invited speakers for their contributions, all members of the Program Committee as well as other reviewers for their careful, critical, and thoughtful reviews, and all local conference organizers and others involved in helping to make MATES 2005 a success. In addition, we would like to explicitly thank our sponsors AgentLink III, Whitestein Technologies, Siemens, and the German Computer Society (GI), whose financial support helped to make this event possible.

We hope the attendees enjoyed MATES 2005 and the Koblenz conference site both scientifically and socially and will continue to support MATES as a conference series with many more successful events to come in the future!

June 2005

Torsten Eymann, Franziska Klügl, Winfried Lamersdorf,
Michael Huhns, Matthias Klusch

Organization

General Co-chairs

Matthias Klusch	DFKI Saarbrücken, Germany
Michael Huhns	University of South Carolina, USA

Program Co-chairs

Torsten Eymann	Bayreuth University, Germany
Franziska Klügl	Würzburg University, Germany
Winfried Lamersdorf	Hamburg University, Germany

Program Committee

Karl Aberer	EPF Lausanne, Switzerland
Elisabeth Andre	University of Augsburg, Germany
Bernhard Bauer	University of Augsburg, Germany
Wolfgang Benn	TU Chemnitz, Germany
Monique Calisti	Whitestein Technologies AG, Zürich, Switzerland
Cristiano Castelfranchi	CNR, Italy
Thomas Christaller	Fraunhofer AIS, Germany
Rosaria Conte	NRC Rome, Italy
Stephen Cranefield	University of Otago, New Zealand
Mehdi Dastani	University of Utrecht, The Netherlands
Yves Demazeau	Leibniz/IMAG, France
Jörg Denziger	University of Calgary, Canada
Klaus Fischer	DFKI Saarbrücken, Germany
Ana Garcia Serrano	TU Madrid, Spain
Fausto Giunchiglia	University of Trento, Italy
Marie-Pierre Gleizes	IRIT Toulouse, France
Rune Gustavsson	TH Blekinge, Sweden
Heikki Helin	TeliaSonera, Helsinki, Finland
Heinrich Hussmann	Universität München, Germany
Toru Ishida	University of Kyoto, Japan
Stefan Kirn	Uiversität Hohenheim, Germany
Ryszard Kowalczyk	TU Swinburne, Australia
Daniel Kudenko	University of York, UK
Jürgen Lind	AgentLab München, Germany
Gabriela Lindemann	HU Berlin, Germany
Jiming Liu	Hong Kong Baptist University, China

Stefano Lodi	University of Bologna, Italy
Beatriz Lopez	University of Girona, Spain
Jörg Müller	Siemens, Germany
Heinz-Jürgen Müller	Berufsakademie Mannheim, Germany
Werner Nutt	Heriot-Watt University, Edinburgh, UK
Andrea Omicini	University of Bologna, Italy
Sascha Ossowski	University of Rey Juan Carlos Madrid, Spain
Michal Pechoucek	TU Prague, Czech Republic
Paolo Petta	OEFAI Vienna, Austria
Stefan Poslad	Queen Mary University of London, UK
Frank Puppe	Universität Würzburg, Germany
Alois Reitbauer	ProFACTOR, Austria
Franz Rothlauf	Universität Mannheim, Germany
Marie-Christine Rousset	Université de Paris-Sud, France
Heiko Schuldt	UMIT Innsbruck, Austria
Onn Shehory	IBM Research, Israel
John Shepherdson	British Telecom, UK
Steffen Staab	Universität Koblenz, Germany
Rudi Studer	TU Karlsruhe, Germany
Ingo Timm	TZI, Universität Bremen, Germany
Robert Tolksdorf	TU Berlin, Germany
Lin Uhrmacher	Universität Rostock, Germany
Rainer Unland	Universität Duisburg-Essen, Germany
Thomas Uthmann	Johannes-Gutenberg Universität Mainz, Germany
Wiebe Van der Hoek	University of Liverpool, UK
Laszlo Zsolt Varga	MTA SZTAKI, Hungary
Daniel Veit	TU Karlsruhe, Germany
Ning Zhong	Maebashi IT, Japan

Additional Reviewers

Lars Braubach
Philippe Cudre-Mauroux
Gianluca Moro
Roman Schmidt
Alexander Pokahr
Jan Broersen
Fabius Klemm
M. Birna van Riemsdijk
Leendert van der Torre
Pere Vila

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 3662: C. Baral, G. Greco, N. Leone, G. Terracina (Eds.), Logic Programming and Nonmonotonic Reasoning. XIII, 454 pages. 2005.

