02 a 05
de outubro 2017
Uberlândia - MG

SBBD
Proceedings
32nd Brazilian Symposium on Databases

Carmem S. Hara, Bernadette F. Lóscio,
Damires Y. S. Fernandes, Ana Carolina B. Salgado,
Humberto L. Razente, Maria Camila N. Barioni (Org.)

Sociedade Brasileira de Computação

SBBD2017

SBBD2017

32nd BRAZILIAN SYMPOSIUM ON DATABASES
October 2nd to 5th, 2017
Uberlândia – MG – Brazil

PROCEEDINGS

Promotion
Sociedade Brasileira de Computação – SBC
Comissão Especial de Banco de Dados (CEBD) da SBC

Organization
Universidade Federal de Uberlândia – UFU

Steering Committee Chair
Javam de Castro Machado, UFC

Local Organization Chairs
Maria Camila Nardini Barioni, UFU
Humberto Luiz Razente, UFU

Program Committee Chairs
Carmem S. Hara, UFPR
Bernadette F. Lóscio, UFPE
Damires Y. S. Fernandes IFPB

ISSN: 2316-5170

Dados Internacionais de Catalogação na Publicação (CIP)

B827s Brazilian Symposium on databases (SBBD) (32. : 2017 : Uberlândia, MG, Brazil)

Proceedings [do] 32nd Brazilian Symposium on databases (SBBD), October, 2nd a 5th, 2017, Uberlândia, Minas Gerais; organizadores: Universidade Federal de Uberlândia, Javam de Castro Machado. - Uberlândia: SBC, 2017.

ISSN: 2316-5170
Inclui bibliografia.
341 p.: il.
Modo de acesso: http://www.sbbd.org.br/2017

1. Computação - Congressos. 2. Bases de Dados - Congressos. I. Hara, Carmem Satie. II. Lóscio, Bernadete Farias. III. Brazilian Symposium on databases (SBBD) (32.: 2017 : Uberlândia, MG, Brazil) IV. Universidade Federal de Uberlândia. V. Sociedade Brasileira de Computação. VI. Título.

CDU: 681.3(061.3)

Message from the Local Organization Committee Chairs

Welcome to the 32nd Brazilian Symposium on Databases and to Uberlândia, Minas Gerais! The Brazilian Symposium on Databases is the official database event of the Brazilian Computer Society (SBC) and the largest venue in Latin America for presentation and discussion of research results in the database domain. The 32nd edition of the symposium (SBBD 2017) was held in Uberlândia, in the state of Minas Gerais, from October 2nd to October 5th, 2017. The local organization was performed by the Federal University of Uberlândia (UFU) through the Computing Faculty (FACOM). This year, for the first time, SBBD had the Brazilian Conference on Intelligent Systems (BRACIS) and the Symposium on Knowledge Discovery, Mining and Learning (KDMiLe) as co-located events providing a rich environment for the discussion of researches of their interrelated areas.

The SBBD 2017 program offers a variety of activities, suited for an audience ranging from undergraduate to Ph.D. students, database professionals, practitioners and researchers. The program includes: 3 invited talks and 3 tutorials, presented by distinguished speakers from Brazil, USA and France; 9 technical sessions; 3 short courses about hot topics in the area, presented by specialists in their research fields; demos and applications session; posters sessions; thesis and dissertations workshop; the biannual thesis and dissertations contest; 2 co-located workshops; the 1st KDD-BR (Brazilian Knowledge Discovery in Databases) competition; and a panel.

The excellence of SBBD 2017 program is the result of the competence and effort of a large community, which we gratefully acknowledge. The various sections of these proceedings list in detail those that contributed to the SBBD 2017 edition. We thank the symposium chairs and our colleagues of the local organization committee who donated their precious time to made SBBD 2017 a reality. We also thank the Computing Faculty (FACOM) of the Federal University of Uberlândia (UFU). We are also grateful to the SBC board for their support and to the steering committee members for their help, advice and support. Further, we thank the program committee members and external reviewers for the high quality reviews, and the authors who submitted their papers to SBBD 2017. Finally, we are grateful to our sponsors. Without their support we would not be able to organize this annual event that brings together our community.

We hope you all enjoy SBBD 2017 in Uberlândia, Minas Gerais!

Maria Camila Nardini Barioni, UFU
Humberto Luiz Razente, UFU
SBBD 2017 Local Organization Committee Chairs

Table of Contents

Editorial

It is a great pleasure to introduce the Proceedings of the Brazilian Symposium on Databases (SBBD) with the full and short papers accepted for presentation at the 32nd edition of the symposium. SBBD 2017 was held in Uberlândia, in the state of Minas Gerais, Brazil, from October 2nd to October 5th, 2017. It was organized by the Federal University of Uberlândia (UFU), and for the first time, SBBD was held in conjunction with the Brazilian Conference on Intelligent Systems (BRACIS) and the Symposium on Knowledge Discovery, Mining and Learning (KDMiLe).

SBBD is the official database event of the Brazilian Computer Society (SBC) and the largest venue in Latin America for presentation and discussion of research results in the databases domain. Along with technical sessions, SBBD includes invited talks, tutorials and short courses given by distinguished speakers from the national and international research communities. SBBD regularly promotes a demos and applications session, and a thesis and dissertations workshop as co-located events. This year, we also promoted the Thesis and Dissertation Contest, a biannually event started in 2015, in which the best Brazilian thesis and dissertations in the database area, defended in 2015 and 2016, were presented and competed for an award.

We also introduced some new activities to the event. For the first time, there were two co-located workshops: the Dataset Showcase Workshop, and the Databases meet Bioinformatics Workshop. Moreover, SBBD joined BRACIS and KDMile to promote the 1st Brazilian Knowledge Discovery in Databases competition (KDD-BR). The competition involved the classification of real images captured by a monitoring station in an astronomy observatory.

All papers presented in technical sessions during the event reported interesting results or proposed novel thought-provoking ideas in several subjects on the databases and related areas. For the 2017 edition, SBBD accepted five categories of submissions: JIDM articles, full papers, short papers, vision papers, and distinguished published papers.

Submissions to the JIDM category can be made throughout the year. The review process is conducted by the editorial board of JIDM, leaded by the editor-in-chief Caetano Traina Jr. Articles accepted by August 8th were invited to be presented at SBBD 2017. This year, only one JIDM article were presented during the event.

Full papers had two cycles of submissions: the first one with deadline in March and the second one with deadline in May. The review process involved a single review round, with a rebuttal phase. Authors were initially notified with the reviews and had a few days for answering the reviewers' comments during the rebuttal phase. After evaluating the rebuttal comments during the discussion period, a final decision was achieved. Out of 45 papers submitted to both the 1st and 2nd cycles, 15 were accepted as full papers (acceptance rate of 33%), and 2 as short papers.

Short and vision papers had a single cycle of submissions with deadline in July. There were no rebuttal phase. However, all program committee members were invited to participate in a discussion period to reach a decision. There were 38 submissions of short papers, out of which 17 were accepted (acceptance rate of 45%). There were no vision papers submissions.

Distinguished Published Papers is a new category of submissions introduced in SBBD 2017. It aimed to attract the best papers of the Brazilian community, published or accepted for publication by a first-class database conference, and give the authors the opportunity to present their work during the event. There were 2 submissions, but none of them were accepted by the SBBD Steering Committee, on the grounds of adherence to the call for papers.

