Lecture Notes in Computer Science **9050**

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

More information about this series at http://www.springer.com/series/7409

Matthias Renz · Cyrus Shahabi
Xiaofang Zhou · Muhammad Aamir Cheema (Eds.)

Database Systems for Advanced Applications

20th International Conference, DASFAA 2015
Hanoi, Vietnam, April 20–23, 2015
Proceedings, Part II

Springer

Editors
Matthias Renz
Universität München
München
Germany

Xiaofang Zhou
University of Queensland
Brisbane
Australia

Cyrus Shahabi
University of Southern California
Los Angeles
USA

Muhammad Aamir Cheema
Monash University
Clayton
Australia

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-18122-6 ISBN 978-3-319-18123-3 (eBook)
DOI 10.1007/978-3-319-18123-3

Library of Congress Control Number: 2015936691

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

It is our great pleasure to welcome you to DASFAA 2015, the 20th edition of the International Conference on Database Systems for Advanced Applications (DASFAA 2015), which was held in Hanoi, Vietnam during April 20–23, 2015. Hanoi (Vietnamese: *Hà Nội*), the capital of Vietnam, is the second largest city in Vietnam and has collected all the essence, unique features, and diversification of Vietnamese culture. The city is preserving more than 4000 historical and cultural relics, architecture and beauty spots, in which nearly 900 relics have been nationally ranked with hundreds of pagodas, temples, architectural works, and sceneries. Handcraft streets and traditional handcraft villages remain prominent and attractive to tourists when visiting Hanoi, many of which centered around the Hoan Kiem Lake in the Old Quarter, close to the conference venue. Hanoi has recently been included on TripAdvisor's list of best destinations in the world, ranked 8th among the world's top 25 destinations.

We are delighted to offer an exciting technical program, including two keynote talks by Amr El Abbadi (University of California, Santa Barbara) and Luc Vincent (Google Inc.); one 10-year best paper award presentation; a panel session on "Big Data Search and Analysis;" a poster session with 18 papers; a demo session with 6 demonstrations; an industry session with 3 full paper presentations; 3 tutorial sessions; and of course a superb set of research papers. This year, we received 287 submissions, each of which went through a rigorous review process. That is, each paper was reviewed by at least three Program Committee members, followed by a discussion led by the meta-reviewers, and a final meta-review prepared for each paper. At the end, DASFA 2015 accepted 63 full papers (the acceptance ratio is 22%).

Two workshops were selected by the Workshop Co-chairs to be held in conjunction with DASFAA 2015. They are the Second International Workshop on Big Data Management and Service (BDMS 2015), and the Second International Workshop on Semantic Computing and Personalization (SeCoP 2015). The workshop papers are included in a separate volume of proceedings also published by Springer in its Lecture Notes in Computer Science series.

The conference received generous financial support from the Hanoi University of Science and Technology (HUST). We, the conference organizers, also received extensive help and logistic support from the DASFAA Steering Committee and the Conference Management Toolkit Support Team at Microsoft.

We are grateful to all conference organizers, Han Su (University of Queensland) and many other people, for their great effort in supporting conference organization. Special thanks also go to the DASFAA 2015 Local Organizing Committee: Tuyet-Trinh Vu, Hong-Phuong Nguyen, and Van Thu Truong, all from the Hanoi University of Science

and Technology, Vietnam. Finally, we would like to take this opportunity to thank all the meta-reviewers, Program Committee members, and external reviewers for their expertise and help in evaluating papers, and all the authors who submitted their papers to this conference.