Vol. 3642: D. Ślezak, J. Yao, J.F. Peters, W. Ziarko, X. Hu (Eds.), Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Part II. XXIV, 738 pages. 2005.

Vol. 3641: D. Ślezak, G. Wang, M.S. Szczuka, I. Düntsch, Y. Yao (Eds.), Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Part I. XXIV, 742 pages. 2005.

Vol. 3632: R. Nieuwenhuis (Ed.), Automated Deduction – CADE-20. XIII, 459 pages. 2005.

Vol. 3626: B. Ganter, G. Stumme, R. Wille (Eds.), Formal Concept Analysis. X, 349 pages. 2005.

Vol. 3625: S. Kramer, B. Pfahringer (Eds.), Inductive Logic Programming. XIII, 427 pages. 2005.

Vol. 3620: H. Muñoz-Avila, F. Ricci (Eds.), Case-Based Reasoning Research and Development. XV, 654 pages. 2005.

Vol. 3614: L. Wang, Y. Jin (Eds.), Fuzzy Systems and Knowledge Discovery, Part II. XLI, 1314 pages. 2005.

Vol. 3613: L. Wang, Y. Jin (Eds.), Fuzzy Systems and Knowledge Discovery, Part I. XLI, 1334 pages. 2005.

Vol. 3607: J.-D. Zucker, L. Saitta (Eds.), Abstraction, Reformulation and Approximation. XII, 376 pages. 2005.

Vol. 3596: F. Dau, M.-L. Mugnier, G. Stumme (Eds.), Conceptual Structures: Common Semantics for Sharing Knowledge. XI, 467 pages. 2005.

Vol. 3593: V. Mařík, R. W. Brennan, M. Pěchouček (Eds.), Holonic and Multi-Agent Systems for Manufacturing. XI, 269 pages. 2005.

Vol. 3587: P. Perner, A. Imiya (Eds.), Machine Learning and Data Mining in Pattern Recognition. XVII, 695 pages. 2005.

Vol. 3584: X. Li, S. Wang, Z.Y. Dong (Eds.), Advanced Data Mining and Applications. XIX, 835 pages. 2005.

Vol. 3581: S. Miksch, J. Hunter, E. Keravnou (Eds.), Artificial Intelligence in Medicine. XVII, 547 pages. 2005.

Vol. 3577: R. Falcone, S. Barber, J. Sabater-Mir, M.P. Singh (Eds.), Trusting Agents for Trusting Electronic Societies. VIII, 235 pages. 2005.

Vol. 3575: S. Wermter, G. Palm, M. Elshaw (Eds.), Biomimetic Neural Learning for Intelligent Robots. IX, 383 pages. 2005.

Vol. 3571: L. Godo (Ed.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty. XVI, 1028 pages. 2005.

Vol. 3559: P. Auer, R. Meir (Eds.), Learning Theory. XI, 692 pages. 2005.

Vol. 3558: V. Torra, Y. Narukawa, S. Miyamoto (Eds.), Modeling Decisions for Artificial Intelligence. XII, 470 pages. 2005.

Vol. 3554: A. Dey, B. Kokinov, D. Leake, R. Turner (Eds.), Modeling and Using Context. XIV, 572 pages. 2005.

Vol. 3550: T. Eymann, F. Klügl, W. Lamersdorf, M. Klusch, M.N. Huhns (Eds.), Multiagent System Technologies. XI, 246 pages. 2005.

Vol. 3539: K. Morik, J.-F. Boulicaut, A. Siebes (Eds.), Local Pattern Detection. XI, 233 pages. 2005.