The topics with more submissions among full and short papers, according to the author's selection from the topics of interest, were: Performance Evaluation and Benchmarking (18 submissions), NoSQL Databases (18 submissions), Data Analytics and Data Visualization (18 submissions), Algorithms and Techniques for Data Mining (14 submissions), Database Techniques to support Data Mining (9 submissions), Database Design and Data Semantics (9 submissions), Data on the Web (9 submissions), Information Retrieval Models and Techniques (8 submissions), and Data Management in Clouds (8 submissions).

JIDM articles, as well as full and short accepted papers were presented in technical sessions during the event. An invited paper has also been included in the technical program: "The collaboration network of the Brazilian Symposium on Databases". It analyzes SBBD's co-authorship network during its 30 years of history. The article has been published in the Journal of the Brazilian Computer Society, in its July 2017 issue.

The best full and short papers received award certificates during the event and will be invited to submit extended versions to JIDM. We have also awarded "Outstanding Reviewer Certificates" to the technical program committee members that excelled in the reviewing process. We do not have enough words to thank all committee members and external reviewers for their commitment and high quality reviews.

The Proceedings of SBBD are the result of the collective effort of a large community, which we gratefully acknowledge. We thank the SBBD 2017 local organization committee and its symposium chairs, who worked hard to guarantee an outstanding event. We are also grateful to the steering committee members for their help, advice and support. Lastly, but most importantly, we are grateful to the authors who submitted their work to SBBD 2017.

Carmem S. Hara, UFPR
SBBD 2017 Program Committee Chair

Bernadette Farias Lóscio, UFPE
Damires Yluska Souza Fernandes, IFPB
SBBD 2017 Short Papers Program Co-Chairs

32th Brazilian Symposium on Databases

full

October 2nd to 5th, 2017
Uberlândia – MG – Brazil

SBBD PROCEEDINGS

FULL PAPERS

Promotion

Brazilian Computer Society – SBC
SBC Special Interest Group on Databases

Organization

Universidade Federal de Uberlândia – UFU

Program Chair

Carmem Hara, UFPR

32nd Brazilian Symposium on Databases

October 2nd to 5th, 2017
Uberlândia – MG – Brazil

Promotion

Brazilian Computer Society – SBC
SBC Special Interest Group on Databases

Organization

Universidade Federal de Uberlândia – UFU

SBBD Steering Commitee

Agma Juci Machado Traina, USP
Bernadette Lóscio, UFPE
Caetano Traina Jr., USP
Carmem Hara, UFPR
Javam Machado, UFC
Mirella M. Moro, UFMG
Vanessa Braganholo, UFF

SBBD 2017 Commitee

Steering Committee Chair
Javam Machado, UFC

Local Organization Chairs
Maria Camila N. Barioni, UFU and Humberto L. Razente, UFU

Program Committee Chair
Carmem S. Hara, UFPR

Short papers Chairs
Bernadette Lóscio, UFPE and Damires Souza, IFPB

Demos and Applications Session Chair
Daniel de Oliveira, UFF

Short Courses Chair
Vaninha Vieira, UFBA

Workshop on Thesis and Dissertations in Databases Chair
Carina Dorneles, UFSC

Tutorials Chair
Ana Carolina Salgado, UFPE

Thesis and Dissertation Contest Chair
Vânia Vidal, UFC

Workshops Chair
Fernanda Baião (UNIRIO)

Local Organization Committee

Maria Camila N. Barioni, UFU
Humberto L. Razente, UFU
José Gustavo de Souza Paiva, UFU
Marcelo Zanchetta do Nascimento, UFU
Elaine Ribeiro de Faria Paiva, UFU
João Henrique de Souza Pereira, UFU

Full Papers Program Committee

Agma Traina, Universidade de São Paulo (ICMC/USP)
Alberto Laender, Universidade Federal de Minas Gerais (UFMG)
Alexandre Plastino, Universidade Federal Fluminense (UFF)
Altigran Soares da Silva, Universidade Federal do Amazonas (UFAM)
Ana Carolina Salgado, Universidade Federal de Pernambuco (UFPE)
André Santanchè, Universidade Estadual de Campinas (Unicamp)
Angelo Brayner, Universidade Federal do Ceará (UFC)
Bernadette Loscio, Universidade Federal de Pernambuco (UFPE)
Caetano Traina Júnior, Universidade de São Paulo (ICMC/USP)
Carina F. Dorneles, Universidade Federal de Santa Catarina (UFSC)
Carmem Hara, Universidade Federal do Paraná (UFPR), Chair
Celso Hirata, Instituto Tecnológico de Aeronáutica (ITA)
Clodoveu Davis, Universidade Federal de Minas Gerais (UFMG)
Cristina Ciferri, Universidade de São Paulo (ICMC/USP)
Daniel de Oliveira, Universidade Federal Fluminense (UFF)
Daniel Kaster, Universidade Estadual de Londrina (UEL)
Divyakant Agrawal, University of California, Santa Barbara (UCSB)
Duncan Ruiz, Pontifícia Universidade Católica do Rio Grande do Sul (PUC/RS)
Edleno Moura, Universidade Federal do Amazonas (UFAM)
Eduardo Ogasawara, Centro Federal de Educação Tecnológica Celso Suckow da Fon-

seca (CEFET/RJ)
Elaine Sousa, Universidade de São Paulo (USP)
Eli Cortez, Microsoft
Fabio Porto, Laboratório Nacional de Computação Científica (LNCC)
Fernanda Baião, Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Humberto Razente, Universidade Federal de Uberlândia (UFU)
Javam Machado, Universidade Federal do Ceará (UFC)
João Eduardo Ferreira, Universidade de São Paulo (USP)
Jonice Oliveira, Universidade Federal do Rio de Janeiro (DCC/IM/UFRJ)
José Palazzo Moreira de Oliveira, Universidade Federal do Rio Grande do Sul (UFRGS)
José Antonio Macêdo, Universidade Federal do Ceará (UFC)
Karin Becker, Universidade Federal do Rio Grande do Sul (UFRGS)
Marco Antonio Casanova, Pontifícia Universidade Católica do Rio de Janeiro (PUC/Rio)
Marcos Gonçalves, Universidade Federal de Minas Gerais (UFMG)
Maria Camila Nardini Barioni, Universidade Federal de Uberlândia (UFU)
Maria Claudia Cavalcanti, Instituto Militar de Engenharia (IME)
Marta Mattoso, Universidade Federal do Rio de Janeiro (COPPE/UFRJ)
Mirella M. Moro, Universidade Federal de Minas Gerais (UFMG)
Mirian Halfeld-Ferrari, Université d'Orleans (LIFO)
Pedro Eugenio Rocha Pedreira, Facebook Inc.
Renata Galante, Universidade Federal do Rio Grande do Sul (UFRGS)
Renato Fileto, Universidade Federal de Santa Catarina (UFSC)
Ricardo Ciferri, Universidade Federal de São Carlos (UFSCar)
Ricardo Torres, Universidade Estadual de Campinas (Unicamp)
Ronaldo Mello, Universidade Federal de Santa Catarina (UFSC)
Sergio Lifschitz, Pontifícia Universidade Católica do Rio de Janeiro (PUC/Rio)
Valéria C. Times, Universidade Federal de Pernambuco (UFPE)
Vanessa Braganholo, Universidade Federal Fluminense (UFF)
Vania Bogorny, Universidade Federal de Santa Catarina (UFSC)
Vaninha Vieira, Universidade Federal da Bahia (UFBA)
Viviane Moreira, Universidade Federal do Rio Grande do Sul (UFRGS)
Wagner Meira Jr., Universidade Federal de Minas Gerais (UFMG)
Wellington Martins, Universidade Federal de Goiás (UFG)

