

Séminaire ETIS : Themis Palpanas

09 Février 2016, 11:00 – 12:00

Titre du séminaire et orateur

Enabling Exploratory Analysis on Very Large Scientific Data.

Themis Palpanas, Université Paris-Descartes.

Date et lieu du séminaire

Mardi 9 février 2016, 11h.

Université de Cergy-Pontoise, site de St-Martin 2, amphithéâtre des colloques.

Abstract

There is an increasingly pressing need, by several applications in diverse domains, for developing techniques able to index and mine very large collections of data series. Examples of such applications come from astronomy, biology, the web, and other domains. It is not unusual for these applications to involve numbers of data series in the order of hundreds of millions to billions.

In this talk, we describe iSAX 2.0 and its improvements, iSAX 2.0 Clustered and iSAX2+, three methods designed for indexing and mining truly massive collections of data series. We show that the main bottleneck in mining such massive datasets is the time taken to build the index, and we thus introduce a novel bulk loading mechanism, the first of this kind specifically tailored to a data series index.

Furthermore, we observe that in several cases scientists, and data analysts in general, need to issue a set of queries as soon as possible, as a first exploratory step of the datasets. Thus, we describe ADS+, an extension of the above techniques that adaptively creates a data series index, and at the same time is able to correctly answer user queries.

We show how our methods allows mining on datasets that would otherwise be completely untenable, including the first published experiments to index one billion data series, and experiments in mining massive data from domains as diverse as entomology, DNA and web-scale image collections.

Bio

[Themis Palpanas](#) is a professor of computer science at the Paris Descartes University (France), where he is a director of the Data Intensive and Knowledge Oriented Systems (diNo) group. He received the BS degree from the National Technical University of Athens, Greece, and the MSc and PhD degrees from the

University of Toronto, Canada. He has previously held positions at the University of Trento and the IBM T.J. Watson Research Center. He has also worked for the University of California, Riverside, and visited Microsoft Research and the IBM Almaden Research Center.

His research solutions have been implemented in world-leading commercial data management products and he is the author of nine US patents. He is the recipient of three Best Paper awards (including ICDE and PERCOM), and the IBM Shared University Research (SUR) Award in 2012, which represents a recognition of research excellence at worldwide level. He has been a member of the IBM Academy of Technology Study on Event Processing, and is a founding member of the Event Processing Technical Society.

He has served as General Chair for VLDB 2013, the top international conference on databases. His research has been supported by the EU, NSF, Hewlett Packard Labs, IBM Research, Telecom Italia, and Facebook.