

DATA 2012

Final Program and

Book of Abstracts

International Conference on
Data Technologies and Applications

Rome, Italy
25 – 27 July, 2012

Sponsored by

INSTICC – Institute for Systems and Technologies of Information, Control
and Communication

Table of Contents

Table of Contents	3
Foreword.....	5
Organizing and Steering Committees.....	6
Program Committee.....	7
Auxiliary Reviewers	8
Panel.....	9
Keynote Lectures.....	10
Issues in Combined Static and Dynamic Data Management.....	10
On Knowledge Discovery and Interactive Intelligent Visualization of Biomedical Data - Challenges in Human–Computer Interaction & Biomedical Informatics	11
Integrated Marine Data Management - Data Management Strategies for Risk Reduction in the Offshore Wind Industry	12
Visualization & Data Mining for High Dimensional Datasets.....	13
Awards	14
Social Event and Banquet	15
General Information	17
Rooms Layout.....	18
Program Layout	19
Final Program and Book of Abstracts.....	21
Contents.....	23
Wednesday Sessions	27
Thursday Sessions	31
Friday Sessions	39
Notes.....	46

Foreword

This book contains the abstracts of the First International Conference on Data Technologies and Applications - DATA 2012. The conference is sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC).

The purpose of DATA is to bring together researchers and practitioners interested in databases, data warehousing, data mining, data management, data security and other aspects of information systems and technology involving advanced applications of data.

DATA 2012 received 66 paper submissions from 32 countries. To evaluate each submission, a double-blind paper evaluation method was used: each paper was reviewed by at least two internationally known experts from the DATA Program Committee. Only 7 papers were selected to be published and presented as full papers, i.e. completed work (10 pages in proceedings / 30' oral presentation). Additionally, 19 papers were accepted as short papers (6 pages / 20' oral presentation) - for a total of 26 oral presentations - and 6 papers as posters. The full-paper acceptance ratio was thus 10,6%, while the total oral paper acceptance ratio was 39,4%.

DATA's program includes a panel to discuss aspects of data technologies and applications from both theoretical and practical perspectives, with the participation of distinguished world-class researchers and practitioners; furthermore, the program is enriched by several keynote lectures delivered by renowned experts in their areas of knowledge. These high points in the conference program definitely contribute to reinforce the overall quality of the DATA conference.

The program for this conference required the dedicated effort of many people. Firstly, we must thank the authors, whose research efforts are herewith recorded. Secondly, we thank the members of the Program Committee and the additional reviewers for their diligent and professional reviewing. Last but not least, we thank the invited speakers for their invaluable contribution and for taking the time to prepare their talks.

A successful conference involves more than paper presentations; it is also a meeting place, where ideas about new research projects and other ventures are discussed and debated. Therefore, a social event - including dinner - has been arranged for the evening of July 26 (Thursday) in order to promote this kind of social networking.

We wish you all an exciting conference and an unforgettable stay in the city of Rome. We hope to meet you again next year for the 2nd DATA, details of which will shortly be made available at <http://www.dataconference.org>.

Markus Helfert, Dublin City University, Ireland

Chiara Francalanci, Politecnico di Milano, Italy

Joaquim Filipe, Polytechnic Institute of Setúbal / INSTICC, Portugal

Organizing and Steering Committees

Conference Chair

Joaquim Filipe, Polytechnic Institute of Setúbal / INSTICC, Portugal

Program Co-chairs

Markus Helfert, Dublin City University, Ireland

Chiara Francalanci, Politecnico di Milano, Italy

Proceedings Production

Helder Coelhas, INSTICC, Portugal

Andreia Costa, INSTICC, Portugal

Patrícia Duarte, INSTICC, Portugal

Bruno Encarnação, INSTICC, Portugal

Liliana Medina, INSTICC, Portugal

Raquel Pedrosa, INSTICC, Portugal

Vitor Pedrosa, INSTICC, Portugal

CD-ROM Production

Pedro Varela, INSTICC, Portugal

Graphics Production and Webdesigner

Daniel Pereira, INSTICC, Portugal

Secretariat

Sérgio Brissos, INSTICC, Portugal

Webmaster

Susana Ribeiro, INSTICC, Portugal

Program Committee

Muhammad Abulaish, King Saud University, Saudi Arabia
Hamideh Afsarmanesh, University of Amsterdam, The Netherlands
Markus Aleksy, ABB Corporate Research Center, Germany
Rafa E. Al-Qutaish, Al Ain University of Science and Technology, U.A.E.
Kenneth Anderson, University of Colorado, U.S.A.
Farhad Arbab, CWI, The Netherlands
Sören Auer, University of Leipzig, Germany
Mortaza S. Bargh, The Netherlands
Bernhard Bauer, University of Augsburg, Germany
Orlando Belo, University of Minho, Portugal
Jorge Bernardino, Institute Polytechnic of Coimbra - ISEC, Portugal
Marko Boškovic, Athabasca University, Canada
Omar Boussaid, Eric Laboratory, University of Lyon 2, France
Mark Van Den Brand, Eindhoven University of Technology, The Netherlands
Cinzia Cappiello, Politecnico di Milano, Italy
Silvana Castano, University of Milan, Italy
Cagatay Catal, Istanbul Kultur University, Turkey
Krzysztof Cetnarowicz, AGH - University of Science and Technology, Poland
Kung Chen, National Chengchi University, Taiwan
Chia-Chu Chiang, University of Arkansas at Little Rock, U.S.A.
Stefan Conrad, Heinrich-Heine University Duesseldorf, Germany
Félix Cuadrado, Queen Mary University of London, U.K.
Decheng Dai, Google, China
María J. Domínguez-Alda, Universidad de Alcalá, Spain
Petros Drineas, Rensselaer Polytechnic Institute, U.S.A.
Artur Dubrawski, The Robotics Institute Carnegie Mellon University, U.S.A.
Juan C. Dueñas, Universidad Politécnica de Madrid, Spain
Sameh Elnikety, Microsoft Research, U.S.A.
Mohamed Y. Eltabakh, Worcester Polytechnic Institute, U.S.A.
Fikret Ercal, Missouri University of Science & Technology, U.S.A.
Chiara Francalanci, Politecnico di Milano, Italy
Rita Francese, Università degli Studi di Salerno, Italy
Faiez Gargouri, Miracl Laboratory, Tunisia
Nikolaos Georgantas, INRIA, France
Paola Giannini, Università del Piemonte Orientale, Italy
J. Paul Gibson, TSP - Telecom SudParis, France
Matteo Golfarelli, University of Bologna, Italy
Cesar Gonzalez-Perez, Institute of Heritage Sciences (Incipit), Spanish National Research Council (CSIC), Spain
Mohand-Said Hacid, Université Claude Bernard Lyon 1, France
Slimane Hammoudi, ESEO, France
Markus Helfert, Dublin City University, Ireland
Jose Luis Arciniegas Herrera, Universidad del Cauca, Colombia
Jose R. Hilera, University of Alcalá, Spain
Jang-eui Hong, Chungbuk National University, Korea, Republic of
Ivan Ivanov, SUNY Empire State College, U.S.A.
Sanpawat Kantabutra, Chiang Mai University, Thailand
Dimitris Karagiannis, University of Vienna, Austria
Panagiotis Karras, Rutgers University, U.S.A.
Maurice van Keulen, University of Twente, The Netherlands
Roger (Buzz) King, University of Colorado, U.S.A.
Jeffrey W. Koch, Tarrant County College Northeast Campus, U.S.A.
Mieczyslaw Kokar, Northeastern University, U.S.A.
Konstantin Läufer, Loyola University Chicago, U.S.A.
Raimondas Lencevicius, Nuance Communications, U.S.A.
Ziyu Lin, Xiamen University, China
Ricardo J. Machado, Universidade do Minho, Portugal

Leszek Maciaszek, Wroclaw University of Economics, Poland / Macquarie University - Sydney, Australia
Antonia Mas, Universitat de les Illes Balears, Spain
Marian Cristian Mihaescu, University of Craiova, Romania
Dimitris Mitrakos, Aristotle University of Thessaloniki, Greece
Valérie Monfort, SOIE Tunis, Tunisia
Mirella M. Moro, Federal University of Minas Gerais (UFMG), Brazil
Paolo Nesi, University of Florence, Italy
Erich Neuhold, Universität Wien, Austria
Paulo Novais, Universidade do Minho, Portugal
Rory O'Connor, Dublin City University, Ireland
Pasi Ojala, Nokia, Finland
Vincenzo Pallotta, University of Business and International Studies - Geneva, Switzerland
George Papastefanatos, RC "Athena", Greece
Rosario Pugliese, Università di Firenze, Italy
Christoph Quix, RWTH Aachen University, Germany
S. Ram, University of Arizona, U.S.A.
Werner Retschitzegger, Johannes Kepler University, Austria
Claudio de la Riva, University of Oviedo, Spain
Colette Rolland, Université Paris 1 Panthéon-Sorbonne, France
Gustavo Rossi, Lafia, Argentina
Gunter Saake, Institute of Technical and Business Information Systems, Germany
Krzysztof Sacha, Warsaw University of Technology, Poland
Manuel Filipe Santos, University of Minho, Portugal
Jean-Marc Seigneur, University of Geneva, Switzerland
Damian Serrano, INRIA, France
Martín Serrano, National University of Ireland Galway - NUIG, Ireland
Lijun Shan, National Digital Switching System Engineering and Technological Research Center, China
Jie Shao, University of Queensland, Australia
Harvey Siy, University of Nebraska at Omaha, U.S.A.
Yeong-tae Song, Towson University, U.S.A.
Cosmin Stoica Spahiu, University of Craiova - Faculty of Automation, Computers and Electronics, Romania
Gianluigi Viscusi, Università Di Milano-bicocca, Italy
Martijn Warnier, Delft University of Technology, The Netherlands
Dietmar Wikarski, FH Brandenburg University of Applied Sciences, Germany
Leandro Krug Wives, Universidade Federal do Rio Grande do Sul, Brazil
Alexander Woehrer, St. Poelten University of Applied Sciences, Austria
Jongwook Woo, California State University, U.S.A.
Yun Xiong, Fudan University, China
Hui Yang, San Francisco State University, U.S.A.
I-Ling Yen, University of Texas at Dallas, U.S.A.
Qing Zhang, Australia e-Health Research Centre / CSIRO ICT Centre, Australia
Xiaokun Zhang, Athabasca University, Canada
Jiakui Zhao, China Electric Power Research Institute, China
Xiangmin Zhou, CSIRO, Australia
Yangyong Zhu, Fudan University, China