February 2015

Quyet-Thang Huynh
Qing Li
Matthias Renz
Cyrus Shahabi
Xiaofang Zhou

Organization

General Co-chairs

Qing Li	City University of Hong Kong, HKSAR, Hong Kong
Quyet-Thang Huynh	Hanoi University of Science and Technology, Vietnam

Program Committee Co-chairs

Cyrus Shahabi	University of Southern California, USA
Matthias Renz	Ludwig-Maximilians-Universität München, Germany
Xiaofang Zhou	University of Queensland, Australia

Tutorial Co-chairs

Arbee L.P. Chen	NCCU, Taiwan
Pierre Senellart	Télécom ParisTech, France

Workshops Co-chairs

An Liu	Soochow University, China
Yoshiharu Ishikawa	Nagoya University, Japan

Demo Co-chairs

Haiwei Pan	Harbin Engineering University, China
Binh Minh Nguyen	Hanoi University of Science and Technology, Vietnam

Panel Co-chairs

Bin Cui	Peking University, China
Katsumi Tanaka	Kyoto University, Japan

Poster Co-chairs

Sarana Nutanong	City University of Hong Kong, China
Tieyun Qian	Wuhan University, China

Industrial/Practitioners Track Co-chairs

Mukesh Mohania	IBM, India
Khai Tran	Oracle, USA

PhD Colloquium

Khoat Than	Hanoi University of Science and Technology, Vietnam
Ge Yu	Northeastern University, China
Tok Wang Ling	National University of Singapore, Singapore
Duong Nguyen Vu	John Von Neumann Institute - VNU-HCMUS, Vietnam

Publication Chair

Muhammad Aamir Cheema	Monash University, Australia

Publicity Co-chairs

Yunjun Gao	Zhejiang University, China
Bao-Quoc Ho	VNU-HCMUS, Vietnam
Jianfeng Si	Institute for Infocomm Research, Singapore
Wen-Chih Peng	National Chiao Tung University, Taiwan

Local Organizing Committee

Tuyet-Trinh Vu	Hanoi University of Science and Technology, Vietnam
Hong-Phuong Nguyen	Hanoi University of Science and Technology, Vietnam
Van Thu Truong	Hanoi University of Science and Technology, Vietnam

Steering Committee Liaison

Stephane Bressan	National University of Singapore, Singapore

Webmaster

Viet-Trung Tran	Hanoi University of Science and Technology, Vietnam

Program Committees

Senior PC members

Ira Assent	Aarhus University, Denmark
Lei Chen	Hong Kong University of Science and Technology (HKUST), China
Reynold Cheng	University of Hong Kong, China
Gabriel Ghinita	University of Massachusetts Boston, USA
Panos Kalnis	King Abdullah University of Science and Technology, Saudi Arabia
Nikos Mamoulis	University of Hong Kong, China
Kyriakos Mouratidis	Singapore Management University, Singapore
Mario Nascimento	University of Alberta, Canada
Dimitris Papadias	Hong Kong University of Science and Technology (HKUST), China
Stavros Papadoupoulos	MIT, USA
Torben Bach Pedersen	Aalborg University, Denmark
Jian Pei	Simon Fraser University, Canada
Thomas Seidl	RWTH Aachen University, Germany
Timos Sellis	RMIT University, Australia
Raymond Wong	Hong Kong University of Science and Technology (HKUST), China

PC Members

Nikolaus Augsten	University of Salzburg, Austria
Spiridon Bakiras	City University of New York, USA
Zhifeng Bao	University of Tasmania, Australia
Srikanta Bedathur	IBM Research, Delhi, India
Ladjel Bellatreche	University of Poitiers, France
Boualem Benatallah	University of New South Wales, Australia
Bin Cui	Peking University, China
Athman Bouguettaya	Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Panagiotis Bouros	Humboldt-Universität zu Berlin, Germany
Selcuk Candan	Arizona State University, USA
Jianneng Cao	A*STAR, Singapore
Marco Casanova	Pontifical Catholic University of Rio de Janeiro, Brazil
Sharma Chakravarthy	University of Texas at Arlington, USA
Jae Chang	Chonbuk National University, Korea
Rui Chen	Hong Kong Baptist University, China
Yi Chen	New Jersey Institute of Technology, USA
James Cheng	The Chinese University of Hong Kong (CUHK), China
Gao Cong	Nanyang Technological University (NTU), Singapore

