



UNIVERSITÀ DI PISA

ALGORITMI PER LA BIOINFORMATICA

NADIA PISANTI

Anno accademico	2016/17
CdS	INFORMATICA
Codice	315AA
CFU	6

Moduli	Settore/i	Tipo	Ore	Docente/i
ALGORITMI PER LA BIOINFORMATICA	INF/01	LEZIONI	48	NADIA PISANTI

Obiettivi di apprendimento

Conoscenze

This course has the goal to give the student an overview of algorithmic methods that have been conceived for the analysis of genomic sequences. We will focus both on theoretical and combinatorial aspects as well as on practical issues such as whole genomes sequencing, sequences alignments, the search of patterns in biological sequences, the inference of repeated patterns and of long approximated repetitions, the computation of genomic distances, and several biologically relevant problems for the management and investigation of genomic data.

Modalità di verifica delle conoscenze

All students are requested to make a final oral report on an assigned topic.

Students understanding is evaluated during the classes as well as with a final oral report.

Students who did not attend the classes will be asked to undergo an oral exam.

The exam (see below for its description) has the main goal of evaluating the students understanding of the problems and the methods described in the course.

Capacità

The students will learn how a scientific paper is like and how to prepare slides for an oral presentation.

Modalità di verifica delle capacità

Besides evaluating student understanding, the exam is additionally meant as a chance to learn how a scientific paper is like, and how to make an oral presentation on scientific/technical topics, that is designed for a specific audience.

Comportamenti

The student will learn how to present a scientific result.

Modalità di verifica dei comportamenti

The lecturer will evaluate the student oral report.

Prerequisiti (conoscenze iniziali)

A basic course on Algorithms.

Programma (contenuti dell'insegnamento)

A brief introduction to molecular biology: DNA, proteins, the cell, the synthesis of a protein.

Sequences Alignments: Dynamic Programming methods for local, global, and semi-local alignments. Optimizations for the alignments of similar sequences. Alignments in linear space. Computing the Longest Common Subsequences. Multiple Alignments.

Pattern Matching on biological sequencing: indexing methods and algorithms based on pattern preprocessing.

Motifs Extraction: KMR Algorithm for the extracion of exact motifs and its modifications for the inference of approximate motifs. Finding Repetitions: Algorithms for the inference of long approximate repetitions. Filters for preprocessing.

Fragment Assembly: Genomes sequencing: some history, scientific opportunities, and practical problems. Some possible approaches for the problem of assembling sequenced fragments. The Greedy solution. Data structures for representing and searching sequencing data. New

Generation Sequencing: Applications of High Throughput Sequencing and its algorithmic challenges. Investigating data types resulting from the existing biotechnologies, and the possible data structures and algorithms for their solution.



UNIVERSITÀ DI PISA

Bibliografia e materiale didattico

Course material is available in the course web site.

Indicazioni per non frequentanti

Students not attending the classes are additionally requested to undergo an oral exam on the topics of the Syllabus.

Modalità d'esame

The Exam assignments works as follows: each student is assigned a paper that is a very recent scientific work on topics related to those of the course (typically it is a paper accepted for publication in the proceedings of an international conference that is going to be held in a few weeks/months). The paper is part of a pool of possible papers selected by the lecturer. The paper assignment follows a brief description of all papers in the pool made by the lecturer, and a bidding phase of the students over such papers. Once the student has his/her paper assigned, the task is to prepare and make a presentation of that work that: (1) describes the results presented in that paper, (2) is suited for the actual audience (that will be the course class) as for comprehension opportunity, (3) sticks to the allowed time slot. Students presentations usually take place all together somewhere at the end of the course. Exceptions are possible upon request for specific needs (e.g. seminar given later than end of the course).

Pagina web del corso

<http://didawiki.di.unipi.it/doku.php/bio/start>

Ultimo aggiornamento 18/05/2017 10:33