Auxiliary Reviewers

Nikos Bikakis, National Technical University of Athens, Greece
Ioannis Konstantinou, National Technical University of Athens, Greece
Alexander Serebrenik, Technische Universiteit Eindhoven, The Netherlands

Panel

Wednesday, 25

15:15 – 16:30

Room: Plenary

Title: *(To be defined)*

Panel Chair:

Markus Helfert, Dublin City University, Ireland

Keynote Lectures

Wednesday 25
16:45 – 17:45
Room: Plenary

Issues in Combined Static and Dynamic Data Management

Daniela Nicklas
Carl von Ossietzky Universitaet Oldenburg
Germany

With the upcoming widespread availability of sensors, more and more applications depend on physical phenomena. Up-to-date real world information is embedded into business processes, in production environments, or in mobile applications, so that such context-aware applications can adapt their behavior to the current situation of their user or environment. Another example are so-called SCADA systems (supervisory control and data acquisition), where complex installations (e.g., energy grids or power plants) are monitored and controlled. In general, data can be managed either by a database management system (DBMS), or directly by the application. The first approach has many advantages: information demands can be declared by queries and kept separately from the application. When the information demand changes, only the query has to be changed (which dramatically decreases software maintenance costs). In addition, a DBMS can optimize the query execution, so that the requested data is retrieved efficiently from the systems. However, if applications depend on real-world data, the amount of data, the update rate, and low latency requirements often prevent the storage in a DBMS. Thus, the data streams from the sensors are managed directly by applications, with all its drawbacks. The goal of data stream management is to provide the same flexibility and data independence for data streams as for stored data. However, despite of performance advantages, DSMS represent no general solution since applications often require the persistent storage of continuous query results as well as a combined processing of current and historical stream data. The performance of integrated database methods, on the other hand, is limited due to the expensive data management of traditional DBMS. A naïve solution would be to federate or integrate both types of systems. However, a closer look shows that such a federation raises many open questions.

Daniela Nicklas is a Junior Professor for Database and Internet technologies at the Carl von Ossietzky Universität Oldenburg since 2008. Besides her research on context-aware applications and context modeling, she works on data stream processing and sensor-based systems within the application domains of transportation, energy, and ambient environments. She received her PhD (Dr. rer. nat.) in Computer Science 2008 from the University of Stuttgart, working within the Center of Collaboration (SFB 627) "Nexus" under the supervision of Prof. Dr. Bernhard Mitschang. The main contribution of her thesis was architecture and a spatial data model for large-scale integration of context data for locationbased, mobile applications. After that, she worked for two years as PostDoc in that Center of Project, co-leading a project on context-aware workflows together with Prof. Dr. Frank Leymann. In 2008, she received an IBM Exploratory Stream Analytics Innovation award for "Data Stream Technology for Future Energy Grid Control". Besides being a committee member in many programme committees of pervasive computing and database conferences and workshops, she helped organizing international conference, like e.g., the Annual IEEE International Conferences on Pervasive Computing and Communications (PerCom) as Vice Program Chair in 2010. She is also a member of the editorial boards of the Elsevier Pervasive Computing and Communication Journal, the International Journal of Pervasive Computing and Communications (Emerald), and the Datenbankspektrum (German Journal on Databases).

Thursday 26
12:15 – 13:15
Room: Plenary

On Knowledge Discovery and Interactive Intelligent Visualization of Biomedical Data - Challenges in Human–Computer Interaction & Biomedical Informatics

Andreas Holzinger
Medical University Graz
Austria

Biomedical Informatics can be defined as “the interdisciplinary field that studies and pursues the effective use of biomedical data, information and knowledge for scientific inquiry, problem solving, and decision Making, motivated by efforts to improve human health.” However, professionals in the life sciences are facing an increasing quantity of highly complex, multi-dimensional and weakly structured data. While researchers in Human-Computer Interaction (HCI) and Knowledge Discovery in Databases (KDD) have for long been working independently to develop methods that can support expert end users to identify, extract and understand information out of this data, it is obvious that an interdisciplinary approach to bring these two fields closer together can yield synergies in the application of these methods to weakly structured complex medical data sets. The aim is to support end users to learn how to interactively analyse information properties and to visualize the most relevant parts – in order to gain knowledge, and finally wisdom, to support a smarter decision making. The danger is not only to get overwhelmed by increasing masses of data, moreover, there is the risk of modelling artifacts.

Andreas Holzinger is head of the Research Unit Human–Computer Interaction, Institute of Medical Informatics, Statistics and Documentation, Medical University Graz, Associate Professor of Applied Informatics at the Faculty of Computer Science, Institute of Information Systems and Computer Media and Lecturer at the Faculty of Electrical and Information Engineering, Institute of Genomics and Bioinformatics at Graz University of Technology. He is chair of the Workgroup Human–Computer Interaction and Usability Engineering (HCI&UE) of the Austrian Computer Society, and founder and leader of the Special Interest Group HCI4MED. Since November 2009 he is Austrian Representative in the International Federation of Information Processing (IFIP), Technical Committee TC 13 Human-Computer Interaction. He serves as consultant for the Austrian, German and Dutch Government and for the German Excellence Initiative and as national expert in the European Commission (Lisbon Delegate 2000). Andreas, born 1963, started as an apprentice in Information Technology in 1978; while working as an industrial engineer, he resumed a parallel second-chance education, finished his PhD in cognitive science in 1997 and completed his second doctorate (habilitation) in applied informatics in 2003. Since 1999 participation in leading positions in 29 R&D multi-national projects, budget 2,8 MEUR; to date 314 publications, 2029 citations; h-index = 23, g-Index = 41; Andreas was Visiting Professor in Berlin, Innsbruck, London, Vienna and Aachen. His research field is in Computing and Information Sciences with application in Life and Medical Sciences and emphasis on Knowledge Management, Multimedia Information Systems, Human-Computer Interaction, Knowledge Discovery/Information Retrieval and Usability Engineering. Homepage: <http://www.hci4all.at>; Current main lecture: http://genome.tugraz.at/medical_informatics.shtml.

Friday 27

09:30 – 10:30

Room: Plenary

Integrated Marine Data Management - Data Management Strategies for Risk Reduction in the Offshore Wind Industry

John Shaw
*Mainstream Renewable Power
Ireland*

There is a compelling need for an Integrated Sea Information System, ISIS, to accelerate the €6.4 Trillion European Offshore Wind Industry.

ISIS will reduce the risk of developing, constructing and operating Offshore Wind to create ISIS we need new EU policies, Regulations, Standards, greater collaboration and Innovative Technologies.

This presentation will explain Mainstream's Vision, Offshore Wind Developers' data needs, Mainstream's Data Management Strategy and provide details on the ISIS Vision, Plan and Consortium.

John Shaw is CIO and Head of Information Services, based at Mainstream's Global Headquarters in Dublin, Ireland. John joined Mainstream in September 2008 to lead the company's Business Technology Strategy. A Chartered Engineer (C Eng MIEI) and a registered European Professional Engineer (Eur Ing) with 20+ years experience. John holds an MBA in Technology Management and a Bachelor's Degree in Electronics Engineer from University College Dublin. A senior member of the IEEE (SMIEEE), joined Mainstream from Pfizer where he was Director of Information Services at one of Pfizer's largest manufacturing facilities, having previously held senior IS executive positions in Accenture and General Electric. John is an active member of technical institutes, including The International Society of Automation, The Institute of Marine Engineering, Science and Technology and TheInnovation Value Institute which researches and develops unifying frameworks to create more value from Business Technology innovation.