External Reviewers

Anderson Chaves Carniel (ICMC-USP)
Eduardo Gonçalves (ENCE-IBGE)
Flavio R. C. Sousa (UFC)
Leonardo Moreira (UFC)
Mariana Tasca (CTSMART)
Rebeca Schroeder Freitas (UDESC)
Troy Kohwalter (UFF)

Table of Contents (JIDM Papers)

Table of Contents (Invited Papers)

Table of Contents (Full Papers)

32th Brazilian Symposium on Databases

short

October 2nd to 5th, 2017
Uberlândia – MG – Brazil

SBBD PROCEEDINGS

SHORT PAPERS

Promotion

Brazilian Computer Society – SBC
SBC Special Interest Group on Databases

Organization

Universidade Federal de Uberlândia – UFU

Program Chairs

Bernadette Farias Lóscio, UFPE
Damires Yluska Souza Fernandes, IFPB

32nd Brazilian Symposium on Databases

October 2nd to 5th, 2017
Uberlândia – MG – Brazil

Promotion

Brazilian Computer Society – SBC
SBC Special Interest Group on Databases

Organization

Universidade Federal de Uberlândia – UFU

SBBD Steering Commitee

Agma Juci Machado Traina, USP
Bernadette Lóscio, UFPE
Caetano Traina Jr., USP
Carmem Hara, UFPR
Javam Machado, UFC
Mirella M. Moro, UFMG
Vanessa Braganholo, UFF

SBBD 2017 Commitee

Steering Committee Chair

Javam Machado, UFC

Local Organization Chairs

Maria Camila N. Barioni, UFU and Humberto L. Razente, UFU

Program Committee Chair

Carmem S. Hara, UFPR

Short papers Chairs

Bernadette Lóscio, UFPE and Damires Souza, IFPB

Demos and Applications Session Chair

Daniel de Oliveira, UFF

Short Courses Chair
Vaninha Vieira, UFBA

Workshop on Thesis and Dissertations in Databases Chair
Carina Dorneles, UFSC

Tutorials Chair
Ana Carolina Salgado, UFPE

Thesis and Dissertation Contest Chair
Vânia Vidal, UFC

Workshops Chair
Fernanda Baião (UNIRIO)

Local Organization Committee

Maria Camila N. Barioni, UFU
Humberto L. Razente, UFU
José Gustavo de Souza Paiva, UFU
Marcelo Zanchetta do Nascimento, UFU
Elaine Ribeiro de Faria Paiva, UFU
João Henrique de Souza Pereira, UFU

Short Papers Program Committee

Alessandreia Oliveira, UFJF
Altigran Soares da Silva, UFAM
Ana Carolina Almeida, UERJ
Anderson Ferreira, UFOP
Angelo Brayner, UFC
Carina F. Dorneles, UFSC
Carlos Eduardo Pires, UFCG
Carmem Hara, UFPR
Celso Hirata, ITA
Clodoveu Davis, UFMG
Cristiano Cervi, UPF
Damires Souza, IFPB
Daniel de Oliveira, UFFS
Daniel Kaster, UEL
Daniel Notari, UCS
Daniela Barreiro Claro, UFBA
Deise Saccol, UFSM
Denio Duarte, UFFS
Duncan Ruiz, PUCRS
Eduardo de Almeida, UFPR

Eduardo Ogasawara, CEFET/RJ
Elaine Sousa, USP
Eveline Sacramento, FUNCEME
Fabio Porto, LNCC
Fernanda Baião, UNIRIO
Flávio R. C. Sousa, UFC
Helena Ribeiro, UCS
Humberto Razente, UFU
João B. Rocha-Junior, UEFS
Jonice Oliveira, DCC/IM/UFRJ
José Palazzo Moreira de Oliveira, UFRGS
José Antonio Macêdo, UFC
José de Aguiar Moraes Filho, UNIFOR
José Maria Monteiro, UFC
Karin Becker, UFRGS
Kelly Braghetto, IME/USP
Luiz Celso Gomes Jr, UTFPR
Marco Antonio Casanova, PUC-Rio
Maria Camila Nardini Barioni, UFU
Maria Claudia Cavalcanti, IME
Marilde Santos, UFSCar
Maristela Holanda, UnB
Mirella M. Moro, UFMG
Moisés Carvalho, UFAM
Pedro Eugenio Rocha Pedreira, Facebook Inc.
Raquel Stasiu, PUCPR/Universidade Tecnológica Federal do Paraná
Raqueline Penteado, UEM
Rebeca Schroeder, UDESC
Renato Fileto, UFSC
Robson Cordeiro, ICMC - USP
Robson Fidalgo, UFPE
Ronaldo Mello, UFSC
Sergio Lifschitz, PUC-Rio
Sergio Mergen, UFSM
Thiago Silva, UTFPR
Ticiana Coelho da Silva, UFC
Valéria C. Times, UFPE
Vania Bogorny, UFSC
Vaninha Vieira, UFBA
Wagner Meira Jr., UFMG
Wellington Martins, UFG

External Reviewers

Carlos Teles, CEFET/RJ
Christian Bones, USP
Demetrio Mestre, UFCG
Flavio Carvalho, CEFET-RJ

Leonardo Moreira, UFC
Marcelo Iury S . Oliveira, UFRPE

Table of Contents (Short Papers)

32th Brazilian Symposium on Databases

tutorials

October 2nd to 5th, 2017
Uberlândia – MG – Brazil

TUTORIALS

Promotion

Brazilian Computer Society – SBC
SBC Special Interest Group on Databases

Organization

Universidade Federal de Uberlândia – UFU

Tutorials Chair

Ana Carolina Salgado, UFPE

32nd Brazilian Symposium on Databases

October 2nd to 5th, 2017
Uberlândia – MG – Brazil

Promotion

Brazilian Computer Society – SBC
SBC Special Interest Group on Databases

Organization

Universidade Federal de Uberlândia – UFU

SBBD Steering Commitee

Agma Juci Machado Traina, USP
Bernadette Lóscio, UFPE
Caetano Traina Jr., USP
Carmem Hara, UFPR
Javam Machado, UFC
Mirella M. Moro, UFMG
Vanessa Braganholo, UFF

SBBD 2017 Commitee

Steering Committee Chair
Javam Machado, UFC

Local Organization Chairs
Maria Camila N. Barioni, UFU and Humberto L. Razente, UFU

Program Committee Chair
Carmem S. Hara, UFPR

Short papers Chairs
Bernadette Lóscio, UFPE and Damires Souza, IFPB

Demos and Applications Session Chair
Daniel de Oliveira, UFF

Short Courses Chair
Vaninha Vieira, UFBA

Workshop on Thesis and Dissertations in Databases Chair
Carina Dorneles, UFSC

Tutorials Chair
Ana Carolina Salgado, UFPE

Thesis and Dissertation Contest Chair
Vânia Vidal, UFC

Workshops Chair
Fernanda Baião (UNIRIO)

Local Organization Committee

Maria Camila N. Barioni, UFU
Humberto L. Razente, UFU
José Gustavo de Souza Paiva, UFU
Marcelo Zanchetta do Nascimento, UFU
Elaine Ribeiro de Faria Paiva, UFU
João Henrique de Souza Pereira, UFU

Tutorials Program Committee

Agma Traina (ICMC - USP)
Caetano Traina Jr. (ICMC - USP)
Carmem S. Hara (UFPR)
Vanessa Braganholo (UFF)

Editorial

Tutorials at the Brazilian Symposium on Databases (SBBD) have the goal to present introductory and advanced discussions on topics within the area of databases. Introductory tutorials target an audience consisting of advanced undergraduate and graduate students, as well as attendees from industry. Advanced tutorials, on the other hand, cover a state-of-the-art topic, motivating and exposing potential research paths.

