

COMP 3820 – Introduction to bioinformatics algorithms

Calendar description: This course is an introduction to the field of bioinformatics, from a computer science perspective. No prior knowledge of biology or biochemistry is required to follow this course. Classical biological problems will be presented along with their computational and algorithmic solutions. Although the biological context will always be presented, the focus will be on the design, programming and analysis of efficient algorithms.

Prerequisite: COMP 2080 (min grade of C).

Outline

- 1) Introduction (2 weeks)
Introduction to biological terms and basic concepts. Introduction to Python and BioPython.
- 2) Exact pattern matching (1 week)
Boyer-Moore-Horspool, Knuth-Morris-Pratt, Aho-Corasick algorithms.
- 3) Sequence alignments (3 weeks)
Local alignment, global alignment, multiple alignment algorithms.
- 4) Homology searching (1 week)
Overview of biological databases, database searching algorithms.
- 5) Phylogenetic tree construction (3 weeks):
Phylogenetics and phylogenetic trees, distance-based and character-based algorithms for phylogenetic tree construction.
- 6) Gene prediction (2 weeks)
Introduction to gene prediction, Markov chains, Hidden Markov Models (HMMs).
- 7) Genome sequencing (1 week)
Introduction to genome sequencing, Sanger sequencing, Next Generation Sequencing (NGS), genome assembly.