Vol. 3538: L. Ardissono, P. Brna, A. Mitrovic (Eds.), User Modeling 2005. XVI, 533 pages. 2005.

Vol. 3533: M. Ali, F. Esposito (Eds.), Innovations in Applied Artificial Intelligence. XX, 858 pages. 2005.

Vol. 3528: P.S. Szczepaniak, J. Kacprzyk, A. Niewiadomski (Eds.), Advances in Web Intelligence. XVII, 513 pages. 2005.

Vol. 3518: T.B. Ho, D. Cheung, H. Liu (Eds.), Advances in Knowledge Discovery and Data Mining. XXI, 864 pages. 2005.

Vol. 3508: P. Bresciani, P. Giorgini, B. Henderson-Sellers, G. Low, M. Winikoff (Eds.), Agent-Oriented Information Systems II. X, 227 pages. 2005.

Vol. 3505: V. Gorodetsky, J. Liu, V. A. Skormin (Eds.), Autonomous Intelligent Systems: Agents and Data Mining. XIII, 303 pages. 2005.

Vol. 3501: B. Kégl, G. Lapalme (Eds.), Advances in Artificial Intelligence. XV, 458 pages. 2005.

Vol. 3492: P. Blache, E. Stabler, J. Busquets, R. Moot (Eds.), Logical Aspects of Computational Linguistics. X, 363 pages. 2005.

Vol. 3488: M.-S. Hacid, N.V. Murray, Z.W. Raś, S. Tsumoto (Eds.), Foundations of Intelligent Systems. XIII, 700 pages. 2005.

Vol. 3487: J. Leite, P. Torroni (Eds.), Computational Logic in Multi-Agent Systems. XII, 281 pages. 2005.

Vol. 3476: J. Leite, A. Omicini, P. Torroni, P. Yolum (Eds.), Declarative Agent Languages and Technologies II. XII, 289 pages. 2005.

Vol. 3464: S.A. Brueckner, G.D.M. Serugendo, A. Karageorgos, R. Nagpal (Eds.), Engineering Self-Organising Systems. XIII, 299 pages. 2005.

Vol. 3452: F. Baader, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning. XI, 562 pages. 2005.

Vol. 3451: M.-P. Gleizes, A. Omicini, F. Zambonelli (Eds.), Engineering Societies in the Agents World V. XIII, 349 pages. 2005.

Vol. 3446: T. Ishida, L. Gasser, H. Nakashima (Eds.), Massively Multi-Agent Systems I. XI, 349 pages. 2005.

Vol. 3445: G. Chollet, A. Esposito, M. Faundez-Zanuy, M. Marinaro (Eds.), Nonlinear Speech Modeling and Applications. XIII, 433 pages. 2005.

Vol. 3438: H. Christiansen, P.R. Skadhauge, J. Villadsen (Eds.), Constraint Solving and Language Processing. VIII, 205 pages. 2005.

Vol. 3430: S. Tsumoto, T. Yamaguchi, M. Numao, H. Motoda (Eds.), Active Mining. XII, 349 pages. 2005.

Vol. 3419: B. Faltings, A. Petcu, F. Fages, F. Rossi (Eds.), Constraint Satisfaction and Constraint Logic Programming. X, 217 pages. 2005.

Vol. 3416: M. Böhlen, J. Gamper, W. Polasek, M.A. Wimmer (Eds.), E-Government: Towards Electronic Democracy. XIII, 311 pages. 2005.

Vol. 3415: P. Davidsson, B. Logan, K. Takadama (Eds.), Multi-Agent and Multi-Agent-Based Simulation. X, 265 pages. 2005.

Vol. 3403: B. Ganter, R. Godin (Eds.), Formal Concept Analysis. XI, 419 pages. 2005.

Vol. 3398: D.-K. Baik (Ed.), Systems Modeling and Simulation: Theory and Applications. XIV, 733 pages. 2005.

Vol. 3397: T.G. Kim (Ed.), Artificial Intelligence and Simulation. XV, 711 pages. 2005.

Vol. 3396: R.M. van Eijk, M.-P. Huget, F. Dignum (Eds.), Agent Communication. X, 261 pages. 2005.