Ugur Demiryurek	University of Southern California (USC), USA
Prasad Deshpande	IBM Research, India
Gill Dobbie	University of Auckland, New Zealand
Eduard Dragut	Temple University, USA
Cristina Dutra de Aguiar Ciferri	Universidade de São Paulo, Brazil
Sameh Elnikety	Microsoft Research Redmond, USA
Tobias Emrich	Ludwig-Maximilians-Universität München, Germany
Johann Gamper	Free University of Bozen-Bolzano, Italy
Xin Gao	King Abdullah University of Science and Technology (KAUST), Saudi Arabia
Chenjuan Guo	Aarhus University, Denmark
Ralf Hartmut Güting	University of Hagen, Germany
Takahiro Hara	Osaka University, Japan
Haibo Hu	Hong Kong Baptist University, China
Yoshiharu Ishikawa	Nagoya University, Japan
Mizuho Iwaihara	Waseda University, Japan
Adam Jatowt	Kyoto University, Japan
Vana Kalogeraki	Athens University of Economy and Business, Greece
Panos Karras	Skoltech, Russia
Norio Katayama	National Institute of Informatics, Japan
Sang-Wook Kim	Hanyang University, Korea
Seon Ho Kim	University of Southern California (USC), USA
Hiroyuki Kitagawa	University of Tsukuba, Japan
Peer Kröger	Ludwig-Maximilians-Universität München, Germany
Jae-Gil Lee	Korea Advanced Institute of Science and Technology (KAIST), Korea
Wang-Chien Lee	Portland State University (PSU), USA
Sang-Goo Lee	Seoul National University, Korea
Hou Leong	University of Macau, China
Guoliang Li	Tsinghua University, China
Hui Li	Xidian University, China
Xiang Lian	University of Texas–Pan American (UTPA), USA
Lipyeow Lim	University of Hawaii, USA
Sebastian Link	University of Auckland, New Zealand
Bin Liu	NEC Laboratories, USA
Changbin Liu	AT & T, USA
Eric Lo	Hong Kong Polytechnic University, China
Jiaheng Lu	Renmin University of China, China
Qiong Luo	Hong Kong University of Science and Technology (HKUST), China
Matteo Magnani	Uppsala University, Sweden
Silviu Maniu	University of Hong Kong (HKU), China
Essam Mansour	Qatar Computing Research Institute, Qatar

Marco Mesiti	University of Milan, Italy
Yasuhiko Morimoto	Hiroshima University, Japan
Wilfred Ng	Hong Kong University of Science and Technology (HKUST), China
Makoto Onizuka	Osaka University, Japan
Balaji Palanisamy	University of Pittsburgh, USA
Stefano Paraboschi	Università degli Studi di Bergamo, Italy
Sanghyun Park	Yonsei University, Korea
Dhaval Patel	IIT Roorkee, India
Evaggelia Pitoura	University of Ioannina, Greece
Pascal Poncelet	Université Montpellier 2, France
Maya Ramanath	Indian Institute of Technology, New Delhi, India
Shazia Sadiq	University of Queensland, Australia
Sherif Sakr	University of New South Wales, Australia
Kai-Uwe Sattler	Ilmenau University of Technology, Germany
Peter Scheuermann	Northwestern University, USA
Markus Schneider	University of Florida, USA
Matthias Schubert	Ludwig-Maximilians-Universität München, Germany
Shuo Shang	China University of Petroleum, Beijing, China
Kyuseok Shim	Seoul National University, Korea
Junho Shim	Sookmyung Women's University, Korea
Shaoxu Song	Tsinghua University, China
Atsuhiro Takasu	National Institute of Informatics, Japan
Kian-Lee Tan	National University of Singapore (NUS), Singapore
Nan Tang	Qatar Computing Research Institute, Qatar
Martin Theobald	University of Antwerp, Belgium
Dimitri Theodoratos	New Jersey Institute of Technology, USA
James Thom	RMIT University, Australia
Wolf Tilo-Balke	University of Hannover, Germany
Hanghang Tong	City University of New York (CUNY), USA
Yongxin Tong	Hong Kong University of Science and Technology (HKUST), China
Kristian Torp	Aalborg University, Denmark
Goce Trajcevski	Northwestern University, USA
Vincent S. Tseng	National Cheng Kung University, Taiwan
Stratis Viglas	University of Edinburgh, UK
Wei Wang	University of New South Wales, Australia
Huayu Wu	Institute for Infocomm Research (I^2R), Singapore
Yinghui Wu	University of California, Santa Barbara (UCSB), USA
Xiaokui Xiao	Nanyang Technological University (NTU), Singapore
Xike Xie	Aalborg University, Denmark
Jianliang Xu	Hong Kong Baptist University, China
Bin Yang	Aalborg University, Denmark