Friday 27

16:30 – 17:30

Room: Plenary

Visualization & Data Mining for High Dimensional Datasets

Alfred Inselberg
Tel Aviv University
Israel

A dataset with M items has 2^M subsets any one of which may be the one which satisfies our objectives. With a good data display and interactivity our fantastic pattern-recognition can cut great swaths searching through this combinatorial explosion and also extract insights from the visual patterns. These are the core reasons for data visualization. With parallel coordinates the search for relations in multivariate datasets is transformed into a 2-D pattern recognition problem. The foundations are developed interlaced with applications. Guidelines and strategies for knowledge discovery are illustrated on several real datasets (financial, process control, credit-score, intrusion-detection etc) one with hundreds of variables. A geometric classification algorithm is presented and applied to complex datasets. It has low computational complexity providing the classification rule explicitly and visually. The minimal set of variables required to state the rule (features) is found and ordered by their predictive value. Multivariate relations can be modeled as hypersurfaces and used for decision support. A model of a (real) country's economy reveals sensitivities, impact of constraints, trade-offs and economic sectors unknowingly competing for the same resources. An overview of the methodology provides foundational understanding; learning the patterns corresponding to various multivariate relations. These patterns are robust in the presence of errors and that is good news for the applications. We stand at the threshold of breaching the gridlock of multidimensional visualization. The parallel coordinates methodology has been applied to collision avoidance and conflict resolution algorithms for air traffic control (3 USA patents), computer vision (1 USA patent), data mining (1 USA patent), optimization, decision support and elsewhere.

Alfred Inselberg received a Ph.D. in Mathematics and Physics from the University of Illinois (Champaign- Urbana) then was Research Professor there until 1966. He held research positions at IBM, where he developed a Mathematical Model of Ear (TIME Nov. 74), concurrently having joint appointments at UCLA, USC and later at the Technion and Ben Gurion University. Since 1995 he is Professor at the School of Mathematical Sciences at Tel Aviv University. He was elected Senior Fellow at the San Diego Supercomputing Center in 1996, Distinguished Visiting Professor at Korea University in 2008 and Distinguished Visiting Professor at National University of Singapore in 2011. Alfred invented and developed the multidimensional system of Parallel Coordinates for which he received numerous awards and patents (on Air Traffic Control, Collision-Avoidance, Computer Vision, Data Mining). The textbook "Parallel Coordinates: VISUAL Multidimensional Geometry and its Applications", Springer (October) 2009, has a full chapter on Data Mining and was acclaimed, among others, by Stephen Hawking.

Awards

Best Paper Awards

A "Best Paper Award" and a "Best Student Paper Award" will be conferred to the author(s) of a full paper presented at the conference, selected by the Program/Conference Chairs based on the best combined marks of paper reviewing, assessed by the Program Committee, and paper presentation quality, assessed by session chairs at the conference venue.

The "Best Student Paper Award" will be given to a paper in which the first author is a registered MSc or PhD student.

The awards will be announced and bestowed at the conference closing session.

Social Event and Banquet

Venue: *Bus Tour in Rome followed by a Dinner at the “Il Borgo di Tragliata”*

Date: *Thursday 26, 18:15 - 23:30*



Located at the entrance to Rome, “IL borgo di Tragliata,” rises above an impressive tufa buttress. Archeological sources provide evidence that this area has been inhabited since ancient times. The discovery of the famous, “Oinochoe of Tagliatella” vase confirms the existence of human settlements since the Etruscan era within an area subject to control by either Ceri or Veio. The place name “Tragliata” takes note of the place names, Talianum Tagliata or Terlata, during medieval times and appears to be derived from “Tagliata, (meaning, “cut”), which is the word given to the paths dug into the tufa by the Etruscans.



The presence of several tombs dug into the tufa along the East slope of the hill on which the village sits, along with several clay artifacts found in the area, are evidence which suggest the presence of a small agricultural settlement. In addition, other documentation reports the findings of the remains of a Roman villa on Tragliata property. It is also known that the two marble memorial stones found in this area have inscriptions dating back to the third century AD.

Midway through the eighth century, this area of the Roman countryside saw a period of repopulation thanks to the intelligence and will of Pope San Zaccaria (741-752) and Pope Adriano I (772-793). Encouraged by political and religious motives, these two Pontiffs presented an energetic revival and control of the territory.



During ninth and tenth centuries the historical scene began to change, the Roman countryside, with less support for the Papacy by the Carolingian empire, was made subject to continual and bloody raids by the Saracen pirates. The system of the “Domuscultae” entered into definite crisis, superceded by a strong defense system of towers and small castles; several coastal light towers were constructed to be used as bright defense signals to alert the inland region upon the pirates’ approach.

The construction of Tragliata’s small castle and tower date back to the ninth century, according to sources at the nearby Boccea castle.

The estate still belongs to the Vatican Basilica, even if time after time it was more or less controlled directly by others. In 1201, for example, it was ruled by a certain Jacobus de Traliata who occupied it, possibly as a lord. Several years later Tragliata, together with nearby “castium” Loterni, became subject to the interests of the turbulent Normanni family.



In 1885, the Chapter granted the Tragliata estate to Mr. Nicola Santovetti as the perpetual leaseholder. Consequently, Santovetti sold the lease to Mr. Domenico Lanza in 1917, (the great grand father of the present proprietor, Andrea de Gallo di Roccagiovane) who then took over as a tenant to finally gain possession of the estate in the following years.



General Information

Welcome Desk/On-site Registration

Tuesday 24 – Open from 15:00 to 17:30

Wednesday 25 – Open from 14:30 to 17:45

Thursday 26 – Open from 8:30 to 16:30

Friday 27 – Open from 9:00 to 17:45

Opening Session

Wednesday 25, at 15:00 in the Plenary room.

Welcome Cocktail

Wednesday 25, at 17:45 in the Picasso Foyer.

Closing Session

Friday 27, at 17:30 in the Plenary room.

Farewell Cocktail

Friday 27, at 17:45 in the Picasso Foyer.

Meals

Coffee-breaks will be served in the Picasso Foyer next to the conference rooms to all registered participants. Lunches will be served in the Murillo room. Please check the hours in the Program Layout.

Communications

Wireless access will be provided free of charge to all registered participants, during the conference business hours.

Secretariat Contacts

DATA Secretariat

Address: Av. D.Manuel I, 27A 2ºesq.

2910-595 Setúbal - Portugal

Tel.: +351 265 520 185

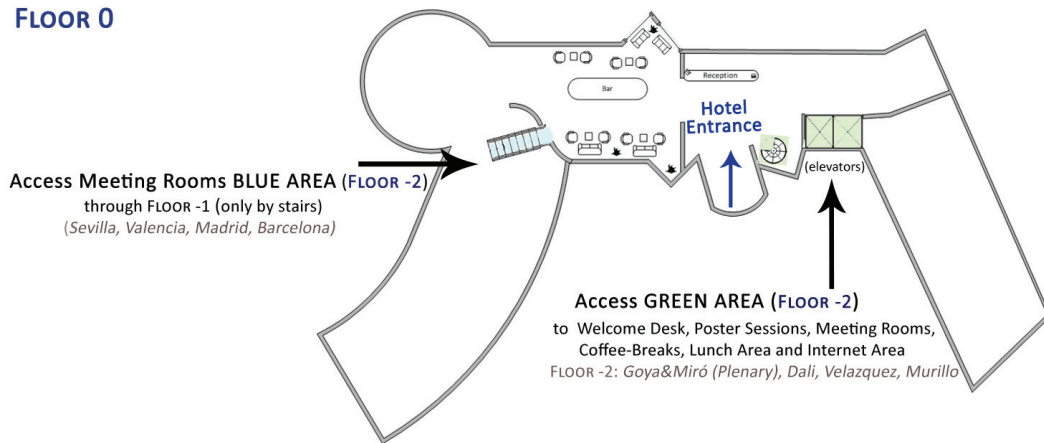
Fax: +44 203 014 5433

e-mail: data.secretariat@insticc.org

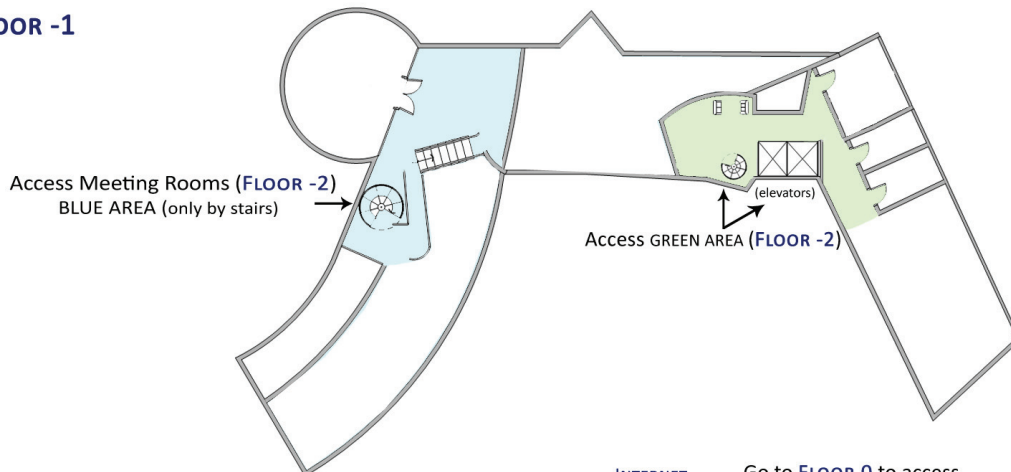
website: <http://www.dataconference.org>