Vol. 3394: D. Kudenko, D. Kazakov, E. Alonso (Eds.), Adaptive Agents and Multi-Agent Systems II. VIII, 313 pages. 2005.

Vol. 3392: D. Seipel, M. Hanus, U. Geske, O. Bartenstein (Eds.), Applications of Declarative Programming and Knowledge Management. X, 309 pages. 2005.

Vol. 3374: D. Weyns, H. V.D. Parunak, F. Michel (Eds.), Environments for Multi-Agent Systems. X, 279 pages. 2005.

Vol. 3371: M.W. Barley, N. Kasabov (Eds.), Intelligent Agents and Multi-Agent Systems. X, 329 pages. 2005.

Vol. 3369: V. R. Benjamins, P. Casanovas, J. Breuker, A. Gangemi (Eds.), Law and the Semantic Web. XII, 249 pages. 2005.

Vol. 3366: I. Rahwan, P. Moraitis, C. Reed (Eds.), Argumentation in Multi-Agent Systems. XII, 263 pages. 2005.

Vol. 3359: G. Grieser, Y. Tanaka (Eds.), Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets. XIV, 257 pages. 2005.

Vol. 3346: R.H. Bordini, M. Dastani, J. Dix, A.E.F. Seghrouchni (Eds.), Programming Multi-Agent Systems. XIV, 249 pages. 2005.

Vol. 3345: Y. Cai (Ed.), Ambient Intelligence for Scientific Discovery. XII, 311 pages. 2005.

Vol. 3343: C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, T. Barkowsky (Eds.), Spatial Cognition IV. XIII, 519 pages. 2005.

Vol. 3339: G.I. Webb, X. Yu (Eds.), AI 2004: Advances in Artificial Intelligence. XXII, 1272 pages. 2004.

Vol. 3336: D. Karagiannis, U. Reimer (Eds.), Practical Aspects of Knowledge Management. X, 523 pages. 2004.

Vol. 3327: Y. Shi, W. Xu, Z. Chen (Eds.), Data Mining and Knowledge Management. XIII, 263 pages. 2005.

Vol. 3315: C. Lemaître, C.A. Reyes, J.A. González (Eds.), Advances in Artificial Intelligence – IBERAMIA 2004. XX, 987 pages. 2004.

Vol. 3303: J.A. López, E. Benfenati, W. Dubitzky (Eds.), Knowledge Exploration in Life Science Informatics. X, 249 pages. 2004.

Vol. 3301: G. Kern-Isberner, W. Rödder, F. Kulmann (Eds.), Conditionals, Information, and Inference. XII, 219 pages. 2005.

Vol. 3276: D. Nardi, M. Riedmiller, C. Sammut, J. Santos-Victor (Eds.), RoboCup 2004: Robot Soccer World Cup VIII. XVIII, 678 pages. 2005.

Vol. 3275: P. Perner (Ed.), Advances in Data Mining. VIII, 173 pages. 2004.

Vol. 3265: R.E. Frederking, K.B. Taylor (Eds.), Machine Translation: From Real Users to Research. XI, 392 pages. 2004.

Vol. 3264: G. Paliouras, Y. Sakakibara (Eds.), Grammatical Inference: Algorithms and Applications. XI, 291 pages. 2004.

Vol. 3259: J. Dix, J. Leite (Eds.), Computational Logic in Multi-Agent Systems. XII, 251 pages. 2004.

Vol. 3257: E. Motta, N.R. Shadbolt, A. Stutt, N. Gibbins (Eds.), Engineering Knowledge in the Age of the Semantic Web. XVII, 517 pages. 2004.

Vol. 3249: B. Buchberger, J.A. Campbell (Eds.), Artificial Intelligence and Symbolic Computation. X, 285 pages. 2004.

Vol. 3248: K.-Y. Su, J. Tsujii, J.-H. Lee, O.Y. Kwong (Eds.), Natural Language Processing – IJCNLP 2004. XVIII, 817 pages. 2005.

