Speaker Biographies for the

Person Validation and Entity Resolution Conference

May 23rd, 2011, Washington D.C.

United States™ Census Bureau

NORC at the UNIVERSITY of CHICAGO

Contents

Session I: International Papers

Antoine Chevrette 4

Glink, A Probabilistic Record Linkage System

Michael Wenzowski 4

Glink, A Probabilistic Record Linkage System

Richard Madden 4

Indigenous Life Expectancy using Multiple Australian Data Sources

Lisa Jackson-Pulver 5

Indigenous Life Expectancy using Multiple Australian Data Sources

Andrew Bate 5

A Hit-Miss Model for Duplicate Detection in the WHO Drug Safety Database

Session II: Fuzzy Matches and Blocking

Theresa L. Lamagni 6

Application of Probabilistic Linkage Methods to Join Infectious Disease Surveillance Records to Death Registrations

Surajit Chaudhuri 6

Error-Tolerant Record Matching

Prasenjit Mitra 7

Identifying Value Mappings for Data Integration: An Unsupervised Approach

Craig Knoblock 7

Learning Blocking Schemes for Record Linkage

Session III: Dynamic Processing

Lise Getoor 8

Collective Entity Resolution

Marianne Winglee 8

A Case Study in Record Linkage

Xin Luna Dong 8

Conflicting Data: The Role of Source Dependence

SESSION I: INTERNATIONAL PAPERS

ANTOINE CHEVRETTE, STATISTICS CANADA

Glink, A Probabilistic Record Linkage System

Antoine Chevrette is the project leader for the development and support of the generalized probabilistic record linkage system (G-Link) at Statistics Canada. Over the years he has developed expertise in combining state of the art software technologies in order to translate mathematical theory into fast, reliable, and easy to use software applications. He was primarily responsible of the complete G-Link redesign implementation, and is now in charge of managing all aspects of G-Link support and enhancements. Finally, he continues to play a key role in facilitating discussion and the exchange of experiences and ideas within the Statistic Canada record linkage user community.

MICHAEL WENZOWSKI, STATISTICS CANADA

Glink, A Probabilistic Record Linkage System

Michael Wenzowski is the manager responsible for the development and support of the generalized systems for both record linkage and computerized coding at Statistics Canada. His long history of software development, coupled with significant experience in designing and building generalized systems, allows him to approach the generalized implementation of such problem domains from both software engineering and practical utility perspectives. Mr. Wenzowski was the architect of Statistics Canada's ACTR coding software, and was the chief architect for the coding (G-Code) and record linkage (G-Link) systems, which were recently completely redesigned, and re-engineered.

RICHARD MADDEN, UNIVERSITY OF SYDNEY

Indigenous Life Expectancy using Multiple Australian Data Sources

Richard Madden is Professor of Health Statistics at the University of Sydney. He is a statistician and an actuary. From 1996 to 2006, he was Director of the Australian Institute of Health and Welfare, following three years as Deputy Australian Statistician. In earlier years, he held a variety of senior posts in health and disability administration in Australia, and was head of the Treasury Department for both the Northern Territory and the Australian Capital territory. Richard holds a PhD from Princeton University.

Lisa Jackson-Pulver, University of New South Wales

Indigenous Life Expectancy using Multiple Australian Data Sources

Lisa is Associate Professor in Aboriginal Health (Development and Research) at the University of New South Wales and heads the Muru Marri Indigenous Health Unit which she co-founded. She is involved in contributing to Indigenous public health policy at a state, national and international level. In 2004, Lisa addressed the UK House of Commons on the state of Australian Indigenous health, an experience she describes as a "defining moment" of her career. Simultaneously, Lisa also works at a grass-roots level to provide more opportunities for Indigenous people. One of her most successful initiatives has been to establish an innovative philanthropic program at the University of New South Wales, funded in part by the sale of Indigenous art, which has provided 11 full scholarships for Indigenous students to study medicine.

Andrew Bate, Pfizer Inc.