Rooms Layout

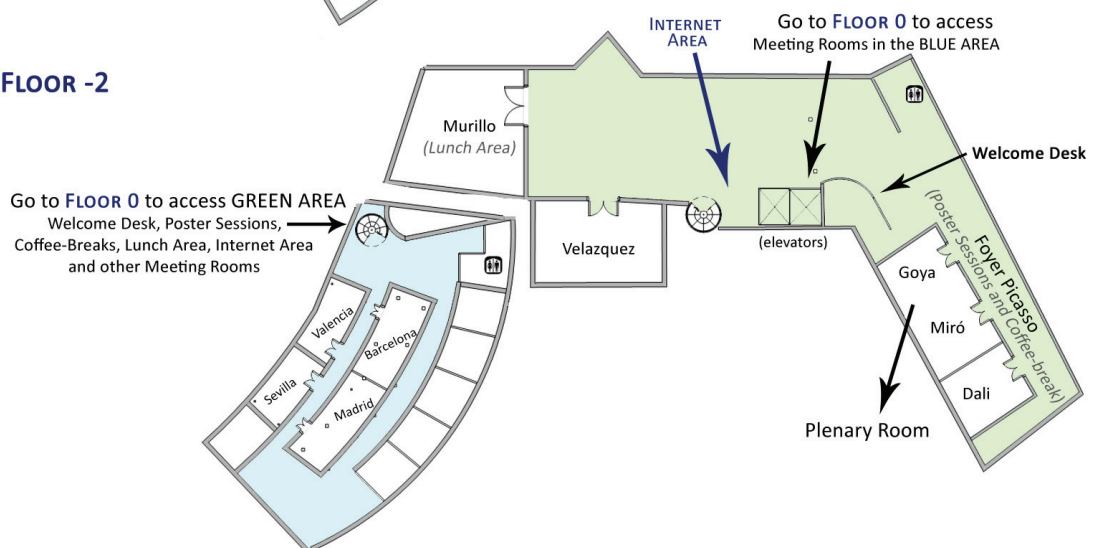
FLOOR 0



FLOOR -1



FLOOR -2



Program Layout

Tuesday 24		Wednesday 25		Thursday 26		Friday 27		
				08:30	Welcome Desk / Registration			
				09:00	Session 1	09:00	Welcome Desk / Registration	
						09:30	Keynote Lecture John Shaw	
				10:30	Coffee-Break	10:30	C.-Break	
				10:45	Session 2	10:45	Posters Session	
						11:30	Session 4	
				12:15	Keynote Lecture Andreas Holzinger			
				13:15	Lunch	13:00	Lunch	
		14:30	Welcome Desk / Registration	14:30	Session 3	14:15	Session 5	
15:00	Welcome Desk / Registration	15:00	Opening Session					16:15
		15:15	Panel					
		16:30	Coffee-Break		16:30		16:30	Keynote Lecture Alfred Inselberg
		16:45	Keynote Lecture Daniela Nicklas				17:30	Closing Session
17:30		17:45	Welcome Cocktail			17:45	Farewell Cocktail	
		18:15		18:15	Buses to Banquet	18:15		
					Social Event and Banquet			
				23:30	Buses back to hotel			

Final Program and Book of Abstracts

Contents

Wednesday Sessions	25
Opening Session (15:00 - 15:15)	27
Room Plenary	27
Panel: DATA (15:15 - 16:30)	27
Room Plenary	27
Keynote Speaker: Daniela Nicklas (16:45 - 17:45)	27
Room Plenary	27
Issues in Combined Static and Dynamic Data Management	27
Thursday Sessions	29
Session 1 (09:00 - 10:30)	31
Room Velazquez - Data Warehousing and Business Intelligence	31
23: Flexible Information Management, Exploration and Analysis in SAP HANA	31
34: Applying Personal and Group-based Trust Models in Document Recommendation	31
Session 2 (10:45 - 12:15)	31
Room Velazquez - Data Warehousing and Business Intelligence	31
17: Cloud based Privacy Preserving Data Mining with Decision Tree	31
33: Inference in Hierarchical Multidimensional Space	32
Keynote Speaker: Andreas Holzinger (12:15 - 13:15)	32
Room Plenary	32
On Knowledge Discovery and Interactive Intelligent Visualization of Biomedical Data - Challenges in Human-Computer Interaction & Biomedical Informatics	32
Parallel Session 3 (14:30 - 16:30)	32
Room Madrid - Ontologies and the Semantic Web	32
19: An Architecture based on Ontologies, Agents and Metaheuristics Applied to the Multimedia Service of the Brazilian Digital Television System	32
22: Effective and Efficient Online Communication - The Channel Model	32
31: Effective Keyword Search via RDF Annotations	33
54: Toward a Product Search Engine based on User Reviews	33
58: Smart Learning Management System Framework	33
Parallel Session 3 (14:30 - 16:30)	34
Room Plenary - Data Management and Quality	34
16: On Continuous Top-k Similarity Joins	34
35: Data Quality Sensitivity Analysis on Aggregate Indicators	34
37: Combining Local and Related Context for Word Sense Disambiguation on Specific Domains	34
70: DBMS meets DSMS - Towards a Federated Solution	35
Friday Sessions	37
Keynote Speaker: John Shaw (09:30 - 10:30)	39
Room Plenary	39
Integrated Marine Data Management - Data Management Strategies for Risk Reduction in the Offshore Wind Industry	39

Poster Session (10:30 - 11:30)	39
Foyer	39
30: Ontology Similarity Measurement Method in Rapid Data Integration	39
43: Ontology-engineered MSME Framework	39
45: Geometric Divide and Conquer Classification for High-dimensional Data	39
47: Graph-based Campaign Amplification in Telecom Cloud	40
51: Create a Specialized Search Engine - The Case of an RSS Search Engine	40
71: Understanding Worldwide Human Information Needs - Revealing Cultural Influences in HCI by Analyzing Big Data in Interactions	40
Session 4 (11:30 - 13:00)	41
Room Dali - Trust, Privacy and Security	41
52: NexusDSS: A System for Security Compliant Processing of Data Streams	41
42: Towards Process Centered Information Security Management - A Common View for Federated Business Processes and Personal Data Usage Processes	41
Parallel Session 5 (14:15 - 16:15)	41
Room Dali - Data Management and Quality	41
5: NRank: A Unified Platform Independent Approach for Top-K Algorithms	41
15: FIND - A Data Cloud Platform for Financial Data Services	42
18: How Do I Manage My Personal Data? – A Telco Perspective	42
20: KIDS - A Model for Developing Evolutionary Database Applications	42
50: Towards a Meaningful Analysis of Big Data - Enhancing Data Mining Techniques through a Collaborative Decision Making Environment	43
65: A Data-Centric Approach for Networking Applications	43
Parallel Session 5 (14:15 - 16:15)	43
Room Velazquez - Data Warehousing and Business Intelligence	43
38: A Virtual Document Approach for Keyword Search in Databases	43
25: Constrained Nonnegative Matrix Factorization based Data Distortion Techniques - Study of Data Privacy and Utility	44
27: Kernel Generations for a Diagnosis Model with GP	44
29: Automatic Subspace Clustering with Density Function	44
Keynote Speaker: Alfred Inselberg (16:30 - 17:30)	45
Room Plenary	45
Visualization & Data Mining for High Dimensional Datasets	45
Closing Session (17:30 - 17:45)	45
Room Plenary	45

Wednesday Sessions

Opening Session 15:00 - 15:15	Room Plenary
----------------------------------	--------------

both types of systems. However, a closer look shows that such a federation raises many open questions.

Panel 15:15 - 16:30	DATA Room Plenary
------------------------	----------------------

16:45 - 17:45	Room Plenary
Issues in Combined Static and Dynamic Data Management	
Keynote Speaker: Daniela Nicklas	

Issues in Combined Static and Dynamic Data Management

Daniela Nicklas

Carl von Ossietzky Universitaet Oldenburg, Oldenburg, Germany

Abstract: With the upcoming widespread availability of sensors, more and more applications depend on physical phenomena. Up-to-date real world information is embedded into business processes, in production environments, or in mobile applications, so that such context-aware applications can adapt their behavior to the current situation of their user or environment. Another example are so-called SCADA systems (supervisory control and data acquisition), where complex installations (e.g., energy grids or power plants) are monitored and controlled.

In general, data can be managed either by a database management system (DBMS), or directly by the application. The first approach has many advantages: information demands can be declared by queries and kept separately from the application. When the information demand changes, only the query has to be changed (which dramatically decreases software maintenance costs). In addition, a DBMS can optimize the query execution, so that the requested data is retrieved efficiently from the systems.

However, if applications depend on real-world data, the amount of data, the update rate, and low latency requirements often prevent the storage in a DBMS. Thus, the data streams from the sensors are managed directly by applications, with all its drawbacks.

The goal of data stream management is to provide the same flexibility and data independence for data streams as for stored data. However, despite of performance advantages, DSMS represent no general solution since applications often require the persistent storage of continuous query results as well as a combined processing of current and historical stream data. The performance of integrated database methods, on the other hand, is limited due to the expensive data management of traditional DBMS.

A naïve solution would be to federate or integrate

Wednesday, 25

Thursday Sessions

Paper 23
09:00 - 10:30 Room Velazquez
Session 1 - Data Warehousing and Business
Intelligence

Flexible Information Management, Exploration and Analysis in SAP HANA

Christof Bornhoevd, Robert Kubis, Wolfgang Lehner, Hannes Voigt and Horst Werner
LLC, Palo Alto, U.S.A.