Vol. 3245: E. Suzuki, S. Arikawa (Eds.), Discovery Science. XIV, 430 pages. 2004.

Vol. 3244: S. Ben-David, J. Case, A. Maruoka (Eds.), Algorithmic Learning Theory. XIV, 505 pages. 2004.

Vol. 3238: S. Biundo, T. Frühwirth, G. Palm (Eds.), KI 2004: Advances in Artificial Intelligence. XI, 467 pages. 2004.

Vol. 3230: J.L. Vicedo, P. Martínez-Barco, R. Muñoz, M. Saiz Noeda (Eds.), Advances in Natural Language Processing. XII, 488 pages. 2004.

Vol. 3229: J.J. Alferes, J. Leite (Eds.), Logics in Artificial Intelligence. XIV, 744 pages. 2004.

Vol. 3228: M.G. Hinchey, J.L. Rash, W.F. Truszkowski, C.A. Rouff (Eds.), Formal Approaches to Agent-Based Systems. VIII, 290 pages. 2004.

Vol. 3215: M.G.. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part III. LVII, 906 pages. 2004.

Vol. 3214: M.G.. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part II. LVIII, 1302 pages. 2004.

Vol. 3213: M.G.. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part I. LVIII, 1280 pages. 2004.

Vol. 3209: B. Berendt, A. Hotho, D. Mladenic, M. van Someren, M. Spiliopoulou, G. Stumme (Eds.), Web Mining: From Web to Semantic Web. IX, 201 pages. 2004.

¥424.80元

Table of Contents

The Dynamics of Knowledge

Methodology and Simulation

Agent Tools and Agent Education

Short Papers

Posters

On the Convergence of Structured Search, Information Retrieval and Trust Management in Distributed Systems*

Karl Aberer, Philippe Cudré-Mauroux, and Zoran Despotovic

School of Computer and Communication Sciences,
EPFL, Lausanne, Switzerland
{karl.aberer, philippe.cudre-mauroux, zoran.despotovic}@epfl.ch

Abstract. The database and information retrieval communities have long been recognized as being irreconcilable. Today, however, we witness a surprising convergence of the techniques used by both communities in decentralized, large-scale environments. The newly emerging field of reputation based trust management, borrowing techniques from both communities, best demonstrates this claim. We argue that incomplete knowledge and increasing autonomy of the participating entities are the driving forces behind this convergence, pushing the adoption of probabilistic techniques typically borrowed from an information retrieval context. We argue that using a common probabilistic framework would be an important step in furthering this convergence and enabling a common treatment and analysis of distributed complex systems. We will provide a first sketch of such a framework and illustrate it with examples from our previous work on information retrieval, structured search and trust assessment.

1 Introduction

The database and information retrieval communities have long been perceived as being irreconcilable. The different ways of how data is represented, interpreted and processed are at the core of the divergence in focus of these communities.

The main problem addressed by the database community can be stated as the efficient management of data represented in some first order logic language and the efficient evaluation of queries specifying information needs unambiguously through logical expressions. Recently this model has been extended in the context of the Semantic Web to deal with distributed, heterogeneous information sources by using shared first order conceptual models (ontologies) and a common Web-based infrastructure.

* The work presented in this paper was supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS) and by the Computational Reputation Mechanisms for Enabling Peer-to-Peer Commerce in Decentralized Networks Project, both supported by the Swiss National Science Foundation under grant number 5005-67322 and 20512-105287/1 respectively.

T. Eymann et al. (Eds.): MATES 2005, LNAI 3550, pp. 1–14, 2005.

On the other hand, the information retrieval community focuses on finding models for retrieving documents in response to incompletely or ambiguously specified information needs by exploiting document features and user relevance feedback. Web search engines are the most prominent incarnation of these techniques for assessing relevance of documents in response to user requests for information, using both textual content of documents and user feedback derived from the link structure of the Web.

