

Twelfth ERCIM Database Research Group Workshop

The Tenth Anniversary meeting provided the opportunity for a number of ERCIM Working Groups to meet. One that did so was the Database Research Group, one of the longest standing. CWI and CLRC-RAL jointly organised a couple of sessions of presentations on current database activity. Two contributions addressed database performance, but the main emphasis during the workshop was on aspects of data mining.

Stefan Manegold (CWI) presented work on techniques within the Monet main memory database system to exploit high level caching to counter the cost of the main memory bottleneck - memory access times are falling behind the relentless increase in CPU speed. The other performance paper was from Peter Bosch who described recent work done at Twente on "Clockwise". The problem tackled is how to schedule real time deadlines for a mix of conventional and bulky (video streams) data on the same disc, given that disc service times are not known.

An important aspect of data mining, namely data preparation, was covered in a joint paper from LGU (London Guildhall University) and RAL. It was presented by Paul Jermyn. He described work in progress to develop an approach to data cleaning based on the concept of "Clean Views" in which alternative sets of data preparation operations are well defined and supported by the system. From CWI, Robert Castello spoke about the development (with Arno Siebes) of graphical models that can help guide insight during data mining, while Menzo Windhouwer explained how multimedia objects are indexed in the Acoi system.

In a more general discussion, Brian Read (RAL) tackled the extent to which conventional knowledge discovery in databases as used on business data can usefully be applied to science. Pattern matching algorithms can be very useful for analysing large data volumes, but it is arguable if data mining alone can yield scientific "truth".

Overall it was an enjoyable Workshop and it explored some common interests. Plans were initiated for some joint work in the data mining area.

Three of the Workshop papers are included here. Stefan Manegold's work appears in the Proceedings of the 25th VLDB Conference (Edinburgh, 1999), while the other presentations will appear separately.

Brian Read -
Tel: +44 1235 44 6492
E-mail: Brian.Read@rl.ac.uk

Arno Siebes
Tel: +31 30 253 3229
E-mail: Arno.Siebes@cwi.nl