Keywords: Schema-flexible Database Management System, Graph Database, Flexible Information Management.

Abstract: Data management is not limited anymore to towering data silos full of perfectly structured, well integrated data. Today, we need to process and make sense of data from diverse sources (public and on-premise), in different application contexts, with different schemas, and with varying degrees of structure and quality. Because of the necessity to define a rigid data schema upfront, fixed-schema database systems are not a good fit for these new scenarios. However, schema is still essential to give data meaning and to process data purposefully. In this paper, we describe a schema-flexible database system that combines a flexible data model with a powerful data query, analysis, and manipulation language that provides both required schema information and the flexibility required for modern information processing and decision support.

Paper 34
09:00 - 10:30 Room Velazquez
Session 1 - Data Warehousing and Business
Intelligence

Applying Personal and Group-based Trust Models in Document Recommendation

Chin-Hui Lai
*Chung Yuan Christian University, Chung-Li, TaoYuan
County, Taiwan*

Duen-Ren Liu, Cai-Sin Lin
National Chiao Tung University, Hsinchu, Taiwan

Keywords: Collaborative Filtering, Document Recommendation, Group Trust, Role Relationship, Personal Trust, Trust-based Recommender System.

Abstract: Collaborative filtering (CF) recommender systems have been used in various application domains to solve the information-overload problem. Recently, trust-based recommender systems have incorporated the trustworthiness of users into CF techniques to improve the quality of recommendation. Some researchers have proposed rating-based trust

models to derive the trust values based on users' past ratings of items, or based on explicitly specified relations (e.g. friends) or trust relationships. The rating-based trust model may not be effective in CF recommendations, due to unreliable trust values derived from very few past rating records. In this work, we propose a hybrid personal trust model which adaptively combines the rating-based trust model and explicit trust metric to resolve the drawback caused by insufficient past rating records. Moreover, users with similar preferences usually form a group to share items (knowledge) with each other, and thus users' preferences may be affected by group members. Accordingly, group trust can enhance personal trust to support recommendation from the group perspective. Eventually, we propose a recommendation method based on a hybrid model of personal and group trust to improve recommendation performance. The experiment result shows that the proposed models can improve the prediction accuracy of other trust-based recommender systems.

Paper 17
10:45 - 12:15 Room Velazquez
Session 2 - Data Warehousing and Business
Intelligence

Cloud based Privacy Preserving Data Mining with Decision Tree

Echo P. Zhang, Yi-Jun He and Lucas C. K. Hui
The University of Hong Kong, Hong Kong, Hong Kong

Keywords: Data Mining, Privacy Preserving Data Mining, Cloud Computing, Privacy in Cloud Computing.

Abstract: Privacy Preserving Data Mining (PPDM) aims at performing data mining among multiple parties, and at the meantime, no single party suffers the threat of releasing private data to any others. Nowadays, cloud service becomes more and more popular. However, how to deal with privacy issues of cloud service is still developing. This paper is one of the first researches in cloud server based PPDM. We propose a novel protocol that the cloud server performs data mining in encrypted databases, and our solution can guarantee the privacy of each client. This scheme can protect client from malicious users. With aid of a hardware box, the scheme can also protect clients from untrusted cloud server. Another novel feature of this solution is that it works even when the database from different parties are overlapping.

Paper 33
10:45 - 12:15 Room Velazquez
Session 2 - Data Warehousing and Business
Intelligence

Inference in Hierarchical Multidimensional Space

Alexandr Savinov

SAP Research, Dresden, Germany

Keywords: Multidimensional Data Models, Data Analytics, Concept-oriented Model, Inference, Query Languages.

Abstract: In spite of its fundamental importance, inference has not been an inherent function of multidimensional models and analytical applications. These models are mainly aimed at numeric analysis where the notion of inference is not well defined. In this paper we define inference using only multidimensional terms like axes and coordinates as opposed to using logic-based approaches. We propose an inference procedure which is based on a novel formal setting of nested partially ordered sets with operations of projection and de-projection.

12:15 - 13:15 Room Plenary
On Knowledge Discovery and Interactive Intelligent
Visualization of Biomedical Data - Challenges in
Human-Computer Interaction & Biomedical
Informatics
Keynote Speaker: Andreas Holzinger

On Knowledge Discovery and Interactive Intelligent Visualization of Biomedical Data Challenges in Human-Computer Interaction & Biomedical Informatics

Andreas Holzinger

Medical University Graz, Graz, Austria

Abstract: Biomedical Informatics can be defined as “the interdisciplinary field that studies and pursues the effective use of biomedical data, information and knowledge for scientific inquiry, problem solving, and decision making, motivated by efforts to improve human health.” However, professionals in the life sciences are faced with an increasing quantity of highly complex, multi-dimensional and weakly structured data. While researchers in Human-Computer Interaction (HCI) and Information Retrieval/Knowledge Discovery in Databases (IR/KDD) have long been independently working to develop methods that can support expert end users to identify, extract and understand information out of this data, it is obvious that an interdisciplinary approach to bring these two fields closer together can yield synergies in the application of these methods to weakly structured complex

medical data sets. The aim is to support end users to learn how to interactively analyse information properties and to visualize the most relevant parts – in order to gain knowledge, and finally wisdom, to support a smarter decision making. The danger is not only to get overwhelmed by increasing masses of data, moreover there is the risk of modelling artifacts.

Paper 19
14:30 - 16:30 Room Madrid
Parallel Session 3 - Ontologies and the Semantic
Web

An Architecture based on Ontologies, Agents and Metaheuristics Applied to the Multimedia Service of the Brazilian Digital Television System

Toni Ismael Wickert and Arthur Tórgo Gómez

*Universidade do Vale do Rio dos Sinos, São Leopoldo,
Brazil*

Keywords: Ontologies, Metaheuristics, Tabu Search, Genetic Algorithm, Software Agents.

Abstract: With the advent of the Brazilian Digital Television System, that arrives on approximately 95% of Brazilian homes, the users will be able to have an interactive channel by the utilization of the digital television. Thus, will be possible to access the multimedia application server, i.e., to send or to receive emails, to access interactive applications, to watch movies or specific news. This paper proposes the development and the implementation of an architecture that includes a module that suggests the content to the user according to his profile and another module to optimize the content that will be transmitted. The implementation was developed using ontologies, software agents, Tabu Search and Genetic Algorithm. The validations of the results are done using a metric.

Paper 22
14:30 - 16:30 Room Madrid
Parallel Session 3 - Ontologies and the Semantic
Web

Effective and Efficient Online Communication The Channel Model

Anna Fensel, Dieter Fensel, Birgit Leiter and
Andreas Thalhammer

University of Innsbruck, Innsbruck, Austria

Keywords: Social Media, Web 2.0, Semantic Web, Dissemination, Communication, Knowledge Management.

Abstract: We discuss the challenge of scalable dissemination approach in a world where the

number of communication channels and interaction possibilities is growing exponentially, particularly on the Web, Web 2.0, and semantic channels. Our goal is to enable smaller organizations to fully exploit this potential. We have developed a new methodology based on distinguishing and explicitly interweaving content and communication as a central means for achieving content reusability and thereby scalability over various, heterogeneous channels. Here, we present in detail the communication channel model of our approach.

Paper 31
14:30 - 16:30 Room Madrid
Parallel Session 3 - Ontologies and the Semantic Web

Effective Keyword Search via RDF Annotations

Roberto De Virgilio and Lorenzo Dolfi
Università Roma Tre, Rome, Italy

Keywords: Visualization of Semantic Applications, Semantic Annotation, Web Browsing, Centrality Indices.

Abstract: Searching relevant information from Web may be a very tedious task. If people cannot navigate through the Web site, they will quickly leave. Thus, designing effective navigation strategies on Web sites is crucial. In this paper we provide and implement centrality indices to guide the user for an effective navigation of Web pages. We get inspiration from well-know location family problems to compute the center of a graph: a joint use of such indices guarantees the automatic selection of the best starting point. To validate our approach, we have developed a system that implements the techniques described in this paper on top of an engine for keyword-based search over RDF data. Such system exploits an interactive front-end to support the user in the visualization of both annotations and corresponding Web pages. Experiments over widely used benchmarks have shown very good results, in terms of both effectiveness and efficiency.

Paper 54
14:30 - 16:30 Room Madrid
Parallel Session 3 - Ontologies and the Semantic Web

Toward a Product Search Engine based on User Reviews

Paolo Fosci and Giuseppe Psaila
Università di Bergamo, Dalmine, Italy

Keywords: Product Search Engine, User Reviews, Itemset Mining.

Abstract: We address the problem of developing a method for retrieving products exploiting product user-reviews that can be found on the internet. For this purpose, we introduce a ranking model based on the concept of itemset mining of frequent terms. The prototype search engine that implements the proposed retrieval model is illustrated, and a preliminary evaluation on a real data set is discussed.

Paper 58
14:30 - 16:30 Room Madrid
Parallel Session 3 - Ontologies and the Semantic Web

Smart Learning Management System Framework

Yeong-Tae Song, Yuanqiong Wang, Sungchul Hong
Towson University, Towson, U.S.A.