Attempts to reconcile the two communities reach far back in history. Even a conference series, the International Conference on Information and Knowledge Management (CIKM), is dedicated to this goal. We were interested to see to which extent the interaction among the communities progressed, and analyzed the program of the years 2003 and 2004. The result is not too impressive. Among 120 research papers we could identify 10 that are at the borderline of databases and information retrieval, whereas the others are quite clearly belonging to the fields of classical database, information retrieval or knowledge management. In 2004, two sessions on databases and information retrieval have been organized. The topics addressed by the borderline papers are on storage management for retrieval systems, processing of XML documents and similarity search in databases. The last two areas in fact indicate one reason why the boundary between the database view of structured data processing and the information retrieval view of content-oriented processing is starting to dissolve. It is the result of processing specific data types that require both structural and content-oriented processing.

In this paper, we argue that recent developments in diverse areas, such as the Semantic Web, peer-to-peer computing, sensor networks, agent technologies and Web retrieval, indicate that the "semantic gap" between traditional logic-based knowledge presentation and processing and the probabilistic approach taken in information retrieval will be rapidly closing, for a very fundamental reason, that goes beyond the requirement of processing specific data types.

In a distributed environment of autonomous information sources, information and information needs can no longer be expressed concisely, as expected by database and semantic web technologies, but have to deal with *numerous sources of uncertainty*, thus requiring a probabilistic view in the processing of data. In information retrieval, one deals with one specific kind of uncertainty, uncertainty about users information needs. We claim that in distributed environments, qualitatively different sources of uncertainty have to be dealt with as well. This will require a structured framework to represent and process the different sources of uncertainty to provide insightful answers to users information needs. This requirement goes well beyond existing capabilities of both database and information retrieval techniques and systems.

We will illustrate this convergence process by providing several important examples of how the uncertainty resulting from autonomy and incomplete knowledge in distributed environments affects information processing. These examples are taken both from our own work and from some typical results found in the literature. We will provide short summaries of these techniques and illustrate by a simple example of a search problem how each of these techniques affects the information processing task for satisfying the search task. By doing this we illustrate how using a probabilistic framework makes it possible to integrate different ways of dealing with uncertainty, just as first order logic is being used as an integration framework for structured representation and reasoning over distributed information sources. This example-based analysis will allow us to derive some basic conclusions on requirements and issues for extending the current Web infrastructure for dealing with uncertainty in a systematic and integrated way.

2 Running Example: Getting Newspaper Articles About Hot Days in Switzerland

To illustrate our claims, we introduce an example which is in our opinion representative of the current challenges emerging in information management today. The example starts as a simple SQL query posed against a relational database but will be enriched throughout the paper as new sources of uncertainty are introduced.

From June to August 2003, unusually high temperatures were reported across Europe, including Switzerland. Imagine a journalist wanting to retrieve all newspaper articles about hot days in Switzerland which appeared exactly on one of those days. In a standard relational databases scenario, this could translate to a SQL query like the following:

```
SELECT article.text
FROM articles, weather WHERE
   article.text like %hot summer days%
   and article.date = weather.date
   and weather.temperature  > 30
```

The query contains three predicates, q_1, q_2 and q_3 representing some condition on the content of articles, their publication date and some temperature record respectively.

From a logical perspective, such a query can be considered as a logical expression q for which we have to find all objects d contained in a database such that the implication $d \rightarrow q$ is true.

Expressing an information need in this form reflects several basic assumptions being made, including the ability of the user to precisely express her information need, the correct interpretation of the schematic information provided by the database and the correctness of the data stored in the database. In practice, as we will demonstrate in the following, none of these assumptions can be taken for granted in realistic, distributed information systems.

3 Uncertainty on Users' Information Needs

Since long it has been recognized that logics is not an appropriate framework for information search when it comes to searching documents with textual content. Boolean retrieval has been an early attempt to apply logics for text search, which has soon found its limitations. Due to the ambiguity of natural language, there exists no strict relationship between queries expressed in natural language against documents containing natural language text. Thus the discipline of information retrieval has developed a rich set of models for assessing the relevance of documents for a given query. These models introduce an element of *uncertainty* into the search process, since result objects are no more included into the result set by virtue of a decidable property (a predicate) but whenever there is indication that they might be relevant to some degree to the users information need. These observations clearly apply to the clause q_1 `== article.text like %hot summer days%` of our example query, which in a current database system (ideally) would not be resolved at the syntactic level searching for the exact phrase, but using an underlying text retrieval system.

