Fundación BBVA

The BBVA Foundation and the Spanish National Research Council (CSIC) are pleased to invite you to the lecture:

Demography Today

A Novel Temporal Clustering Technique and Quality Evaluation Measure for Group Record Linkage

Prof. Peter Christen

Research School of Computer Science, The Australian National University, Canberra, Australia

Summary:

Research in the social sciences is increasingly based on large and complex data collections, where individual data sets from different domains need to be linked to allow advanced data analysis. A popular type of data used in such a context are historical registries containing birth, death, and marriage certificates. Once such data sets are linked, pedigrees for full populations can be constructed. These will facilitate novel studies to, for example, investigate how education, health, mobility, and employment have influenced the lives of people over several generations. In this talk I will present our recently developed novel temporal clustering approach which is aimed at linking records for a group of individuals, such as all births by the same mother, and where temporal constraints need to be enforced, such as intervals between births. We then present a novel cluster quality evaluation measure that categorizes each individual record according to the quality of the cluster the record has been linked into. Experiments on a real Scottish data set show the superiority of our novel temporal clustering approach over a previous approach for group record linkage, while also highlighting the need for novel quality evaluation measures for group record linkage. This work was conducted with Ms Charini Nanayakkara and Dr Thilina Ranbaduge.

For details about the temporal techniques see: https://link.springer.com/chapter/10.1007/978-3-030-16145-3_41

Biography:

Peter Christen is a professor at the Research School of Computer Science at the Australian National University. He received his Diploma in Computer Science Engineering from ETH Zurich in 1995 and his PhD in Computer Science from the University of Basel in 1999. His research interests are in data mining and record linkage, with a focus on machine learning and privacy-preserving techniques for record linkage. He has published over 140 articles in these areas, including in 2012 the book *Data Matching* published by Springer. For more details see: http://cs.anu.edu.au/people/Peter.Christen/

Charini Nanayakkara is currently working as a PhD student at the Australian National University (ANU), where the focus of her research is on record linkage techniques for complex historical birth, marriage, death, and census data. She received her BSc (Hons) degree in Computer Science from the University of Colombo School of Computing, Sri Lanka, in 2016. Prior to joining the ANU as a PhD student in March 2018, she was employed as a software engineer at WSO2 Lanka Pvt. Ltd for two years. Charini's research is part of the Digitising Scotland project (https://www.lscs.ac.uk/projects/digitising-scotland/). Her publications can be found at: https://scholar.google.com.au/citations?user=bQwdgp0AAAAJ&hl=en

Thilina Ranbaduge is a research fellow at the Australian National University (ANU) Research School of Computer Science. His research interests are in data mining, and in multidatabase and privacy-preserving record linkage. He received his PhD in Computer Science from the ANU in 2018 and completed his PG.Dip and BSc (Hons) at the University of Moratuwa, Sri Lanka, in 2013 and 2009 respectively.

The **Demography Today** lecture series aims to **promote and communicate** scientific work on demography through the dissemination of research and the **training of specialists** in issues related to demography, Big Data, longitudinal records and health, while informing society, in an accessible way, about issues currently in the foreground of scientific and political debate, such as the limits to longevity, pension systems, ageing, emerging diseases, migration and low fertility.

This lecture series enjoys the exclusive support of the BBVA Foundation and has been co-organized with the Spanish National Research Council and the LONGPOP project (Methodologies and Data Mining Techniques for the Analysis of Big Data based on Longitudinal Population and Epidemiological Registers).The LONGPOP project has received funding from the European Union's Horizon 2020 research and innovation program under a Marie Sklodowska-Curie grant.

All **lectures** are **available for viewing** on the interactive platform: **www.demografia.tv**

The lecture series also forms part of the Postgraduate Courses run by the Spanish National Research Council (CSIC).

Information and contact:

e-mail: demografia@cchs.csic.es Director of series: Diego Ramiro Fariñas http://demografia.iegd.csic.es

Tel: (34) 916022403 @demografia_csic

Tuesday, July 2 at 9:00. The Lecture will be followed by a 3h Training Course on 'Record Linkage – Introduction, Recent Advances, and Privacy Issues'

Instituto de Estadística y Cartografía de Andalucía. Calle Leonardo Da Vinci, 21. Isla de la Cartuja 41071-Sevilla

REQUEST ATTENDANCE

Please confirm attendance. Limited seating
e-mail: demografia@cchs.csic.es
The lecture will be delivered in English without translation