Yong-Ik Yoon
Sookmyung University, Seoul, Korea, Republic of

Keywords: Smart Learning Management System, Learning Objects, User Profile, Ontology, RDF, Semantic Search.

Abstract: Thanks to modern networking technologies and advancement of social networks, people in the modern society need more and more information just to be in the game. With such environment, the importance of learning and information sharing cannot be overemphasized. Even though plethora of information is available on various sources such as the web, libraries, and any learning material repositories, if it is not readily available and meets the needs of the user, it may not be utilized. For that, we need a system that can help provide customized information – matches with user's level and interest - to the user. Such system should understand what the user's interests are, what level the user belongs for the topic, and so on. In this paper, we are proposing a framework for smart learning management system (SLMS) that utilizes user profiles and semantically organized learning objects so only the relevant information can be delivered to the user. The SLMS maintains user profiles – continuously updating whenever there is a change – and learning objects that are organized by building ontology. Upon user's request, the system fetches relevant learning materials based on the user's profile. The delivered learning materials are suitable for the user's topic and the level for the requested topic sorted by relevancy ranking.

Paper 16
14:30 - 16:30 Room Plenary
Parallel Session 3 - Data Management and Quality

On Continuous Top-k Similarity Joins

Da Jun Li

National Tsing Hua University, Hsinchu, Taiwan

En Tzu Wang

Industrial Technology Research Institute, Hsinchu, Taiwan

Yu-Chou Tsai

Institute for Information Industry, Taipei, Taiwan

Arbee L. P. Chen

National Chengchi University, Taipei, Taiwan

Keywords: Data Stream, Similarity Join, Continuous Query, Top-K Query.

Abstract: Given a similarity function and a threshold σ within a range of $[0, 1]$, a *similarity join* query between two sets of records returns pairs of records from the two sets, which have similarity values exceeding or equaling σ . Similarity joins have received much research attention since it is a fundamental operation used in a wide range of applications such as duplicate detection, data integration, and pattern recognition. Recently, a variant of similarity joins is proposed to avoid the need to set the threshold σ , i.e. *top-k similarity joins*. Since data in many applications are generated as a form of continuous data streams, in this paper, we make the first attempt to solve the problem of top-k similarity joins considering a dynamic environment involving a data stream, *named continuous top-k similarity joins*. Given a set of records as the query, we continuously output the top-k pairs of records, ranked by their similarity values, for the query and the most recent data, i.e. the data contained in the sliding window of a monitored data stream. Two algorithms are proposed to solve this problem. The first one extends an existing approach for static datasets to find the top-k pairs regarding the query and the newly arrived data and then keep the obtained pairs in a candidate result set. As a result, the top-k pairs can be found from the candidate result set. In the other algorithm, the records in the query are preprocessed to be indexed using a novel data structure. By this structure, the data in the monitored stream can be compared with all records in the query at one time, substantially reducing the processing time of finding the top-k results. A series of experiments are performed to evaluate the two proposed algorithms and the experiment results demonstrate that the algorithm with preprocessing outperforms the other algorithm extended from an existing approach for a static environment.

Paper 35
14:30 - 16:30 Room Plenary
Parallel Session 3 - Data Management and Quality

Data Quality Sensitivity Analysis on Aggregate Indicators

Mario Mezzanzanica, Roberto Boselli, Mirko Cesarini

University of Milan Bicocca, Milan, Italy

Fabio Mercorio

University of Milano-Bicocca, Milan, Italy

Keywords: Data Quality, Data Cleansing, Sensitivity Analysis, Inconsistent Databases, Aggregate Indicators, Uncertainty Assessment.

Abstract: Decision making activities stress data and information quality requirements. The quality of data sources is frequently very poor, therefore a cleansing process is required before using such data for decision making processes. When alternative (and more trusted) data sources are not available data can be cleansed only using business rules derived from domain knowledge. Business rules focus on fixing inconsistencies, but an inconsistency can be cleansed in different ways (i.e. the correction can be not deterministic), therefore the choice on how to cleanse data can (even strongly) affect the aggregate values computed for decision making purposes. The paper proposes a methodology exploiting Finite State Systems to quantitatively estimate how computed variables and indicators might be affected by the uncertainty related to low data quality, independently from the data cleansing methodology used. The methodology has been implemented and tested on a real case scenario providing effective results.

Paper 37
14:30 - 16:30 Room Plenary
Parallel Session 3 - Data Management and Quality

Combining Local and Related Context for Word Sense Disambiguation on Specific Domains

Franco Rojas-Lopez, Ivan Lopez-Arevalo and Victor Sosa-Sosa

Cinvestav-Tamaulipas, Victoria, Mexico

Keywords: Semantic Similarity, Local and Related Context, Word Sense Disambiguation.

Abstract: In this paper an approach for word sense disambiguation in documents is presented. For Word Sense Disambiguation (WSD), the local and related context for an ambiguous word is extracted, such context is used for retrieve second order vectors from WordNet. Thus two graphs are built at the same time and evaluated individually, finally both results are

combined to automatically assign the correct sense for the ambiguous word. The proposed approach was tested on the task #17 of the SemEval 2010 international competition producing promising results compared to other approaches.

Paper 70
14:30 - 16:30 Room Plenary
Parallel Session 3 - Data Management and Quality

**DBMS meets DSMS
Towards a Federated Solution**

Andreas Behrend
University of Bonn, Bonn, Germany

Dieter Gawlick
Oracle Redwood City, Redwood City, U.S.A.

Daniela Nicklas
*Carl von Ossietzky Universität Oldenburg, Oldenburg,
Germany*

Keywords: Data Stream Management, Event Processing, Active Databases, Database Architecture, Query Processing.

Abstract: In this paper, we describe the requirements and benefits for integrating data stream processing with database management systems. Currently, these technologies focus on very different tasks; streams systems extract instances of patterns from streams of transient data, while database systems store, manage, provide access to, and analyze persistent data. Many applications, e.g., patient care, program trading, or flight supervision, however, depend on the functionality and operational characteristics of both types of systems. We discuss how to design a federated system which provides the benefits of both approaches.

Thursday, 26

Friday Sessions

09:30 - 10:30 Room Plenary
 Integrated Marine Data Management - Data
 Management Strategies for Risk Reduction in the
 Offshore Wind Industry
 Keynote Speaker: John Shaw

Integrated Marine Data Management Data Management Strategies for Risk Reduction in the Offshore Wind Industry

John Shaw

Mainstream Renewable Power, Dublin, Ireland

Abstract: There is a compelling need for an Integrated Sea Information System, ISIS, to accelerate the €6.4 Trillion European Offshore Wind Industry. ISIS will reduce the risk of developing, constructing and operating Offshore Wind To create ISIS we need new EU policies, Regulations, Standards, greater collaboration and Innovative Technologies. This presentation will explain Mainstream's Vision, Offshore Wind Developers' data needs, Mainstream's Data Management Strategy and provide details on the ISIS Vision, Plan and Consortium.

Paper 30 Foyer
 10:30 - 11:30
 Poster Session

Ontology Similarity Measurement Method in Rapid Data Integration

Juebo Wu, Chen-Chieh Feng and Chih-Yuan Chen
National University of Singapore, Singapore, Singapore

Keywords: Ontology Similarity, Rapid Data Integration, Ontology Mapping.

Abstract: Rapid data integration has been a challenging topic in the field of computer science and its related subjects, widely used in data warehouse, artificial intelligence, biological medicine, and geographical information system etc. In this paper, we present a method of ontology similarity measurement in rapid data integration, by means of semantic ontology from high level perspective. The edit distance algorithm is introduced as the basic principle for ontology similarity calculation. A case study is carried out and the result shows that the presented method is feasible and effective.

Paper 43 Foyer
 10:30 - 11:30
 Poster Session

Ontology-engineered MSME Framework

M. Saravanan

Ericsson India Private Ltd, Chennai, India

M. Amirtha Varshini, Brindha S.

Meenakshi Sundararajan Engineering College, Chennai, India

Keywords: Micro Small and Medium Scale Enterprises, Middlemen, Ontology Engineered Framework, JSON, Android, Cloud Environment.

Abstract: Micro, Small and Medium scale Enterprises (MSMEs) hold an unflinching distinction of being pillars of equitable economic growth. Lack of proper business platforms and knowledge of marketing strategies render MSMEs vulnerable to middlemen exploitation. The Web has extended e-business platforms, e-commerce and micro financing solutions to assist MSMEs. These web-based solutions fail to obliterate intermediation. In view of the advancements and customers' growth in the telecommunications field, we utilize the mobile platform to offer trading solutions to MSMEs. In this paper, we propose a mobile phone-based ontology engineered framework for MSMEs that can achieve disintermediation. The framework has been tested on mobile cloud akin to EC2 cloud environment and integrated with an android application that provides easy access - anytime, anywhere. The envisioned framework will boost MSME margins, build healthy business-ties and transform MSMEs into self-sufficient establishments equipped with full-fledged trading systems that operate in mobile phone enabled environment.

Paper 45 Foyer
 10:30 - 11:30
 Poster Session

Geometric Divide and Conquer Classification for High-dimensional Data

Pei Ling Lai, Yang Jin Liang

Southern Taiwan University, Tainan, Taiwan

Alfred Inselberg

Tel Aviv University, Tel Aviv, Israel

Keywords: Classification, Divide and Conquer, Parallel Coordinates, Visualization.