3.1 Running Example: Accounting for the Uncertainty on Information Needs Through Probabilistic Retrieval

Since we are aiming at a probabilistic framework for dealing with uncertainty in modern information systems, we provide here a short overview of information retrieval from a probabilistic perspective, which follows the exposition given by [5]. From a logical perspective, answering a query q with document d amounts to proving that the implication $d \rightarrow q$ is true. In Boolean retrieval this means that all terms of a (conjunctive) query q would appear in d. In contrast, probabilistic retrieval adopts the following notion for answering a query q: the conditional probability $P(q|d)$ indicates of how relevant document d is to query q.

For computing this probability usually a concept space C of disjoint concepts $c \in C$ is introduced with a probability density function $P(.)$ over C. Queries and documents are considered as concept sets. Then the query answer can be represented as follows:

$$P(q|d) = \frac{P(q \cap d)}{P(d)}, \; P(d) = \sum_{c \in d} P(c), \; P(q \cap d) = \sum_{c \in q, c \in d} P(c)$$

A popular type of concepts are terms taken from a vocabulary. Since the concepts are considered as being independent we can further derive

$$P(q|d) = \frac{P(q \cap d)}{P(d)} = \frac{\sum_{c \in C} P(d \cap q \cap c)}{P(d)} = \frac{\sum_{c \in C} P(d \cap q|c)P(c)}{P(d)}$$

If the concept space consists of the terms of a vocabulary, we may assume that the probabilities $P(d|c)$ and $P(q|c)$ are known from analyzing the text collection. For computing a query answer, a standard assumption that is made in probabilistic retrieval is the *maximum entropy principle*, which states the following independence:

$$P(d \cap q|c) = P(d|c)P(q|c).$$

Using this assumption we get

$$P(q|d) = \frac{\sum_{c\in C} P(d \cap q|c)P(c)}{P(d)}$$
$$= \frac{\sum_{c\in C} P(d|c)P(q|c)P(c)}{P(d)}$$
$$= \sum_{c\in C} P(q|c)P(c|d)$$

The last expression can be interpreted as the classical model of vector space retrieval, the predominant model for modern text retrieval. Under this interpretation, $P(c|d)$ corresponds to the term weight for a document representation, which is typically computed using a (heuristic) tf-idf scheme and gives the probability that a term is characteristic for a given document. $P(q|c)$ corresponds to the query term weight and gives the probability that a term is characteristic for the result set of query q.

In summary, a predicate such as q_1 `== article.text like %hot summer days%` corresponds in a search model that is considering uncertainty on users' information needs to a random variable q_1 for which we have a method to compute $P(q_1|d)$, the probability that a document is relevant to the predicate. The method to compute this probability relies on an intermediary concept (or feature) space C, for which we assume to have probabilistic models for $P(q_1|c)$ and $P(c|d)$ for a random variable c over the concept space. The computation of $P(q_1|d)$ is then performed by marginalization of the joint probability distribution $P(q_1, c, d)$ exploiting the separation of the random variables q_1 and d through c.

From a practical perspective, using a retrieval engine within a logics-based query language such as SQL poses the question of how to reflect the probabilistic evaluation of $q1$ into the query result. Two solutions are applicable: either only result documents are included that exceed a certain threshold probability. This seems to be problematic with respect to the interpretation of the result. Alternatively the probability values are included into the result table. This might raise efficiency concerns as the result set might become unacceptably large. As we will show in the following, this is a problem that is not confined to the case of dealing with users' uncertainty on information need, but with dealing with uncertainty in general.

4 Uncertainty on Knowledge Conceptualizations

Traditionally, knowledge representations have been based on subsets of first-order logic in computer science. Indeed, it is widely recognized that knowledge can be efficiently captured by characterizing classes of objects and their inter-relationships. Databases have long used dialects derived from first-order logic to represent or query data, while description logic, a subset of first-order logic, has been chosen to back-up standards for the Semantic Web.