Yin Yang	Advanced Digital Sciences Center, Singapore
Man-Lung Yiu	Hong Kong Polytechnic University, China
Haruo Yokota	Tokyo Institute of Technology, Japan
Jeffrey Yu	The Chinese University of Hong Kong (CUHK), China
Zhenjie Zhang	Advanced Digital Sciences Center (ADSC), Singapore
Xiuzhen Zhang	RMIT University, Australia
Kevin Zheng	University of Queensland, Australia
Wenchao Zhou	Georgetown University, USA
Bin Zhou	University of Maryland, Baltimore County, USA
Roger Zimmermann	National University of Singapore (NUS), Singapore
Lei Zou	Beijing University, China
Andreas Züfle	Ludwig-Maximilians-Universität München, Germany

External Reviewers

Yeonchan Ahn	Seoul National University, Korea
Cem Aksoy	New Jersey Institute of Technology, USA
Ibrahim Alabdulmohsin	King Abdullah University of Science and Technology, Saudi Arabia
Yoshitaka Arahori	Tokyo Institute of Technology, Japan
Zhuowei Bao	Facebook, USA
Thomas Behr	University of Hagen, Germany
Jianneng Cao	A*STAR, Singapore
Brice Chardin	LIAS/ISAE-ENSMA, France
Lei Chen	Hong Kong Baptist University, China
Jian Dai	The Chinese Academy of Sciences, China
Ananya Dass	New Jersey Institute of Technology, USA
Aggeliki Dimitriou	National Technical University of Athens, Greece
Zhaoan Dong	Renmin University of China, China
Hai Dong	RMIT University, Australia
Zoé Faget	LIAS/ISAE-ENSMA, France
Qiong Fang	Hong Kong University of Science and Technology (HKUST), China
ZiQiang Feng	Hong Kong Polytechnic University, China
Ming Gao	East China Normal University, China
Azadeh Ghari-Neiat	RMIT University, Australia
Zhian He	Hong Kong Polytechnic University, China
Yuzhen Huang	The Chinese University of Hong Kong, China
Stéphane Jean	LIAS/ISAE-ENSMA, France
Selma Khouri	LIAS/ISAE-ENSMA, France
Hanbit Lee	Seoul National University, Korea
Sang-Chul Lee	Carnegie Mellon University, USA