The three accepted tutorials this year are related to relevant nowadays topics. The first one is entitled "Mobile Crowdsourcing for Smart City Applications" and will discuss mobile crowdsourcing issues for smart city applications, including topics like data quality, data analytics and trust issues. It will be presented by Sanjay Madria, a full tenured Professor at the Missouri University of Science and Technology.

The second tutorial, "Adaptivity in Database Kernels", will be presented by Javam Machado (Full Professor), Paulo Amora (MSc. student) and Elvis Teixeira (MSc. student) from Federal University of Ceara (UFC). They will talk about adaptivity, i.e., problems related to database physical design optimization for scenarios where the workload is unknown and immediate availability is a requirement. Adaptivity can also be applied on data storage layout in order to optimize relevant data exchange between memory hierarchy layers.

The third tutorial is entitled "Social Professional Networks: Taxonomy, Metrics and Analyses of Relationship Strength". It will present a deeper understanding of the social professional networks types, definitions, features, analysis and applications while providing a useful taxonomy about their use. It also discusses the strength of ties, a central aspect that allows studying the roles of relationships. This tutorial will be presented by Mirella Moro (Associate Professor) and Michele Brandão ("professor substituto") from the Computer Science department at UFMG.

This year we had six excellent submissions and I would like to thank all the authors of submitted proposals. Also, I would like to invite all of you to attend and take advantage of the selected tutorials.

Ana Carolina Salgado, UFPE
SBBD 2017 Tutorials Chair

Table of Contents (Tutorials)

Mobile Crowdsourcing for Smart City Applications

Sanjay Kumar Madria
Professor, Department of Computer Science and Director, Web and Wireless Computing Lab, Missouri Institute of Science and Technology, USA
E-mail: madrias@mst.edu

Abstract: This tutorial will discuss mobile crowdsourcing issues for smart city applications. In particular, it will motivate the use of incentives for better crowd participation from people carrying mobile devices, data quality, data analytics and trust issues among others. It will also discuss some open research problems in that domain.

Introduction: Ever-increasing prevalence of social networking using mobile devices has catalyzed the growth of interesting and innovative new-age mobile crowdsourcing applications, which work at the intersection of human-centric computation (e.g., economic incentive management and social computing) and dynamic management of information and content in wireless networks. The prevalence and proliferation of mobile devices coupled with popularity of social media and increasingly technology-savvy users have fuelled the growth of mobile crowdsourcing and participatory sensing. In particular, participatory sensing can occur in various ways by means of devices (e.g., mobile phones, PDAs, laptops and various types of sensors) or by including humans in the loop or both. Notably, participatory sensing can also potentially act as a key enabling technology for various applications involving smarter cities initiatives.

Incidentally, large-scale collection of city-related event data is crucial to effective planning and decision-making for improving city management. Examples of city-related event data include traffic congestion, illegal parking, accidents, dysfunctional streetlights, broken pavements, potholes, planned road construction works, public rallies, waterlogging, uncleared garbage, and the like. Notably, existing sensor-based data collection mechanisms cannot always take human judgment and the context of the event into consideration, and the costs of deploying them across all city locations would be prohibitively expensive. Hence, event data collection by users can be used to complement sensor-based data collection. Observe that a vast majority of the users can be reasonably expected to carry mobile devices. Since mobile devices often come equipped with various kinds of sensors, resident-driven data collection is also well-aligned with current technological trends. However, note that incentives need to be provided to users for encouraging them to contribute event data.

In this environment, research issues include incentives & economic models for crowd participation, large-scale data management, resource discovery, replica allocation and consistency, mobile cloud, analytics on the collected data, resource access, indexing & query processing, mobile computing, and trust. These issues have generated a significant amount of interest in academia as well as in industry. We will discuss some of the open research issues in this area and provide our perspectives on those issues. This tutorial intends to foster discussions on the key research challenges as well as the design issues of key enabling technologies that need to be addressed to make scalable next-generation mobile crowdsourcing and human computation effective in a real-world application.

1. Mobile Crowd sourcing architectures; Mobile P2P, Mobile Cloud, etc.
2. Incentive Models for Mobile Crowdsourcing
3. Dynamic data management and querying in Crowdsourcing
4. Sensor data management for smart city applications
5. Content management in smart transportation
6. Privacy of Trajectory Data for Smart Transportation

References

Ilarri, Sergio, Ouri Wolfson, and Thierry Delot. "Collaborative sensing for urban transportation." IEEE Data Engineering Bulletin 37 (2014): 3-14.

Nilesh Padhariya, Anirban Mondal, Sanjay Madria, Efficient Processing of Mobile Crowdsourcing Queries with Multiple Sub-tasks for Facilitating Smart Cities, in SmartCities workshop with ACM Middleware, 2016, Italy.

Nilesh Padhariya, Anirban Mondal, Sanjay Kumar Madria: Top-k query processing in mobile-P2P networks using economic incentive schemes. Peer-to-Peer Networking and Applications 9(4): 731-751 (2016)

Nilesh Padhariya, Ouri Wolfson, Anirban Mondal, Varun Gandhi, Sanjay Kumar Madria: E-VeT: Economic Reward/Penalty-Based System for Vehicular Traffic Management. MDM (1) 2014: 99-102

Dejun Yang, Guoliang (Larry) Xue, Xi Fang and Jian Tang, Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing, Conference: Proceedings of the 18th annual international conference on Mobile computing and networking, 2012.

Shuo Ma, Yu Zheng, Ouri Wolfson: Real-Time City-Scale Taxi Ridesharing. IEEE Trans. Knowl. Data Eng. 27(7): 1782-1795 (2015)

Y. Wang, X. Jia, Q. Jin, and J. Ma, Mobile crowdsourcing: Architecture, applications, and challenges, Concurrency and Computation: Practice and Experience, 2016.

Kan Yang, Kuan Zhang, Ju Ren, and Xuemin Shen. Security and privacy in mobile crowdsourcing networks: challenges and opportunities. Communications Magazine, IEEE, 53(8):75–81, 2015.

Katrina Ward, Dan Lin and Sanjay Madria, "MELT: Mapreduce-based Efficient Large-scale Trajectory Anonymization", to appear in 29th International Conference on Scientific and Statistical Database Management (SSDBM, 2017), USA.

