

Prelegent: dr François Coste, Inria Rennes

Data: 24/05/2016 godz. 13.15

Miejsce: 304b/D-1

Tytuł: "Learning automata and grammars to model protein families"

Abstrakt:

Given a set of protein sequence performing the same function, we are interested in learning rules characterizing the function at the sequence level that can be used to predict new members of the functional family. After a short presentation of state-of-the-art in Pattern/Motif Discovery which is limited to position-specific characterizations, I will show how we overcome this limitation in Dyliss team by the introduction of partial local multiple sequence alignments to learn automata [1,2,3] and grammars [4,5] from protein sequences.

[1] "A Similar Fragments Merging Approach to Learn Automata on Proteins", F. Coste and G. Kerbellec, ECML 2005

[2] "Learning Automata on Protein Sequences", F. Coste and G. Kerbellec, JOBIM 2006

[3] "Apprentissage d'automates modélisant des familles de séquences protéiques", G. Kerbellec, PhD thesis, Jun 2008.

[4] "Local Substitutability for Sequence Generalization", F. Coste, G. Garet, J. Nicolas. ICGI 2012.

[5] "A bottom-up efficient algorithm learning substitutable languages from positive examples", F. Coste, G. Garet, J. Nicolas, ICGI 2014