



## Enabling Exploratory Analysis on Very Large Scientific Data



**Dr. Θέμης Παλπάνας**  
**University of Trento**

### Περίληψη

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. We discuss extensions of our previous techniques that adaptively create data series indexes, and at the same time are 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.

### Σύντομο Βιογραφικό

Themis Palpanas is a professor of computer science at the University of Trento, Italy. 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. Before joining the University of Trento, he worked at the IBM T.J. Watson Research Center. He has also been a Visiting Professor at the National University of Singapore, 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 eight US patents, three of which are part of commercial products in multi-billion dollar markets. He is the recipient of three Best Paper awards. 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 is General Chair for VLDB 2013, has served on the program committees of several top database and data mining conferences, and also serves as a reviewer for the European Commission Framework Programme, the Natural Sciences and Engineering Research Council of Canada (NSERC), the Netherlands Organisation for Scientific Research (NWO), and the Qatar National Research Fund (QNRF).

**Πληροφορίες:** Παναγιώτης Τσαπάρας

**Πέμπτη, 10 Ιανουαρίου 2013 – ώρα 14:00 - 15:00**  
**Αίθουσα Σεμιναρίων, Κτίριο Πληροφορικής**  
**Πανεπιστήμιο Ιωαννίνων**