Biography: Sanjay Kumar Madria received his Ph.D. in Computer Science from Indian Institute of Technology, Delhi, India in 1995. He is a full tenured Professor, Department of Computer Science, at the Missouri University of Science and Technology (formerly, University of Missouri-Rolla), USA. Earlier he was Visiting Assistant Professor in the Department of Computer Science, Purdue University, West Lafayette, USA. He has published more than 230 Journal and conference papers in the areas of mobile computing, sensor networks, security, cloud computing, etc. He won five IEEE best paper awards in conferences including IEEE SRDS 2015, IEEE MDM in 2011 and 2012. He co-authored a book entitled "Web Data Management: A Warehouse Approach" published by Springer-verlag. He guest edited WWW Journal, several Data and Knowledge Engineering Sp. Issues on Web data management and Data warehousing. He was founding Program Chair for EC&WEB conference series. He served as a general co-chair of Mobile Data Management conference in 2010, IEEE Symposium on Reliability in Distributed Systems in 2012 and PC co-chair of MDM 2015. He serves in steering committees of IEEE SRDS and IEEE MDM. He is serving/served as PC member of various conferences such as VLDB, MDM, CIKM, ICDCS, and reviewer for many reputed journals such as IEEE TKDE, IEEE Computer, ACM Internet Computing, IEEE TMC etc. Dr. Madria has given tutorials on mobile data management in many international conferences like Middleware, MDM and SRDS. He is regular invited panelist in NSF, NSERC (Canada), Hong Kong Research Council and Sweden Council of Research. He received UMR faculty excellence award in 2007, 2009, 2011, 2013 and 2015, Japanese Society for Promotion of Science invitational fellowship in 2006, and Air Force Research Lab's visiting faculty fellowship from 2008 to 2016. He was honored with NRC fellowship in 2012-2013 by National Academies. His research is supported by multiple grants from several agencies such as NSF, DOE, NIST, AFRL, ARL, UM research board and from industries such as Boeing. He is IEEE Senior Member, served as IEEE Distinguished speaker and currently he is a speaker under ACM Distinguished Visitor program. He is ACM Distinguished Scientist and IEEE Golden Core Awardee.

Adaptivity in Database Kernels

Javam Machado[1], **Elvis Teixeira**[1], **Paulo Amora**[1]

[1]LSBD/DC – Universidade Federal do Ceará (UFC)
Fortaleza – CE – Brazil

{javam.machado,elvis.teixeira,paulo.amora}@lsbd.ufc.br

Abstract

Adaptivity addresses a class of problems related to database physical design optimization for scenarios where the workload is unknown and immediate availability is a requirement. The general strategy is to improve physical design by means of incremental changes, each guided by the current workload request. For instance, adaptive indexing builds partial indexes through steps during query processing rather than building full indexes. Adaptivity can also be applied on data storage layout in order to optimize relevant data exchange between memory hierarchy layers. Instead of having a fixed layout, adaptive storage redesigns data organization to answer incoming queries incrementally based on the current requests or recent workload pattern.

Introduction

Analytical processing of large datasets is currently a daily task for many business and for the scientific community. For these tasks it is often the case that users and data scientists do not have a well defined set of queries and data modification procedures, that is, the workload is not known or modeled. This scenario is typical of data exploration environments [Idreos et al. 2015] for data mining or scientific work found in industry or academia.

Another issue on modern data exploration that can be addressed with adaptive approaches is the treatment of hot and cold data. Hot data is characterized as data which is being currently used or accessed by insert and update commands, as well as recent data, which is more suitable to suffer those changes, is usually the data present in an OLTP database system. Cold data is defined as historical, usually accessed by aggregation queries and used for information extraction, like the data accessed by an OLAP database system and stored in data warehouses. So, the more recent the data is, the more valuable it is. This means that the amount of time it takes to load the dataset and optimize its physical layout is a concern, and that traditional database system design is not well suited.

New database setups for such tasks are needed for two main reasons: first, schema, physical design tuning and index selection and creation take time. Second, even if there is time available for database tuning the lack of knowledge of the workload turns hard problems like data model and index selection into even greater challenges since an inappropriate set of indexes or a storage layout unfit for the access patterns provide performance well below optimal, rendering data exploration impractical.

Adaptive systems address both issues by making layout operations less atomic so that they are able to change direction if required by the actual requests, thus optimizing the access to data that is currently being requested. Initially the system is made available to process user queries as soon as new data arrives [Idreos et al. 2007] with a physical design composed of simple and plain data structures and without any optimization. Again, this decision is made based on the premises that immediate availability is desired and no knowledge is available to lead to good decisions on how to tune the physical design.

This tutorial touches the most recent literature on adaptivity in various aspects of database architecture discussing and comparing their strengths and weaknesses, and presents the architecture of some well established adaptive indexing and storage techniques. Focus will be directed towards the lower levels of a DBMS, that is, the storage and data access layers. Additionally, storage and data access is where most of the important recent developments occurred and pioneering works on adaptivity in higher level database components, such as transaction management [Graefe et al. 2014] will be presented as guidelines for valuable future work.

Adaptive Storage

The way data is stored is usually defined by the system administrator in a manual process, with the objective of optimizing query response times. This data organization takes into account the application domain, how that data is accessed, and other factors. These constraints have one common factor: they all necessitate previous knowledge to be fulfilled, be it about query workload or semantic relationships between attributes of a schema. The DBA is assumed to have this knowledge, but most database systems rely on a static data layout. Therefore, adapting these layouts as queries are executed to allow optimal data organization without previous knowledge is a desirable feature, enabling the database to answer queries more quickly, without external interference and without requiring any knowledge about the data or how it is accessed.

In this tutorial we focus on three adaptive storage approaches: PelotonDB [Arulraj et al. 2016], H_2O [Alagiannis et al. 2014] and HyPer [Lang et al. 2016].

H_2O is a OLAP storage system which leverages queries access patterns to reorganize data layouts, incurring in several layouts to the same data, each optimized to a different access pattern. This self-designing approach requires different physical operators to access the different data organizations, and this is done by creating these operators on-the-fly, through code templates.

PelotonDB is a HTAP (Hybrid Transactional Analytic Processing) oriented database that fragments data into horizontal and vertical partitions called physical tiles as data becomes colder, for example, when it ceases to be accessed in a transactional way. A table can keep several data layouts, called tile groups. Data access is done by a single execution engine, through the abstraction of logical tiles, which contain pointers to the data in physical tiles. It also employs special techniques for recovery and logging aimed at novel Non-Volatile Memory hardwares.

HyPer is another HTAP oriented database that provides hardware manipulation techniques. The main process handles the OLTP workload and when a OLAP workload shows up, the main process forks itself. The forked process then have access to the data through a virtual memory snapshot. At first, data was categorized within four categories, Hot, Cooling, Cold and Frozen. Hot, Cooling and Cold items had only conceptual differences among them, all of them were stored in small memory pages and were uncompressed. The Frozen items, however, were compressed and stored in huge pages. To track the temperature of items, special in-memory flags, such as *dirty* and *young* are manipulated by the database, overriding the OS, through a special memory allocation. Nowadays, HyPer proposes a structure named Data Blocks, which is based on the same principle of splitting data in hot, uncompressed parts and cold, immutable, compressed parts. It makes use of Positioned Small Materialized Aggregates (PSMA) as an index to speed up scan operations. By having a maximum and a minimum for each block, it can save *scan* operations on blocks where the value searched isn't present for sure.

Adaptive Indexing

Adaptive indexing came up as the first promise that a workload-driven, self-tuning database system architecture is possible. The fundamental idea is tuning the access structures to provide faster response times for key ranges which have already been queried, and its vicinity, by indexing them at query processing time. The underlying assumption is that some key ranges will be of more interest than others so indexing them only when queried would be better since it is faster than building full indexes and the focus of indexing can change as soon as the query focus changes.