A Hit-Miss Model for Duplicate Detection in the WHO Drug Safety Database

Andrew is Senior Director, Analytics Team Lead in Epidemiology at Pfizer Inc. This position oversees the provision of methodological and analytic expertise to the Epidemiology group in support of drug development and safety evaluation activities worldwide. Prior to joining Pfizer in 2009, Andrew was at the WHO Collaborating Centre for International Drug Monitoring for more than 12 years, where he was responsible for research at the institute. Andrew holds a Masters degree in Chemistry from Oxford University, and a PhD in Clinical Pharmacology from Umea University, Sweden. Additionally, he is a Visiting Professor of Computing and Mathematics at Brunel University, London and an Adjunct Associate Professor of Clinical Pharmacology at New York University (NYU) as well as serving on the faculty of the Center for Health Informatics and Bioinformatics at NYU. Andrew's research career has focused on the development for methods and tools to protect patient safety through the analysis of diverse types of observational health care data including spontaneous reports of adverse drug events, adverse events, Electronic Medical Records (EMRs) and insurance claims data. He has over 30 publications in the Lancet, British Medical Journal, Statistics in Medicine, Data Mining and Knowledge Discovery and other peer reviewed international journals, in addition to many book chapters and full peer reviewed conference proceedings and has made over 50 conference presentations at international symposia. Andrew has contributed to several inter-national initiatives and partnerships associated with medicinal safety. He is an invited member of the FDA Science Board Subcommittee on Pharmacovigilance charged by Janet Woodcock, Director of CDER with conducting a review of CDER's current and planned PV practices. Andrew was an expert adviser to the Committee for Medicinal Products for Human Use (CHMP), European Agency for the Evaluation of Medicinal Products from 2003 to 2009.Additionally, he served as a member of Council for International Organizations of Medical Sciences (CIOMS), the international, non-governmental, non-profit organization established by WHO and UNESCO, on working group VIII which produced a report on practical aspects of signal detection of adverse effects of medicinal products Andrew is a member of the Advisory Board of OMOP (Observational Medical Outcomes Partnership), an FNIH coordinated public-private partnership to help improve the monitoring of drugs for safety.

SESSION II: Fuzzy Matching and Blocking

Theresa L. Lamagni, Health Protection Agency, U.K.

Application of Probabilistic Linkage Methods to Join Infectious Disease Surveillance Records to Death Registrations

Dr. Theresa Lamagni graduated in Psychology in 1994 and then worked in the voluntary-sector (drug misuse services) for a year. Theresa came to work for the Public Health Laboratory Service Communicable Disease Surveillance Centre in 1995, working on various HIV and STI surveillance programmes. During this time she also completed an MSc in Epidemiology at the London School of Hygiene & Tropical Medicine. In May 2001, she joined the newly formed Healthcare Associated Infection & Antimicrobial Resistance Department, now part of the Health Protection Agency Centre for Infections, as a senior epidemiologist. She is involved in a range of national and international projects concerned with developing new surveillance initiatives, research programmes and public health policies relating to streptococcal, fungal and healthcare-associated infections. She completed her PhD on the epidemiology of severe group A streptococcal infections in Europe at the University of Helsinki in 2008.

Surajit Chaudhuri, Microsoft

Error-Tolerant Record Matching

Surajit Chaudhuri is a Research Area Manager at Microsoft Research, Redmond. He started the AutoAdmin project on self-tuning database systems. Surajit has also worked in the area of data cleaning. Their research on both physical database design and data cleaning has been incorporated in Microsoft products and services such as SQL Server and Bing. Surajit did his Ph.D. from Stanford University and he is an ACM Fellow. He was awarded the ACM SIGMOD Contributions award in 2004 and a 10-year VLDB Best paper Award in 2007.