Feng Li	Microsoft Research, Redmond, USA
Yafei Li	Hong Kong Baptist University, China
Jinfeng Li	The Chinese University of Hong Kong, China
Xin Lin	East China Normal University, China
Yu Liu	Renmin University of China, China
Yi Lu	The Chinese University of Hong Kong, China
Nguyen Minh Luan	A*STAR, Singapore
Gerasimos Marketos	University of Piraeus, Greece
Jun Miyazaki	Tokyo Institute of Technology, Japan
Bin Mu	City University of New York, USA
Johannes Niedermayer	Ludwig-Maximilians-Universität München, Germany
Konstantinos Nikolopoulos	City University of New York, USA
Sungchan Park	Seoul National University, Korea
Youngki Park	Samsung Advanced Institute of Technology, Korea
Jianbin Qin	University of New South Wales, Australia
Kai Qin	RMIT University, Australia
Youhyun Shin	Seoul National University, Korea
Hiroaki Shiokawa	NTT Software Innovation Center, Japan
Masumi Shirakawa	Osaka University, Japan
Md. Anisuzzaman Siddique	Hiroshima University, Japan
Reza Soltanpoor	RMIT University, Australia
Yifang Sun	University of New South Wales, Australia
Erald Troja	City University of New York, USA
Fabio Valdés	University of Hagen, Germany
Jan Vosecky	Hong Kong University of Science and Technology (HKUST), China
Jim Jing-Yan Wang	King Abdullah University of Science and Technology, Saudi Arabia
Huanhuan Wu	The Chinese University of Hong Kong, China
Xiaoying Wu	Wuhan University, China
Chen Xu	Technische Universität Berlin, Germany
Jianqiu Xu	Nanjing University of Aeronautics and Astronautics, China
Takeshi Yamamuro	NTT Software Innovation Center, Japan
Da Yan	The Chinese University of Hong Kong, China
Fan Yang	The Chinese University of Hong Kong, China
Jongheum Yeon	Seoul National University, Korea
Seongwook Youn	University of Southern California, USA
Zhou Zhao	Hong Kong University of Science and Technology (HKUST), China
Xiaoling Zhou	University of New South Wales, Australia

Tutorials

Scalable Learning Technologies for Big Data Mining

Gerard de Melo[1] and Aparna S. Varde[2]

[1] Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China
gdm@demelo.org
[2] Department of Computer Science, Montclair State University, Montclair, NJ, USA
vardea@montclair.edu

Abstract. As data expands into big data, enhanced or entirely novel data mining algorithms often become necessary. The real value of big data is often only exposed when we can adequately mine and learn from it. We provide an overview of new scalable techniques for knowledge discovery. Our focus is on the areas of cloud data mining and machine learning, semi-supervised processing, and deep learning. We also give practical advice for choosing among different methods and discuss open research problems and concerns.

Keywords: Big Data · Cloud Data Mining · Deep Learning · Semi-Supervised Learning

1 Cloud Data Mining and Machine Learning

When moving from data to big data, we often need fundamentally different technologies for mining and learning. For data mining, one can rely on SQL over Big Data systems such as Hive, Cloudera's Impala, and Hortonworks' Stinger for efficient querying. One can also opt for more general processing frameworks like Scalding and Apache Spark, or stream-based alternatives like Apache Storm and Flink.

For cloud-based machine learning, Apache Mahout provides open source implementations of recommendation, clustering and classification, with MLBase and its MLLib component as a Spark-based alternative. Applications of this include news gathering, email classification, and recommender systems. Additionally, custom algorithms can be used for specific applications such as scalable rule mining and frequent item-set mining.

2 Semi-Supervised Processing and Deep Learning

Since training data is hard to obtain, semi-supervised learning and distant supervision can be used to exploit large amounts of unlabeled data, e.g. for sentiment analysis on Twitter. Additional signals for certain tasks can come from specific sources such as Wikipedia. Web-scale statistics, e.g. from Google's Web N-Grams, improve performance in co-reference analysis and information extraction.

Deep Learning and representation learning aim at learning multiple layers of abstraction from original features to capture more complex interactions. One can exploit unlabeled textual data to improve over bag-of-words feature representations or to create multimodal embeddings that combine words and images in the same space.

Overall, we see numerous new possibilities arising from the availability of Big Data in conjunction with such novel methods for exploiting it.

Acknowledgments. Gerard de Melo's research is in part by the National Basic Research Program of China Grants 2011CBA00300, 2011CBA00301, and NSFC Grants 61033001, 61361136003.

Large Scale Video Management Using Spatial Metadata and Their Applications

Seon Ho Kim[1] and Roger Zimmermann[2]

[1] Integrated Media Systems Center, Univ. of Southern California, Los Angeles, USA
seonkim@usc.edu
[2] Interactive & Digital Media Institute, National University of Singapore, Singapore
rogerz@comp.nus.edu.sg

1 Introduction

Recently, there has been research to utilize spatial metadata of videos for the management of large numbers of videos. For example, the location of a camera from GPS and the viewing direction from an embedded digital compass are captured at the recording time so that they form the field of view (FOV) model for the coverage of a viewable scene. These spatial metadata effectively convert the challenging video search problem to well known spatial database problems, which can greatly enhance the utility of collected videos in many emerging applications. However, there are still many open fundamental research questions in utilizing spatial metadata for data collection, indexing, searching, etc., especially in the presence of a large video dataset, as well as more sophisticated questions such as whether a video database search can accommodate human friendly views.