The first technique to be extensively tested was Database Cracking [Idreos et al. 2007], an approach that applies an incremental quicksort-like partitioning scheme to optimize the access for key ranges which are being focus of current query attention. This technique was conceived and implemented in the context of Column Store database systems. The process on indexing consists of choosing the boundaries of range query predicates as pivots for the next partition, so the expected behavior of the response time is to decrease from that of a linear scan towards that of a full index scan.

Adaptive Merging [Graefe and Kuno 2010b] is other distinctive technique focused on traditional tuple-based storage layouts, and it harnesses the well known behavior of B+-trees or, more specifically, partitioned B+-trees for index data structure. The first query instructs the system about index selection, the heuristic used is creating the index for the attribute that mostly benefits this first query. The process of index building starts in this first query by performing a full linear scan in the table and sorting it piece by piece in main memory. After the first query is done, the data will be again in disk but in the form of sorted runs ready to be turned into the leaves of the tree by an external merge sort process. Subsequent queries are processed by searching within each run and merging their key range into the final structure.

The main advantage of adaptive merging is the fact that its underlying data structure is cache friendly, that is, units of data storage and transfer are naturally paged, as in any B-tree related structure. This is ideal for databases stored in block devices like hard disks or flash drives, or even for the exchange of data between higher cache hierarchies like main memory and processor caches. Further developments that point out ways to go beyond adaptivity [Halim et al. 2012] [Petraki et al. 2015] Will be briefly touched in the tutorial.

About the authors:

Javam C. Machado is a full professor at the Federal University of Ceara (UFC), Brazil. He obtained a MsC in Computer Science from the Federal University of Rio Grande do Sul, Brazil and a PhD degree in Computer Science from the University of Grenoble, France. For 8 years, he was the manager of the UFC's IT infrastructure and, in 2011, he became the vice-director of the College of Science at the same University. Since 1995, he has coordinated research projects on the area of Database and Distributed Systems and he has advised MsC and PhD candidates. Javam is the coordinator of the Database Special Committee of the Brazilian Computer Society (2017) and he was the chair of the 2016 edition of the Brazilian Databases Systems Symposium. He has served as a member of the program committee of different conferences on database and on cloud computing. In 2010 Javam founded the Databases and Systems Laboratory (LSBD) of the Computer Science Department which he coordinates since than. Lately he is interested more particularly in data management and data privacy but also in adaptive database systems and cloud computing.

Paulo Roberto Pessoa Amora is a Computer Science MSc. student at UFC - Federal University of Ceará under Prof. Javam Machado. He is a member of the Databases and Systems Laboratory (LSBD) currently working on adaptive storages in main-memory databases. Paulo earned a BSc. in Computer Engineering at IFCE - Instituto Federal do Ceará with a sandwich year at University of Pittsburgh. He has interest in the following research topics: Adaptive Databases, Main-Memory Databases, Storage management.

Elvis Marques Teixeira is a Computer Science MSc. student at UFC - Federal University of Ceará under Prof. Javam Machado. He is also a member of the Databases and Systems Laboratory (LSBD) currently working on adaptive indexing techniques. Elvis earned a BSc. in Physics at UFC - Federal University of Ceará, meanwhile conducted research on mineralogy data processing. He has interest in the following research topics: Adaptive Indexes, data analysis and applications of machine learning on the design and adaptation of data access methods.

Referências

Alagiannis, I., Idreos, S., and Ailamaki, A. (2014). H2O: a hands-free adaptive store - read. In *International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, 2014*, pages 1103–1114.

Arulraj, J., Pavlo, A., and Menon, P. (2016). Bridging the archipelago between row-stores and column-stores for hybrid workloads. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, 2016*, pages 583–598.

Graefe, G., Halim, F., Idreos, S., Kuno, H. A., Manegold, S., and Seeger, B. (2014). Transactional support for adaptive indexing. *VLDB J.*, 23(2):303–328.

Graefe, G. and Kuno, H. A. (2010a). Adaptive indexing for relational keys. In *Workshops Proceedings of the 26th International Conference on Data Engineering, ICDE 2010, Long Beach, California, USA*, pages 69–74.

Graefe, G. and Kuno, H. A. (2010b). Self-selecting, self-tuning, incrementally optimized indexes. In *EDBT 2010, 13th International Conference on Extending Database Technology, Lausanne, Switzerland, 2010, Proceedings*, pages 371–381.

Halim, F., Idreos, S., Karras, P., and Yap, R. H. C. (2012). Stochastic database cracking: Towards robust adaptive indexing in main-memory column-stores. *PVLDB*, 5(6):502–513.

Idreos, S., Kersten, M. L., and Manegold, S. (2007). Database cracking. In *CIDR 2007, Third Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, 2007, Online Proceedings*, pages 68–78.

Idreos, S., Papaemmanouil, O., and Chaudhuri, S. (2015). Overview of data exploration techniques. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, 2015*, pages 277–281.

Lang, H., Mühlbauer, T., Funke, F., Boncz, P. A., Neumann, T., and Kemper, A. (2016). Data blocks: Hybrid OLTP and OLAP on compressed storage using both vectorization and compilation. In *SIGMOD Conference*, pages 311–326. ACM.

Petraki, E., Idreos, S., and Manegold, S. (2015). Holistic indexing in main-memory column-stores. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, 2015*, pages 1153–1166.

Social Professional Networks:

Taxonomy, Metrics and Analyses of Relationship Strength

Michele A. Brandão, UFMG
micheleabrandao@dcc.ufmg.br
www.dcc.ufmg.br/~micheleabrandao

Mirella M. Moro, UFMG
mirella@dcc.ufmg.br
www.dcc.ufmg.br/~mirella

Abstract

Social professional networks provide features not available in other networks. For example, LinkedIn facilitate professional networking, and GitHub enables committing and sharing code. Such networks also provide data on users, behaviors and interactions. Here, we foster a deeper understanding of the social professional networks types, definitions, features, analyses and applications while providing a useful taxonomy about their use. We also study the strength of ties, a central aspect that allows studying the roles of relationships. Therefore, besides analyzing the strength of co-authorship ties, we also present a set of metrics and algorithms to measure such strength in different contexts.

Outline

1. Introduction

Social Networks Analysis has evolved from a Social Sciences area to a Computer Science-based Multidisciplinary research area. Despite the many analyses possible, there are two main aspects to any research at both perspectives (Social and Computer Science): (*i*) how to collect and manage social data, and (*ii*) how to build and analyze the social networks derived from such data.

Furthermore, extracting and analyzing relevant information from social networks provide many challenges for developers, user, and technology. For developers, after collecting data from collaborators, it is necessary to model, store and manage them within databases with proper interface to whatever application uses them. For users, when they need to obtain relevant knowledge from these networks. For technology, which should provide the necessary support for implementation of methodologies. For instance, exploring collaborative relations can improve the accuracy and quality of existing methods that combine bibliometry, altmetrics and academic social analysis.

A specific perspective of evaluation is given by *academic social networks*, in which nodes represent researchers and edges their co-authorships and academic relations. Building the structure of such networks is relatively simple, as the nodes are given by any set of researchers who are connected through their common published work, for example. However, one central aspect of more complex analysis is the *strength of the ties* among researchers, as pairs of researchers have stronger or weaker connections depending on the degree of academic relationship. Such degree of relationship (or *tie strength*) may be defined according to Granovetter's theory: the ties are *weak* when they serve as bridges in the network by connecting users from different groups, and *strong* when they link individuals in the same group (community) [1].

