Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti"
Consiglio Nazionale delle Ricerche

colloquia@iasi

> *"La mente è come un paracadute.*
> *Funziona solo se si apre"*
> *A. Einstein*

Alignment-free Classification and Comparison of Biological Sequences and Structures

Prof. **Raffaele Giancarlo**
Dipartimento di Matematica ed Informatica, Università di Palermo

Abstract

Similarity of sequences is a key mathematical notion for Classification and Phylogenetic studies in Biology. It is currently primarily handled using alignments. However, the alignment methods seem inadequate for post-genomic studies since they do not scale well with data set size and they seem to be confined only to genomic and proteomic sequences. Therefore, alignment-free similarity measures are actively pursued. Those measures offer the additional advantage of being applicable to structural data, as long as the structure is encoded into a string. A series of recent ground-breaking results have given evidence that those techniques are valid alternatives to classic methods, based on alignments, for classification and comparison of biological sequences and structures. In this talk, the state of the art in this area is presented, starting from some basic notions related to alignment algorithms. In depth attention will be given to the Universal Similarity Metric, a methodology based on Kolmogorov Complexity. In particular, the first set of extensive quantitative experiments showing that USM is successfully applicable in Biology will be presented. Moreover, recent related results will also be discussed, namely, (a) extensions of the technique to network comparison; relation between Biological Complexity and Classic Notions of Computational Complexity; (c) Data Compression as a pillar for Computational Biology.

Aula Conferenze, Viale Manzoni 30, Roma Giovedì 6 Ottobre 2011 alle ore 11:00

Aula Conferenze, Viale Manzoni 30, Roma Giovedì 6 Ottobre 2011 alle ore 11:00