Abstract: From the **Nested Cavities** (abbr. **NC**) classifier (Inselberg and Avidan, 2000) a powerful new classification approach emerged. For a dataset

P and a subset $S \subset P$ the classifier constructs a rule distinguishing the elements of S from those in $P - S$. The **NC** is a geometrical algorithm which builds a sequence of nested unbounded parallelpipeds of minimal dimensionality containing disjoint subsets of P , and from which a hypersurface (the rule) containing the subset S is obtained. The partitioning of $P - S$ and S into *disjoint* subsets is very useful when the original rule obtained is either too complex or imprecise. As illustrated with examples, this separation reveals exquisite insight on the dataset's structure. Specifically from one of the problems we studied *two different types of watermines were separated*. From another dataset, *two distinct types of ovarian cancer* were found. This process is developed and illustrated on a (sonar) dataset with 60 variables and two categories ("mines" and "rocks") resulting in significant understanding of the domain and simplification of the classification rule. Such a situation is generic and occurs with other datasets as illustrated with a similar decompositions of a financial dataset producing two sets of conditions determining gold prices. The divide-and-conquer extension can be automated and also allows the classification of the sub-categories to be done in parallel.

Paper 47
10:30 - 11:30
Poster Session Foyer

Graph-based Campaign Amplification in Telecom Cloud

M. Saravanan, Sandeep Akhouri

Ericsson India Global Services, Chennai, India

Loganath Thamizharasu

Anna University, Chennai, India

Keywords: Telecom Cloud, Campaign Management, NoSQL Graph Database, Map Reduce Framework.

Abstract: Majority of telecom operators are making a transition from a monolithic, stove-pipe approach of creating services to a more flexible architecture that provides them agility to rapidly develop and deploy services. New revenue streams require an ability to rapidly identify and target dynamic shifts in traffic patterns and subscriber behaviour. As subscriber behaviour morphs with plans, promotions, devices, location and time, this presents challenges and opportunities for an operator to create and launch targeted campaigns. The enormous volume of data being generated requires a scalable platform for processing massive xDR (eg. Call Detail Records). This paper proposes graph databases in a telecom cloud environment for quickly identifying trends, isolating a targeted subscriber base and

rapidly launching campaigns. We also highlight the limitations of a conventional relational database in terms of capturing complex relationships as compared to a NoSQL graph database and the benefits of automatic provisioning and deployment in the cloud environment.

Paper 51
10:30 - 11:30
Poster Session Foyer

Create a Specialized Search Engine The Case of an RSS Search Engine

Robert Viseur

Université de Mons, Mons, Belgium

Keywords: API, Atom, Crawler, Indexer, Mashup, Open Source, RSS, Search Engine, Syndication.

Abstract: Several approaches are possible for creating specialized search engines. For example, you can use the API of existing commercial search engines or create engine from scratch with reusable components such as open source indexer. RSS format is used for spreading information from websites, creating new applications (mashups), or collecting information for competitive or technical watch. In this paper, we focus on the study case of an RSS search engine development. We identify issues and propose ways to address them.

Paper 71
10:30 - 11:30
Poster Session Foyer

Understanding Worldwide Human Information Needs

Revealing Cultural Influences in HCI by Analyzing Big Data in Interactions

Rüdiger Heimgärtner

Intercultural User Interface Consulting (IUIIC), Undorf, Germany

Keywords: Culture, Relation, Structure Equation Model, Neural Networks, Factor Analysis, Data Analysis, Global, Local, Intercultural Interaction Analysis, Tools, Big Data, Modeling, Data Mining, Statistics.

Abstract: Understanding human information needs worldwide requires the analysis of much data and adequate statistical analysis methods. Factor Analysis and Structural Equation Models (SEM) are a means to reveal structures in data. Data from empirical studies found in literature regarding cultural human computer interaction (HCI) was analyzed using these methods to develop a model of culturally influenced HCI. There are significant differences in HCI style depending on the cultural imprint of

the user. Having knowledge about the relationship between culture and HCI using this model, the local human information needs can be predicted for a worldwide scope.

Paper 52
11:30 - 13:00 Room Dali
Session 4 - Trust, Privacy and Security

NexusDSS: A System for Security Compliant Processing of Data Streams

Nazario Cipriani¹, Christoph Stach¹, Oliver Dörler²
and Bernhard Mitschang¹

¹ *Universität Stuttgart, Stuttgart, Germany*

² *Steinäcker 54, Aichwald, Germany*

Keywords: Accessibility of Data, Privacy Policies, Data Stream Processing.

Abstract: Technological advances in microelectronic and communication technology are increasingly leading to a highly connected environment equipped with sensors producing a continuous flow of context data. The steadily growing number of sensory context data available enables new application scenarios and drives new processing techniques. The growing pervasion of everyday life with social media and the possibility of interconnecting them with moving objects' traces, leads to a growing importance of access control for this kind of data since it concerns privacy issues. The challenge is twofold: First mechanisms to control data access and data usage must be established and second efficient and flexible processing of sensible data must be supported. In this paper we present a flexible and extensible security framework which provides mechanisms to enforce requirements for context data access and beyond that support safe processing of sensible context data according to predefined processing rules. In addition and in contrast to previous concepts, our security framework especially supports fine-grained control to contextual data.

Paper 42
11:30 - 13:00 Room Dali
Session 4 - Trust, Privacy and Security

Towards Process Centered Information Security Management

A Common View for Federated Business Processes and Personal Data Usage Processes

Erik Neitzel and Andreas Witt

Otto-von-Guericke University, Magdeburg, Germany

University of Applied Sciences Brandenburg, Brandenburg, Germany

Keywords: Security, Privacy, Federated Business Processes, Social Networks, Information Security (IS), Information Security Management Systems (ISMS).

Abstract: While comparing the progress of our two research projects of developing an information security management system (ISMS) for federated business process landscapes and the enhancement of security of social networks, we discovered a fundamental view congruency concerning the way information security can be handled. This paper deals with a conceptual framework which uses the ISO 27001 and the German BSI IT-Grundschutz Framework as a base for determining a methodology for a process based point of view towards information security management for both federated business processes within business applications and personal data usage processes within social networks. The proposed layers are (1) process layer, (2) application layer, (3) network layer, (4) IT systems layer and (5) infrastructure layer.

Paper 5
14:15 - 16:15 Room Dali
Parallel Session 5 - Data Management and Quality

NRank: A Unified Platform Independent Approach for Top-K Algorithms

Martin Čech

Lekis s.r.o., Benešov, Czech Republic

Jaroslav Pokorný

Charles University, Prague, Czech Republic

Keywords: Top-K Query, Relational Database, Rank Join.

Abstract: Due to increasing capacity of storage devices and speed of computer networks during last years, it is still more required to sort and search data effectively. A query result containing thousands of rows from a relational database is usually useless and unreadable. In that situation, users may prefer to define constraints and sorting priorities in the query,

and see only several top rows from the result. This paper deals with top-k queries problems, extension of relational algebra by new operators and their implementation in a database system. It focuses on optimization of operations join and sort. The work also includes implementation and comparison of some algorithms in standalone .NET library NRank.

Paper 15
14:15 - 16:15 Room Dali
Parallel Session 5 - Data Management and Quality

FIND A Data Cloud Platform for Financial Data Services

Zhicheng Liao, Yun Xiong and Yangyong Zhu
Fudan University, Shanghai, China

Keywords: Cloud Computing, Data Cloud, Securities Data Integration, Data Service.

Abstract: In recent years, researchers have paid more interest in dealing with large scale of data. However, it is difficult to discover patterns from various sources of big data flexibly and efficiently. In this paper, we design a data cloud platform for financial data services (FIND), and implement a prototype system to evaluate the performance and usability of the data cloud. FIND consists of a cloud infrastructure, a data resource center and a data service portal. FIND provides high performance computation capability, high quality integrated financial data, sophisticated data mining algorithms, and powerful data services.

Paper 18
14:15 - 16:15 Room Dali
Parallel Session 5 - Data Management and Quality

How Do I Manage My Personal Data? – A Telco Perspective

Corrado Moiso
Telecom Italia, Torino, Italy
Fabrizio Antonelli, Michele Vescovi
Telecom Italia, Trento, Italy

Keywords: Personal Data, Personal Data Ecosystems, Personal Data Stores and Services, Privacy.

Abstract: Personal data are considered the core of digital services. Data Privacy is the main concern in the currently adopted “organization-centric” approaches for personal data management: this affects the potential benefits arising from a smarter and more valuable use of personal data. We introduce a “user-centric” model for personal data

management, where individuals have control over the entire personal data lifecycle from acquisition to storage, from processing to sharing. In particular, the paper analyses the features of a personal data store and discusses how its adoption enables new application scenarios.

Paper 20
14:15 - 16:15 Room Dali
Parallel Session 5 - Data Management and Quality

KIDS A Model for Developing Evolutionary Database Applications

Zhen Hua Liu¹, Andreas Behrend², Eric Chan¹,
Dieter Gawlick¹ and Adel Ghoneimy¹
¹ 500 Oracle ParkWay, Redwood Shores, U.S.A.
² Universität Bonn, Bonn, Germany

Keywords: Knowledge Management, Information, Data Management, Event Processing, Workflow.