Formally, tie strength may be measured by a combination of the amount of time, the cooperation intensity and the reciprocal services that characterize the tie [1]. Such strength may also be measured by using the neighborhood overlap metric (also known as topological overlap), a numerical quantity that captures the total number of collaborations between the two ends of each edge. Note that neighborhood overlap has been used for uncovering the community structure and analyzing structural properties of a large network of mobile phone users, besides measuring tie strength [2].

As for practical applications, the study of social ties has lead to building rigorous models that reveal the evolution of social networks and their dynamism. Indeed, the strength of ties has been largely explored in different contexts, such as information diffusion, analyses of patterns in communication logs and evaluation of scientific researchers productivity [1-6]. Moreover, analyzing tie strength allows investigating how distinct relationships play different roles and identifying impact at micro-macro levels in a network.

Finally, properly measuring the strength of co-authorship ties may help to identify which collaborations are more influent to each researcher. For example, if a researcher *A* collaborates with other researchers *B* and *C*, the strength of ties reveals which one is more important to *A*, then allowing different studies, such as team formation analyses. Also, researchers that form mostly weak (or strong) ties in the social network may indicate different collaboration patterns. Among others, a researcher who has many collaborators through single papers, i.e., that person has collaborated only once with many people.

2. Background and Main Definitions

This part of the tutorial goes over the background necessary to understand the tutorial. Specifically: an overview of general concepts for social networks (building process, modeling, metrics and properties) and social professional networks examples and features.

3. General Taxonomy for Professional Networks

We present a taxonomy based on the tasks and issues of social networks. By analyzing the publications on the area, we have identified (*i*) two main tasks: analysis and applications; and (*ii*) two main issues: data acquisition and preparation, and data storage. After covering such a taxonomy, we further describe a specific analysis (clustering) and two important applications (recommendation and ranking). This part of the tutorial is mostly based on the paper [3].

4. Tie Strength: Concepts and Existing Metrics

Tie strength in social networks has been addressed with diverse goals such as measuring the strength of weak ties [1], co-authorship ties [2], contact ties [4], friendship ties [5] and work ties [6]. Such studies contextualize the importance of measuring tie strength in an appropriate manner: relationships play different roles and should be distinctly qualified as well through (for example) their strength. Indeed, studies show that a relationship has large impact at micro-macro levels in the network, depending on its strength, and influences the patterns of communications [1, 2, 5]. Tie strength can be calculated by considering topological and/or semantic properties. This part of the tutorial goes over such properties and presents an analysis over co-authorships social networks of three different research areas (Medicine, Computer Science and Sociology), mostly based on the paper [2].

5. New Metrics for Tie Strength and their Applications

This part goes over new metrics for calculating tie strength over static networks as well as temporal networks, and is divided in (mostly based on the papers [7, 8, 9, 10]):

a. An analysis of problems of measuring tie strength using solely neighborhood overlap or co-authorship frequency.

b. A new tie strength metric to non-temporal social networks.

c. An analysis of tie persistence and transformation in temporal social networks by using an existing algorithm and a new one.

6. Open Problems and Possible Future Work

Finally, besides open problems and possible future work, we go over our current research over social coding networks. This part includes the paper *Collaboration Strength Metrics and Analyses on GitHub*, recently accepted for publication at IEEE/WIC/ACM Web Intelligence.

References

[1] M. Granovetter. The strength of weak ties. american joumal af sociology, 78:1360-1380. Dietz. Pugh. and Wiley, 91(2004):423--433. 1973.

[2] Michele A. Brandão and Mirella M. Moro. Analyzing the strength of co-authorship ties with neighborhood overlap. In *Proceedings of the International Conference on Database and Expert Systems Applications* (DEXA), pages 527-542. 2015.

[3] Michele A. Brandão and Mirella M. Moro. Social professional networks: A survey and taxonomy. *Computer Communications* 100: 20–31. 2017.

[4] J. Wiese et al., You never call, you never write: Call and sms logs do not always indicate tie strength. In *Procs. of CSCW*, pages 765--774. 2015.

[5] M. Zignani, S. Gaito, and G. P. Rossi. Predicting the link strength of newborn links. In *Proceedings of International Conference on World Wide Web - Companion Volume*, pages 147--148. 2016.

[6] D. Castilho, P. O. Vaz de Melo, and F. Benevenuto. The strength of the work ties. *Information Sciences*, 375:155--170. 2017.

[7] Michele A. Brandão and Mirella M. Moro. The strength of co-authorship ties through different topological properties. *Journal of the Brazilian Computer Society* 23:5. 2017.

[8] Michele A. Brandão, Pedro O. S. Vaz de Melo and Mirella M. Moro. Tie Strength Persistence and Transformation. In *Proceedings of the Alberto Mendelzon Workshop* (AMW), 2017.

[9] Michele A. Brandão and Mirella M. Moro. Strength of Co-authorship Ties in Clusters: a Comparative Analysis. In *Proceedings of the Alberto Mendelzon Workshop* (AMW), 2017.

[10] Michele A. Brandão, Pedro O. S. Vaz de Melo, Mirella M. Moro. Tie Strength Dynamics over Temporal Co-authorship Social Networks. In *Proceedings of IEEE/WIC/ACM Web Intelligence* (accepted for publication), 2017.

Authors' Bio

Michele A. Brandão has recently finished her PhD in Computer Science at UFMG (Belo Horizonte, Brazil), with title *Tie strength in co-authorship social networks: analyses, metrics and a new computational model*. She holds a Masters degree from UFMG, and a Bachelor from Universidade Estadual de Santa Cruz, both in Computer Science. She is currently at UFMG as staff (*"professor substituto"*). Her research interests are mostly in data science, social networks, recommender systems and link prediction. Recent publications include papers: at journals Computer Communications, Journal of the Brazilian Computer Society, and JIDM; as well as at conferences IEEE/WIC/ACM Web Intelligence, BRASNAM, AMW, DEXA, and SBBD.

Mirella M. Moro is associate professor at the Computer Science department at UFMG (Belo Horizonte, Brazil). She holds a Ph.D. in Computer Science (University of California Riverside - UCR, 2007), and MSc and BSc in Computer Science as well (UFRGS, Brazil). She is a member of the ACM Education Council, ACM SIGMOD, ACM SIGCSE, ACM-W, IEEE, IEEE WIE, SBC, and MentorNet. She was the Education Director of SBC (Brazilian Computer Society, 2009-2015), the editor-in-chief of the electrocnic magazine SBC Horizontes (2008-2012), and associated editor of JIDM (2010-2012). Mirella has been working with research in Computer Science in the area of Databases since 1997. Her research interests include social networks analysis, query optimization, and hybrid SQL/NoSQL modeling. Her recent publications include papers on prestigious venues such as Scientometrics, Data & Knowledge Engineering, ACM Hypertext, IEEE/WIC/ACM Web Intelligence, TPDL, JCDL as well as JIDM and SBBD.