Prasenjit Mitra, The Pennsylvania State University

Identifying Value Mappings for Data Integration: An Unsupervised Approach

Prasenjit Mitra is an Associate Professor in the College of Information Sciences and Technology; he serves on the graduate faculty of the Department of Computer Sciences and Engineering and is an affiliate faculty member of the Department of Industrial and Manufacturing Engineering at The Pennsylvania State University. His major research interests are in exploring issues in information extraction, information integration and information visualization. His research is being supported by the NSF CAREER Award. Additionally, his research has been supported by the NSF, Microsoft Corporation, DoD, DHS, DoE, NGA, and DTRA. He obtained his Ph.D. in Electrical Engineering from Stanford University in 2004. Prior to that, he obtained his M.S. in Computer Science from The University of Texas at Austin in 1994 and his B. Tech. (Hons.) from the Indian Institute of Technology, Kharagpur in 1993. From 1995 to 2000, he was a Senior Member of the Technical Staff at Oracle Corporation in the Oracle Parallel Server and Languages and Relational Technologies groups in the Server Technologies division. He also serves in the Board of Advisors of Global IDs, Inc. Mitra has co-authored over sixty articles at top conferences and journals. His work along (with his co-authors) resulted in a visual analytics system that was awarded the IEEE VAST '08 Grand Challenge award in the Data Integration area. He has served as the co-chair of three workshops including WIDM'09 and served on the PC of several conferences including SIGMOD, AAAI, IJCAI, WWW, CIKM, and ICDM.

Craig Knoblock, University of Southern California

Learning Blocking Schemes for Record Linkage

Craig Knoblock is a Research Professor in Computer Science at the University of Southern California (USC) and the Director of Information Integration at the USC Information Sciences Institute. He received his Bachelor of Science degree from Syracuse University, and his Master's and Ph.D. from Carnegie Mellon University, all in computer science. Dr. Knoblock is also a founder of Fetch Technologies, a web extraction and integration provider, and of Geosemble Technologies, which develops geospatial data integration solutions. At USC, Dr. Knoblock leads a team of about 20 researchers, staff and students in developing techniques for rapid, efficient information integration. He focuses on constructing distributed, integrated applications from online sources through information extraction, source modeling, record linkage, constraint reasoning and other technologies for geospatial and bioinformatics data integration. Dr. Knoblock is a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI), a Distinguished Scientist of the Association of Computing Machinery (ACM), a Trustee of the International Joint Conference on Artificial Intelligence (IJCAI), and past President of the International Conference on Automated Planning and Scheduling (ICAPS). He has served on the Senior Program Committee of the National Artificial Intelligence Conference, among others, and is conference chair for the 2011 International Joint Conference on AI (IJCAI). Dr. Knoblock has published Generating Abstraction Hierarchies (Kluwer Academic Publishers, 1993), along with more than 200 journal articles, book chapters and conference papers. He serves on the Boards of several journals, including Artificial Intelligence and the Journal of Web Semantics.

SESSION III: DYNAMIC PROCESSING

LISE GETOOR, UNIVERSITY OF MARYLAND, COLLEGE PARK

Collective Entity Resolution

Lise Getoor is an associate professor in the Computer Science Department at the University of Maryland, College Park. She received her PhD from Stanford University in 2001. Her current work includes research on link mining, statistical relational learning and representing uncertainty in structured and semi-structured data. She has also done work on social network analysis and visual analytics.

She has published numerous articles in machine learning, data mining, database, and artificial intelligence forums. She was awarded an NSF Career Award, is an action editor for the Machine Learning Journal, is a JAIR associate editor, has been a member of AAAI Executive council, and has served on a variety of program committees including AAAI, ICML, IJCAI, KDD, SIGMOD, UAI, VLDB, and WWW.

MARIANNE WINGLEE, WESTAT

A Case Study in Record Linkage

Marianne Winglee is a senior statistician at Westat, a consulting company in Maryland. Marianne works in the Statistical Group specializing in Survey Design and Data Quality.

XIN LUNA DONG, AT&T

Integrating Conflicting Data: The Role of Source Dependence

Dr. Xin Luna Dong is a researcher at AT&T Labs-Research. She received a Ph.D. in Computer Science and Engineering from University of Washington in 2007, received a Master's Degree in Computer Science from Peking University in China in 2001, and received a Bachelor's Degree in Computer Science from Nankai University in China in 1998. Her research interests include databases, information retrieval and machine learning, with an emphasis on data integration, data cleaning, personal information management, and web search. She has led the Solomon project, whose goal is to detect copying between structured sources and to leverage the results in various aspects of data integration, and the Semex personal information management system, which got the Best Demo award (one of top-3) in Sigmod'05. She co-chaired WebDB'10, and has served in the program committee of Sigmod'11, VLDB'11, PVLDB'10, WWW'10, ICDE'10, VLDB'09, etc.