Abstract

A primary concern of the medical e-research community is the availability of suitable data sets for their analysis requirements. The quantity and dubious quality of data present significant barriers to the application of many automated analysis technologies, including data mining, to the medical and health domain. Publicly available data is frequently poorly coded, incomplete, out-of-date or simply not applicable to the analysis or algorithm being applied. Work has been done to overcome these issues through the application of data linking processes but further complications have been encountered resulting in slow progress.

The use of locally held medical data is difficult enough due to its structural complexity and non-standardised language, however linking data from disparate electronic sources adds the challenges of privacy, security, semantic compatibility, provenance, and governance, each with its own inherent issues. A focal requirement is a mechanism for the sharing of medical and health data across multiple sites which incorporates careful management of the semantics and limitations of the data sets whilst maintaining functional relevance for the end user. Our paper addresses this requirement by exploring recent conceptual modeling and data evaluation methodologies that facilitate effective data linking whilst ensuring the semantics of the data are maintained and the individual needs of the end user are met.

About the speakers

Anna is currently a PhD student in the Knowledge Discovery and Intelligent Systems Laboratory at Flinders University. Her thesis, on the nature of interest in the medical domain and its application to the evaluation of medical data patterns is due to be completed in August 2007.
Denise holds a PhD in Computer Science from Flinders University and is currently working as a postdoctoral research fellow with the Knowledge Discovery and Intelligent Systems Laboratory at Flinders University. Her interests are in database systems, primarily in the development of the mesodata techniques for handling complex domain structures. Her industry work includes contracts with the DSTO and with the AFC.

John is currently the SACITT Chair in Information Technology and Associate Head (Research) in the School of Informatics and Engineering at Flinders University. He joined Flinders in 2000 after 15 years at the Universities of South Australia and Tasmania, which followed 10 years experience in the computing industry as (progressively) a programmer, analyst, project leader and consultant. Prof. Roddick's research specialises in the fields of Data Mining and Knowledge Discovery (specifically in temporal and spatial data mining and as applied to medical and health data, and to defence and security) and in Conceptual Modelling (specifically in enhanced database systems semantics such as schema evolution and temporal and spatial systems design and use).