

Course guide

270540 - TEB - Techniques and Tools for Bioinformatics

Last modified: 12/07/2022

Unit in charge: Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science.

Degree: MASTER'S DEGREE IN INFORMATICS ENGINEERING (Syllabus 2012). (Optional subject).

Academic year: 2022 **ECTS Credits:** 3.0 **Languages:** Catalan, Spanish

LECTURER

Coordinating lecturer: XAVIER MESSEGUER PEYPOCH

Others:

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:

CDG1. Capability to integrate technologies, applications, services and systems of Informatics Engineering, in general and in broader and multidisciplinary contexts.

CTE7. Capability to understand and to apply advanced knowledge of high performance computing and numerical or computational methods to engineering problems.

CTE9. Capability to apply mathematical, statistical and artificial intelligence methods to model, design and develop applications, services, intelligent systems and knowledge-based systems.

Generical:

CG4. Capacity for mathematical modeling, calculation and simulation in technology and engineering companies centers, particularly in research, development and innovation tasks in all areas related to Informatics Engineering.

Transversal:

CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.

Basic:

CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

TEACHING METHODOLOGY

The course is proposed with 16h of theoretical sessions, to understand the theoretical foundations of the techniques used, and 10h of computer laboratory sessions to know the main tools of the topic.

The theoretical sessions (and problems) will be given on whiteboard with slides..

The laboratory sessions will be devoted to practice with programs designed for the subject.

LEARNING OBJECTIVES OF THE SUBJECT



STUDY LOAD

Type	Hours	Percentage
Hours large group	13,5	18.00
Hours small group	13,5	18.00
Self study	48,0	64.00

Total learning time: 75 h

CONTENTS

Efficient search algorithms and data structures

Description:

The most efficient string matching algorithms according to the alphabet, the length and number patterns will be given. Also we will discuss useful data structures for comparing genomes as suffix trees. Finally we will study strategies for exhaustive search in very long DNA sequences (Gb).

Sequence alignment

Description:

The dynamic programming algorithm will be given and its application to some cases: the edit distance between two words, the approximate string matching and the best alignment between two sequences. Finally we introduce the multiple sequence alignment.

Data base searching: BLAST

Description:

The computational and statistical foundations of the BLAST algorithm and its use for approximate searches in databases will be introduced.

Phylogenetic algorithms

Description:

The fundamental algorithms to estimate the evolution of individuals within a species will be explained.

Hidden Markov Models

Description:

The Hidden Markov Models and its applications to bioinformatics will be explained.

GRADING SYSTEM

The course note (NF) has two contributions: a laboratory grade (NL), which becomes from the evaluations of some given exercises in laboratory sessions, and a final exam grade (NE), then

$$NF = NE * 0.7 + NL * 0.3$$



BIBLIOGRAPHY

Basic:

- Mount, D.W. Bioinformatics: sequence and genome analysis. 2nd ed. Cold Spring Harbor Laboratory Press, 2004. ISBN 0879697121.
- Navarro, G.; Raffinot, M. Flexible pattern matching in strings: practical on-line search algorithms for texts and biological sequences. Cambridge University Press, 2002. ISBN 978-0-521-81307-5.