BASIC DETAILS:

Subject:	BASES DE DATOS PARA BIOINFORMÁTICA
Id.:	33301
Programme:	GRADUADO EN BIOINFORMÁTICA. PLAN 2019 (BOE 06/02/2019)
Module:	BIOINFORMÁTICA
Subject type:	OBLIGATORIA

Year:	3	**Teaching period:**	Primer Cuatrimestre
Credits:	3	**Total hours:**	75
Classroom activities:	30	**Individual study:**	45
Main teaching language:	Inglés	**Secondary teaching language:**	Castellano
Lecturer:		**Email:**	

PRESENTATION:

This subject presents the principles needed to integrate different biological data sources. It introduces the most common approaches through a review of their architecture. These architectures are described through practical case studies of currently integrated biological repositories. This subject will provide an overview of XML language and related tools (XLST, XPath, XQuery and XML Schema) used to manage and retrieve information from biological databases and handle the meta-data that enable the different integration architectures. Finally the subject will cover the main matching and mapping techniques that make possible the semantic and syntactic integration of the information.

PROFESSIONAL COMPETENCES ACQUIRED IN THE SUBJECT:

General programme competences	G02	Perform the analysis and synthesis of problems of their professional activity and apply them in similar environments.
	G04	Reason critically based on information, data and lines of action and their application on relevant issues of a social, scientific or ethical nature.
	G05	Communicate professional topics in Spanish and / or English both orally and in writing.
	G06	Solve complex or unforeseen problems that arise during the professional activity within any type of organisation and adapt to the needs and demands of their professional environment.
	G09	Apply information and communication technologies in the professional field.
	G10	Apply creativity, independence of thought, self-criticism and autonomy in the professional practice.
Specific programme competences	E02	Develop the use and programming of computers, databases and computer programs and their application in bioinformatics.
	E04	Program applications in a robust, correct, and efficient way, choosing the paradigm and the most appropriate programming languages, applying knowledge about basic algorithmic procedures and using the most appropriate types and data structures.
	E05	Implement well-founded applications, previously designed and analysed, in the characteristics of the databases.
	E07	Apply the principles, methodologies and life cycles of software engineering to the development of a project in the field of bioinformatics.
	E13	Apply omics technologies for the extraction of statistically significant information and for the creation of relational databases of biodata that can be updated and publicly accessible to the scientific community.
Learning outcomes	R01	Definir los sistemas de información y conocer sus componentes fundamentales.
	R02	Reconocer patrones de arquitecturas de integración (como cliente-servidor), sus reglas, modelos.
	R03	Aplicar patrones de arquitecturas de integración a la anotación "ómica".
	R04	Determinar la necesidad de integrar o no datos provenientes de bases de datos diversas en diferentes entornos.
	R05	Desarrollar programas para gestionar archivos codificados en XML directamente o a través de bases de datos.

PRE-REQUISITES:

It is recommended that students have a global vision about main biological databases and understand basic SQL

syntax.

SUBJECT PROGRAMME:

Subject contents:

| **1 - Introduction** |
| 1.1 - Overview of information systems in bioinformatics |
| 1.2 - Types and requirements |
| **2 - Architectures for Information Integration** |
| 2.1 - Data Warehouse |
| 2.2 - Federated Databases |
| 2.3 - Mediator-base Databases |
| 2.4 - Peer-to-peer Databases |
| **3 - XML language applied to Bioinformatics** |
| 3.1 - Introduction to XML |
| 3.2 - XPath and XSLT |
| 3.3 - DTD and XML Schema |
| 3.4 - XQuery |
| **4 - Schema and meta-data management at information integration systems** |
| 4.1 - Matching techniques |
| 4.2 - Mapping techniques |

Subject planning could be modified due unforeseen circumstances (group performance, availability of resources, changes to academic calendar etc.) and should not, therefore, be considered to be definitive.

TEACHING AND LEARNING METHODOLOGIES AND ACTIVITIES:

Teaching and learning methodologies and activities applied:

Theory Sessions: Lectures will be used to explain the basis of the different chapters. Wherever possible,explanations will be accompanied by images, text or sounds to be used as practical examples and discussion topics. During the sessions, the lecturer will propose activities or to look for information out of the class and she will resolve doubts.

Practical activities: During these sessions, student will see real examples of different integration architectures explained in class, available through different websites, and they will learn to take advantage of mining tools offered at each site to retrieve data. Also, they will apply concepts explained in class with hands-on practice creating XML files. They should be able to expand it with the content explained in class and other bibliographic resources.

The lecturer will be available to students during the tutorial schedule to help them in all matters concerning the course. On request, group tutorials may be programmed to control the work of the group. The concepts explained in one chapter will be used in the followings.

Student work load:

Teaching mode	Teaching methods	Estimated hours
Classroom activities	Master classes	15
	Other theory activities	5
	Practical work, exercises, problem-solving etc.	6
	Coursework presentations	2
	Assessment activities	2
Individual study	Tutorials	3
	Individual study	10

	Individual coursework preparation	15
	Project work	6
	Compulsory reading	5
	Recommended reading	3
	Information research	3
	Total hours:	75

ASSESSMENT SCHEME:

Calculation of final mark:

Written tests:	20	%
Individual coursework:	15	%
Final exam:	45	%
Individual project:	20	%
TOTAL	100	%

*Las observaciones específicas sobre el sistema de evaluación serán comunicadas por escrito a los alumnos al inicio de la materia.

BIBLIOGRAPHY AND DOCUMENTATION:

Basic bibliography:

CERAMI, Ethan. XML for Bioinformatics. Springer, 2005.
LACROIX, Zoe and CRITCHLOW, Terence. Bioinformatics: Managing Scientific Data (The Morgan Kaufmann Series in Multimedia Information and Systems). Morgan Kaufmann Publishers Inc., 2003
Silberschatz, Korth and Sudarshan. Database System Concepts. Mcgraw-Hill S.A, 2007.

Recommended bibliography:

ABITEBOUL, Serge and MANOLESCU, Ioana. Web Data Managment. Cambridge University Press, 2011
CHEN, Ming and HOFESTÄDT, Ralf. Approaches in Integrative Bioinformatics (Towards the Virtual Cell). Springer, 2014.

Recommended websites:

About data warehouse	https://docs.oracle.com/database/121/DWHSG/toc.htm
BaseX	https://basex.org/
BioMart	https://m.ensembl.org/info/data/biomart/index.html
NeuroImaging tools and resources collaboratory	https://www.nitrc.org/
TargetMine	https://targetmine.mizuguchilab.org/
W3C Extensible Markup Language (XML) 1.0	https://www.w3.org/TR/xml/
XML Schema	https://www.w3.org/TR/xmlschema-0/
XPath and XQuery Data Model 3.1	https://www.w3.org/TR/xpath-datamodel-31/
XSL Transformations (XSLT) Version 3.1	https://www.w3.org/TR/xslt-30/

* Guía Docente sujeta a modificaciones