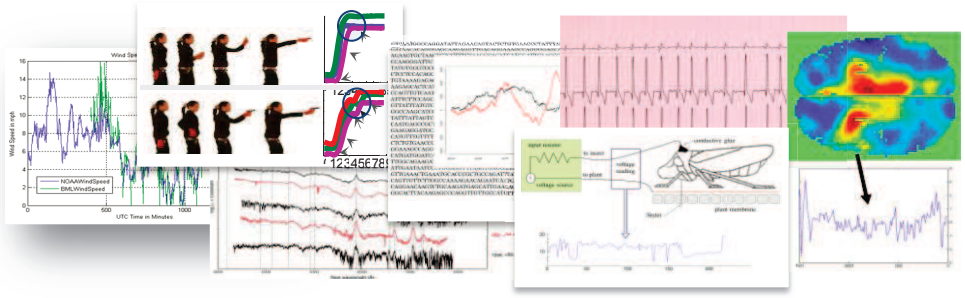
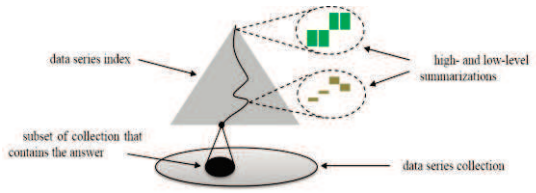
Ultra-Fast Visualizations for Data Series Analytics

Advisors: Prof. Themis Palpanas (Paris Descartes University), Dr. CR1 Theophanis Tsandilas (INRIA Saclay)
themis@mi.parisdescartes.fr *fanis@lri.fr*

The development of sensor technologies in a wide range of domains (e.g., earth observation, astronomy, genome sequencing) has led to an explosion in monitoring activities, which provide a very large amount of data series (i.e., ordered sequences of values).



In order to efficiently process and analyze large volumes of data series, we have to operate on their summaries (or approximations). Several techniques have been proposed in the literature for the approximation of data series [1], including Discrete Fourier Transform (DFT), Piecewise Aggregate Approximation (PAA), Discrete Wavelet Transform (DWT), Symbolic Aggregate approXimation (SAX), and others. Based on these approximations, we can then build indexes that help us answer fast similarity queries on massive collections of data series. Our group has developed the current state of the art data series indexes [2][3]: we have been able to experimentally demonstrate scalability to dataset sizes of 1 billion data series, which is 2-3 orders of magnitude more than the previous approaches. Nevertheless, very little attention has been paid on problems related to visualization and human-computer interaction (HCI) techniques specifically designed for analyzing large data series collections.



The goal of this project is to develop interactive visualization techniques for massive data series collections that are both fast and expressive. While previous work has introduced user interfaces for identifying data series patterns [4], the proposed solution does not scale with the number of data series. We will work on novel HCI operations (e.g., similarity queries, where different points in the query sequence have different weights), as well as back-end algorithms that can support these operations for very large collections of data series.

Accepting this project will make you part of an enthusiastic team working on real, challenging problems!
Prerequisites: experience with file and data structures, excellent programming skills.

References

[1] Themis Palpanas, Michail Vlachos, Eamonn Keogh, Dimitrios Gunopulos. Streaming Time Series Summarization Using User-Defined Amnesic Functions. TKDE 20(7), 2008.
[2] Alessandro Camerra, Jin Shieh, Themis Palpanas, Thanawin Rakthanmanon, Eamonn Keogh. Beyond One Billion Time Series: Indexing and Mining Very Large Time Series Collections with iSAX2+. KAIS 39(1), 2014.
[3] Kostas Zoumpatianos, Stratos Idreos, Themis Palpanas. RINSE: Interactive Data Series Exploration with ADS+. VLDB 2015. http://daslab.seas.harvard.edu/rinse/
[4] Christian Holz and Steven Feiner. 2009. Relaxed selection techniques for querying time-series graphs. UIST 2009.