This tutorial covers and describes existing technologies and methods in harnessing spatial metadata of videos in spatial databases for video applications, especially for mobile videos from smartphones. The discussion includes types of metadata, their acquisition, indexing, searching, and potential applications of geo-tagged videos in scale. Furthermore, a spatial crowdsourcing of mobile media contents is discussed with regard to the use of geographical information in crowdsourcing. Some use cases of the implemented techniques are presented too. This tutorial can be a good survey and summary of the techniques which bridge the spatial database and videos. The target audience can be any researchers and practitioners who are interested in designing and managing a large video databases and Big Data in social media.

2 Tutorial Outline

Motivation: why are spatial metadata of videos important?
Types of Geospatial Metadata: types and methods of acquisition.
Modeling the Coverage Area of Scene: Field of View model and its use.
Extracting Textual Keywords Using Metadata: methods to extract.
Spatial Crowdsourcing: crowdsourcing media content.
Indexing and Searching: how to handle a large number of FOVs.
Applications: use cases and futuristics applications.
Open Discussion and Q&A

Querying Web Data

Andrea Calì

[1] Dept. of Computer Science
University of London, Birkbeck College
[2] Oxford-Man Institute of Quantitative Finance
University of Oxford
andrea@dcs.bbk.ac.uk

Abstract. A vast amount of information is available on the Web, not only in HTML static pages (documents), but also and especially in more structured form in databases accessible on the Web in different forms. In the Semantic Web [1], ontologies provide shared conceptual specifications of domains, so that the data can be enhanced with intensional semantic information to provide answers to queries that are consistent with an ontology. Ontological query answering [4] amounts to returning the answers to a query, that are logically entailed by the union of a data set (a set of membership assertions) and an ontology, where the latter is a set of logical assertions. We give an overview of the most important ontology formalisms for the Semantic Web, and we illustrate the most relevant query answering techniques, with particular emphasis on their efficiency [3]. Then we illustrate techniques specifically developed to access Hidden Web data, which can be accessed only according to specific patterns, therefore making query answering a complex and generally inefficient task [2]; we show techniques to improve efficiency by determining whether an access to a certain source is relevant to a given query. Finally we show techniques to query Linked Data sets, and to source relevant data to a certain query in such sets. We show how Linked Data can be employed in semantic search and recommender systems [5], by extracing ontological information from Linked Data sources.

Acknowledgments. The author acknowledges support by the EPSRC project "Logic-based Integration and Querying of Unindexed Data" (EP/E010865/1).

References

1. A. Artale, D. Calvanese, R. Kontchakov, and M. Zakharyaschev. The DL-Lite family and relations. *J. Artif. Intell. Res.*, 36:1–69, 2009.
2. M. Benedikt, G. Gottlob, and P. Senellart. Determining relevance of accesses at runtime. In *Proc. of PODS 2011*, pages 211–222.
3. A. Calì, G. Gottlob, and A. Pieris. Towards more expressive ontology languages: The query answering problem. *Artif. Intell.*, 193:87–128, 2012.
4. R. Kontchakov, C. Lutz, D. Toman, F. Wolter, and M. Zakharyaschev. The combined approach to ontology-based data access. In *Proc. of IJCAI*, pages 2656–2661, 2011.
5. V. C. Ostuni, T. D. Noia, E. D. Sciascio, and R. Mirizzi. Top-N recommendations from implicit feedback leveraging linked open data. In *Proc. of RecSys 2013*, pages 85–92.

Contents – Part II

Contents – Part I

Social Networks I

Information Integration and Data Quality

Information Retrieval and Summarization

Security and Privacy