

Department of Computer Science & Software Engineering

Master Thesis Defense

| | |
|----------------------|--|
| Speaker: | Philon Nguyen |
| Supervisor: | Dr. Shiri |
| Examining Committee: | Drs. Bui, Grahne, and Dr. Harutyunyan (Chair) |
| Title: | Fast and Scalable Similarity and Correlation Queries on Time Series Data |
| Date: | Wednesday April 1, 2009 |
| Time: | 2:00 pm |
| Place: | EV3.101 |

ABSTRACT

Time series are ubiquitous in many fields ranging from financial applications such as the stock market to scientific applications and sensor data. Hence, there has been an increasing interest in time series indexing over the past years because there has been an increasing need for fast methods for analyzing and querying these datasets that are often too big for practical brute force analysis. We start with the main contributions to the field over the past decade and a half. We will then proceed by describing new solutions to correlation analysis on time series datasets using an existing index called the Compact Multi-Resolution Index (CMRI). We describe new algorithms for indexed correlation analysis using Pearson's product moment coefficient and using the multidimensional correlation coefficient and introduce a new measure called Dynamic Time Warping Correlation (DTWC) based on Dynamic Time Warping (DTW). In addition to these linear correlation algorithms, we propose an algorithm called rank order correlation on a non-linear/monotonic measure. To support these algorithms, we revised the Compact Multi-Resolution Index (CMRI) and propose a new index for time series datasets which improves over the sizes, speed and precision of CMRI. We call this index the reduced Compact Multi-Resolution Index (rCMRI). We evaluate the performance of rCMRI compared to CMRI for range queries and range query based queries.