

Σ Ε Μ Ι Ν Α Ρ Ι Ο Τ Μ Η Μ Α Τ Ο Σ

ΟΜΙΛΗΤΗΣ:

Panos Kalnis
Associate Professor of Computer
Science in the King Abdullah University
of Science and Technology (KAUST)

ΗΜΕΡΟΜΗΝΙΑ:

Παρασκευή, 5 Δεκεμβρίου 2014

ΩΡΑ:

12:00

ΑΙΘΟΥΣΑ:

Αίθουσα Σεμιναρίων (ισόγειο I11)
Κτήριο Τμήματος Μηχανικών Η/Υ & Πληροφορικής

Θ έ μ α

« Using 16,000 Processors to Extract Frequent Patterns from Very Long Sequences »

Περίληψη

Modern applications, including bioinformatics, time series, and web log analysis, require the extraction of frequent patterns, called motifs, from one very long sequence. Existing approaches are either heuristics that are error-prone; or exact combinatorial methods that are extremely slow; therefore applicable only to very small sequences, no more than a few megabytes.

This talk introduces ACME, a combinatorial approach for motif extraction that scales to gigabyte long sequences. ACME is a versatile parallel system that can be deployed on desktop multi-core systems, or on thousands of CPUs in the cloud. However, merely using more compute nodes does not guarantee efficiency, because of the related overheads. To this end, ACME introduces an automatic tuning mechanism that suggests the appropriate number of CPUs to utilize, in order to meet the user constraints in terms of run time, while minimizing the financial cost of cloud resources. Our experiments show that, compared to the state of the art, ACME: (i) supports 3 orders of magnitude longer sequences (e.g., DNA for the entire human genome); (ii) handles large alphabets (e.g., English alphabet for Wikipedia); (iii) scales out to 16,384 CPUs on a supercomputer; and (iv) is the only method to support elastic deployment on the cloud.

Short bio:

Panos Kalnis is associate professor of Computer Science in the King Abdullah University of Science and Technology (KAUST). He is leading the InfoCloud group that focuses on Big Data. In 2009 he was visiting assistant professor in Stanford University. Before that, he was assistant professor in the National University of Singapore (NUS). In the past he was involved in the designing and testing of VLSI chips in the Computer Technology Institute, Greece. He also worked in several companies on database designing, e-commerce projects and web applications. He is an associate editor for the IEEE Transactions on Knowledge and Data Engineering and serves on the editorial board of The VLDB Journal. He received his Diploma in Computer Engineering from the Computer Engineering and Informatics Dept., University of Patras, Greece in 1998 and his PhD from the Computer Science Dept., Hong Kong University of Science and Technology (HKUST) in 2002. His research interests include Big Data, Analytics, Cloud Computing, Distributed Systems, Large Graphs and Sequences.

More information can be found at: <http://web.kaust.edu.sa/faculty/PanosKalnis/>