Abstract: Database applications enable users to handle the ever-increasing amount and complexity of data, knowledge as well as the dissemination of information to ensure timely response to critical events. However, the very process of human problem solving, which requires understanding and tracking the evolution of data, knowledge, and events, is still handled mostly by human and not by databases and their applications. In this position paper, we propose KIDS as a model that reflects the way human are solving problems. We propose to use KIDS as a blueprint to extend database technologies to manage data, knowledge, directives and events in a coherent, self-evolving way. Our proposal is based on our experience of building database centric applications that require comprehensive interactions among facts, information, events, and knowledge.

Paper 50
14:15 - 16:15 Room Dali
Parallel Session 5 - Data Management and Quality

Towards a Meaningful Analysis of Big Data Enhancing Data Mining Techniques through a Collaborative Decision Making Environment

Nikos Karacapilidis¹, Stefan Rueping², Georgia Tsiliki³ and Manolis Tzagarakis¹

¹ University of Patras, Rio Patras, Greece

² Fraunhofer IAIS, Schloss Birlinghoven, Sankt Augustin, Germany

³ Academy of Athens, Athens, Greece

Keywords: Big Data, Data Mining, Decision Support Systems, Modeling and Managing Large Data Systems.

Abstract: Arguing that dealing with data-intensive settings is not a technical problem alone, we propose a hybrid approach that builds on the synergy between machine and human intelligence to facilitate the underlying sense-making and decision making processes. The proposed approach, which can be viewed as an innovative workbench incorporating and orchestrating a set of interoperable services, is illustrated through a real case concerning collaborative subgroup discovery in microarray data. Evaluation results, validating the potential of our approach, are also included.

Paper 65
14:15 - 16:15 Room Dali
Parallel Session 5 - Data Management and Quality

A Data-Centric Approach for Networking Applications

Ahmad Ahmad-Kassem¹, Christophe Bobineau², Christine Collet², Etienne Dublé³, Stéphane Grumbach¹, Fuda Ma⁴, Lourdes Martinez² and Stéphane Ubéda¹

¹ INRIA, Villeurbanne, France

² Grenoble Institute of Technology, Grenoble, France

³ CNRS, Grenoble, France

⁴ INRIA, INSA-Lyon, Villeurbanne, France

Keywords: Declarative Networking, Programming Abstraction, Case-based Distributed Query Optimization.

Abstract: The paper introduces our vision for rapid prototyping of heterogeneous and distributed applications. It abstracts a network as a large distributed database providing a unified view of "objects" handled in networks and applications. The applications interact through declarative queries including declarative networking programs (e.g.

routing) and/or specific data-oriented distributed algorithms (e.g. distributed join). Case-Based Reasoning is used for optimization of distributed queries by learning when there is no prior knowledge on queried data sources and no related metadata such as data statistics.

Paper 38
14:15 - 16:15 Room Velazquez
Parallel Session 5 - Data Warehousing and Business Intelligence

A Virtual Document Approach for Keyword Search in Databases

Jaime I. Lopez-Veyna, Victor J. Sosa-Sosa and Ivan Lopez-Arevalo

Center of Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV), Cd. Victoria, Mexico

Keywords: Keyword Search, Indexing, Databases, Top-k, Virtual Documents.

Abstract: It is clear that in recent years the amount of information available in a variety of data sources, like those found on the Web, has presented an accelerated growth. This information can be classified based on its structure in three different forms: unstructured (free text documents), semi-structured (XML documents) and structured (a relational database or XML database). A search technique that has gained wide acceptance for use in massive data sources, such as the Web, is the keyword based search, which is simple to people who are familiar with the use of Web search engines. Keyword search has become an alternative to users without any knowledge about formal query languages and schema used in structured data. There are some traditional approaches to perform keyword search over relational databases such as Steiner Trees, Candidate Networks and recently Tuple Units. Nevertheless these methods have some limitations. In this paper we propose a Virtual Document (VD) approach for keyword search in databases. We represent the structured information as graphs and propose the use of an index that captures the structural relationships of the information. This approach produce fast and accuracy results in search responses. We have conducted extensive experiments on large-scale real databases and the results demonstrates that our approach achieves high search efficiency and high accuracy for keyword search in databases.

Paper 25
14:15 - 16:15 Room Velazquez
Parallel Session 5 - Data Warehousing and
Business Intelligence

**Constrained Nonnegative Matrix Factorization
based Data Distortion Techniques
Study of Data Privacy and Utility**

Nirmal Thapa¹, PengPeng Lin¹, Lian Liu¹, Jie
Wang² and Jun Zhang¹

¹ University of Kentucky, Lexington, U.S.A.

² Indiana University Northwest, Gary, U.S.A.

Keywords: Data Distortion, Nonnegative Matrix
Factorization, Constrained NMF, SVD, NMF.

Abstract: With the rise of data mining techniques came across the problem of privacy disclosure, that is why it has become one of the top priorities as far as designing the data mining techniques is concerned. In this paper, we briefly discuss the Nonnegative Matrix Factorization (NMF) and the motivation behind using NMF for data representation. We provide the mathematical derivation for NMF with some additional constraints. Based on the mathematical derivations, we propose a couple of novel data distortion strategies. The first technique is called the Constrained Nonnegative Matrix Factorization (CMF) and the second one is Sparsified CNMF. We study the distortion level of each of these algorithms with the other matrix based techniques like SVD and NMF. K-means is used to study the data utility of the two proposed methods. Our experimental results show that, in comparison with standard data distortion techniques, the proposed schemes are very effective in achieving a good tradeoff between data privacy and data utility, and affords a feasible solution to protect sensitive information and promise higher accuracy in decision making. We investigate utility of the perturbed data based on the results from the original data.

Paper 27
14:15 - 16:15 Room Velazquez
Parallel Session 5 - Data Warehousing and
Business Intelligence

**Kernel Generations for a Diagnosis Model with
GP**

Jongseong Kim and Hoo-Gon Choi

Sungkyunkwan University, Suwon, Korea, Republic of

Keywords: Diagnosis Model, Genetic Programming,
Absorbing Evolution, Accuracy and Recall Rate,
Computing Time.

Abstract: An accurate diagnosis model is required to

diagnose the medical subjects. The subjects should be diagnosed with high accuracy and recall rate by the model. The laboratory test data are collected from 953 latent subjects having type 2 diabetes mellitus. The results are classified into patient group and normal group by using support vector machine kernels optimized through genetic programming. Genetic programming is applied for the input data twice with absorbing evolution, which is a new approach. The result shows that new approach creates a kernel with 80% accuracy, 0.794 recall rate and 28% reduction of computing time comparing to other typical methods. Also, the suggested kernel can be easily utilized by users having no and little experience on large data.

Paper 29
14:15 - 16:15 Room Velazquez
Parallel Session 5 - Data Warehousing and
Business Intelligence

**Automatic Subspace Clustering with Density
Function**

Jiwu Zhao and Stefan Conrad

Heinrich-Heine University, Duesseldorf, Germany

Keywords: Subspace Clustering, Density Function,
High Dimension.

Abstract: Clustering techniques in data mining aim to find interesting patterns in data sets. However, traditional clustering methods are not suitable for large, high-dimensional data. Subspace clustering is an extension of traditional clustering that enables finding clusters in subspaces within a data set, which means subspace clustering is more suitable for detecting clusters in high-dimensional data sets. However, most subspace clustering methods usually require many complicated parameter settings, which are always troublesome to determine, and therefore there are many limitations for applying these subspace clustering methods. In this article, we develop a novel subspace clustering method with a new density function, which computes and represents the density distribution directly in high-dimensional data sets, and furthermore the new method requires as few parameters as possible.

16:30 - 17:30 Room Plenary
Visualization & Data Mining for High Dimensional
Datasets
Keynote Speaker: Alfred Inselberg

Visualization & Data Mining for High Dimensional Datasets

Alfred Inselberg

Tel Aviv University, Tel Aviv, Israel

Abstract: A dataset with M items has 2^M subsets any one of which may be the one which satisfies our objectives. With a good data display and interactivity our fantastic pattern-recognition can cut great swaths searching through this combinatorial explosion and also extract insights from the visual patterns. These are the core reasons for data visualization. With parallel coordinates the search for relations in multivariate datasets is transformed into a 2-D pattern recognition problem. The foundations are developed interlaced with applications. Guidelines and strategies for knowledge discovery are illustrated on several real datasets (financial, process control, credit-score, intrusion-detection etc) one with hundreds of variables. A geometric classification algorithm is presented and applied to complex datasets. It has low computational complexity providing the classification rule explicitly and visually. The minimal set of variables required to state the rule (features) is found and ordered by their predictive value. Multivariate relations can be modeled as hypersurfaces and used for decision support. A model of a (real) country's economy reveals sensitivities, impact of constraints, trade-offs and economic sectors unknowingly competing for the same resources. An overview of the methodology provides foundational understanding; learning the patterns corresponding to various multivariate relations. These patterns are robust in the presence of errors and that is good news for the applications. We stand at the threshold of breaching the gridlock of multidimensional visualization. The parallel coordinates methodology has been applied to collision avoidance and conflict resolution algorithms for air traffic control (3 USA patents), computer vision (1 USA patent), data mining (1 USA patent), optimization, decision support and elsewhere.

Closing Session
17:30 - 17:45 Room Plenary