32th Brazilian Symposium on Databases

talks

October 2nd to 5th, 2017
Uberlândia – MG – Brazil

INVITED TALKS

Promotion

Brazilian Computer Society – SBC
SBC Special Interest Group on Databases

Organization

Universidade Federal de Uberlândia – UFU

Table of Contents (Invited Talks)

paper:talk1

A Holistic View of Human Factors in Crowdsourcing

Sihem Amer-Yahia

For over 40 years, organization studies have examined human factors in physical workplaces and their influence on the ability of an individual to perform a task, or a set of tasks, alone or in collaboration with others. In a virtual marketplace, the crowd is typically volatile, its arrival and departure asynchronous, and its levels of attention and accuracy diverse. This has generated a wealth of new research ranging from studying workers' fatigue in task completion to examining the role of motivation in task assignment. I will review such work and argue that we need a holistic view to take full advantage of human factors such as skills, expected wage and motivation, in improving the performance of a crowdsourcing platform.

Sihem Amer-Yahia is a CNRS Research Director in Grenoble where she leads the SLIDE team. Her interests are at the intersection of large-scale data management and data analytics. Before joining CNRS, she was Principal Scientist at QCRI, Senior Scientist at Yahoo! Research and Member of Technical Staff at at&t Labs. Sihem served on the SIGMOD Executive Board, the VLDB Endowment, and the EDBT Board. She is Editor-in-Chief of the VLDB Journal. Sihem is PC co-chair for VLDB 2018. Sihem received her Ph.D. in CS from Paris-Orsay and INRIA in 1999, and her Diplôme d'Ingénieur from INI, Algeria.

paper:talk2

What Non-Volatile Memory Means for the Future of Database Management Systems

Andy Pavlo

The advent of non-volatile memory (NVM) will fundamentally change the dichotomy between memory and durable storage in database management systems (DBMSs). These new NVM devices are almost as fast as DRAM, but all writes to it are potentially persistent even after power loss. Existing DBMSs are unable to take full advantage of this technology because their internal architectures are predicated on the assumption that memory is volatile. That means when NVM finally arrives, just like when you finally passed that kidney stone after three weeks, everyone will be relieved but the transition will be painful. Many of the components of legacy DBMSs will become unnecessary and will degrade the performance of data intensive applications.

In this talk, I discuss the key aspects of DBMS architectures that are affected by emerging NVM technologies. I then describe how to adapt in-memory DBMS architectures for NVM. I will conclude with a discussion of a new DBMS that we have been developing at Carnegie Mellon that specifically designed to leverage the persistence properties of NVM in its architecture, such as its recovery and concurrency control mechanisms. Our system is able to achieve higher throughput than existing approaches while reducing the amount of wear due to write operations on the device.

Andy Pavlo is an Assistant Professor of Databaseology in the Computer Science Department at Carnegie Mellon University. At CMU, he is a member of the Database Group and the Parallel Data Laboratory. His work is also in collaboration with the Intel Science and Technology Center for Big Data and Visual Computing Systems. He was a recipient of the 2014 ACM SIGMOD Jim Gray Dissertation Award.

Pesquisador Homenageado do SBBD 2017

José Palazzo Moreira de Oliveira

Como professor e pesquisador, desde minha formatura na Escola de Engenharia da UFRGS, tive a rara sorte de acompanhar o desenvolvimento da Computação e do ensino de Banco de Dados nas universidades brasileiras. Minha ontogênese acadêmica acompanhou o percurso da história do SBBD. Esta distinção foi uma grande alegria e surpresa, em uma época em que a avaliação de um pesquisador é constituída quase exclusivamente por índices bibliométricos, em receber um reconhecimento pelo *conjunto da obra*. Algo muito relevante para mim foi que os jovens colegas se lembraram de uma carreira de 48 anos com forte dedicação à área de Sistemas de Informação e Banco de Dados. Ao longo da carreira desenvolvi atividades em múltiplas dimensões, 81 alunos de pós-graduação já orientados, muitas disciplinas ministradas, forte interação internacional e um consistente número de boas publicações. Tinha que decidir o formato desta apresentação, uma alternativa seria descrever tecnicamente minhas pesquisas, representadas pelas publicações, isto seria enfadonho e traria pouca contribuição para os jovens membros a comunidade. Pensei melhor e então resolvi apresentar as áreas de pesquisa em que tenho trabalhado e sua evolução ao longo destes anos, sem entrar em profundos detalhes técnicos. Este andamento seguiu muito de perto a evolução do SBBD. Após apresento uma perspectiva do futuro dos Bancos de Dados e os perigos que corremos. Uma das atividades realizadas na Comissão Especial de BD e que considero importante foi a implementação do 1° Concurso de Teses e Dissertações em Banco de Dados. Desejo que a apresentação seja útil para os jovens pesquisadores conhecerem melhor o caminho percorrido até aqui pela nossa comunidade e para que entrevejam o possível futuro e seus desafios. A vida acadêmica não pode ser uma *Torre de Marfim*, a preocupação e engajamento com a comunidade é essencial. Nesta apresentação vocês terão a oportunidade de conhecer, de forma agradável, o desenvolvimento de nossa área no Brasil em paralelo com uma análise do que considero essencial para uma carreira equilibrada no ensino e na pesquisa. A história dos Bancos de Dados inicia com a estruturação de arquivos tradicionais e chega aos complexos sistemas atuais. As noções de transação, recuperação e outras são essenciais para a maioria das aplicações transacionais. Hoje há uma revolta contra tudo isto propondo alternativas como o NoSQL, mas diferentes aplicações exigem diversos modelos de SGBDs. Talvez estejamos exagerando nas *customizações*. O que nos reserva o futuro? Como vamos estruturar nossas carreiras em um período turbulento?

José Palazzo Moreira de Oliveira é Professor Titular por concurso público do Instituto de Informática da UFRGS, atualmente atuando como docente convidado. Possui graduação em Engenharia Elétrica pela Universidade Federal do Rio Grande do Sul (1968), mestrado em Ciências da Computação pela Universidade Federal do Rio Grande do Sul (1976) e doutorado em Informática pelo Instituto Nacional Politécnico de Grenoble (1984). Tem experiência na área de Ciência da Computação, com ênfase em Sistemas de Informação, atuando principalmente nos seguintes temas: ontologia, modelagem conceitual, ensino a distância, banco de dados, sistemas de informação e sistemas na Web. É conselheiro da Sociedade Brasileira de Computação. Foi Coordenador do PPGC/UFRGS, participou da criação dos programas de doutorado em Computação e Administração da UFRGS, foi vice-presidente e membro da Câmara de PG da UFRGS, membro do Comitê Assessor em Ciência da Computação do CNPq – CA-CC, coordenador do Comitê de Matemática, Estatística e Computação – MEC – da Fundação de Amparo à Pesquisa do RS – FAPERGS, implantou o Curso de Informática Instrumental para professores do Ensino Médio oferecido pela UFRGS para a UAB. Tendo sido coordenador e membro da Comissão Especial de Banco de dados da SBC e membro da Comissão de Educação da SBC.

REALIZAÇÃO

SBC
Sociedade Brasileira de Computação

ORGANIZAÇÃO

UFU

Faculdade de Computação

APOIO

ProPP

UBERLÂNDIA
CONVENTION & VISITORS BUREAU

FOMENTO

CAPES

FAPEMIG

CNPq

PATROCÍNIO DIAMANTE

Itaú

NVIDIA.

PATROCÍNIO OURO

Algar Telecom

PATROCÍNIO PRATA

Google

IBM Research | Brasil

PATROCÍNIO BRONZE

SEBRAE

neppo

SANKHYA
GESTÃO DE NEGÓCIOS

ClickPerformance®